COMBINATION AND CONFUSION: A DEFENSE OF CONSTITUTIVE MICROPSYCHISM

By

Forrest J.S. Schreick

Submitted to
Central European University
Department of Philosophy

In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Supervisor:
Michael V. Griffin

Vienna, Austria

2025

Author's Declaration

I, the undersigned, **Forrest J.S. Schreick**, candidate for the PhD degree in Philosophy declare herewith that the present thesis is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography. I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person's or institution's copyright. I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree.

Forrest J.S. Schreick
Vienna, 28 February 2025

Copyright Notice

[1]

[1] Icon by Font Awesome: https://fontawesome.com/

Abstract

Constitutive micropsychism is the view that fundamental physical role-playing properties are realized by microsubjects who aggregate to constitute macrosubjects like us. This form of Russellian monism combines all the theoretical strengths of materialism and dualism while avoiding the traditional problems of both views. Instead, it faces the many forms of the seemingly intractable combination problem or the problems of explaining how microsubjects and their mental states could aggregate to produce novel subjects. This dissertation provides two solutions to the combination problem, both reject the traditionally popular view that introspection accurately represents the complete essential nature of mental states. My first solution is only available to the materialist constitutive micropsychist. I argue she can rebut two combination problems: the subject-summing and subject decomposition problems. However, while many panpsychists claim to be materialists, I show that the notion of materiality is deeply problematic. My second solution is consistent with any form of constitutive micropsychism. I argue for a restricted form of G.W. Leibniz's confused ideas thesis, that introspection misrepresents the structure of mental states. I show how the thesis enables the constitutive micropsychist to rebut almost all major combination problems.

Acknowledgments

First and foremost, I want to thank my supervisor, Mike Griffin. He has consistently been supportive and enthusiastic about the project. His criticisms have been incredibly helpful in developing these ideas. Despite the abundance of crazy views, philosophy can be a field where minority views are made to feel especially weird. As an advisor, Mike created a fantastic environment to pursue ideas. Further, through a globally, politically, and personally tumultuous handful of years, Mike has been a great source of encouragement and support. I am incredibly grateful for Mike's supervision. On top of all this, he's always down for a game of chess (even if he beats me every time).

In early 2022, I was a visiting doctoral candidate at Cornell University in Ithaca, New York. There I was fortunate to be advised by Derk Pereboom. Derk read and gave comments on almost every chapter. I am very grateful for his time, thoughts, conversation, and criticisms.

I would also like to thank flatmates in Ithaca: Rianna Thomas and Stars, her cat. You both made my time there so much better I almost forgot about how dangerous our apartment was. Thanks also to my best friend, Peter Whitehead, who took the train up to New York to see me and cheer me up when it was desperately needed.

For reading and commenting on various chapters, I would like to thank Dávid Bartha, Sara Boljević, Tim Crane, Howard Robinson, and the members of CEU's Doctoral Work-In-Progress seminar. A special thanks to Jeffrey Watson for his comments and going out of the way to help a former student. Special thanks also to Ruben Noorloos for his friendship and camaraderie in bringing the insights of the early modern rationalists to today's problems in the metaphysics of mind.

For conversation about topics relevant to this research, I am grateful to Dilara Boğa, Kerem Eroğlu, Kati Farkas, Rob Hoveman, Ferenc Huoranzski, Valentina Martinis, Dave McElhoes, Rosamaria Romanelli, Marta Santuccio, Peter Whitehead, and the many others I'm sure that I'm forgetting.

Thanks to my MA advisors, Bernie Kobes, Steven Reynolds, and Ángel Pinillos. It was in their seminars that I first became fascinated with consciousness, skepticism, and Leibniz—all major topics of this dissertation! The way I think about these issues was helpfully shaped by their teaching over four years of undergrad and two years of a masters. Thanks also to Cheshire Calhoun whose advice on writing I've followed for, oh my, ten years now!

Thanks to my parents, Carole Phillips and John Schreick, for their continuing support. Without them, my entire journey into the mysteries of philosophy wouldn't have been possible. Thanks also to Ira and Kathy Latto for their encouragement and support.

Finally, thanks to my partner, Aubrey Spivey. What she hasn't read of the dissertation, she's listened to me talk about endlessly. More than that, she's been my biggest cheerleader,

my greatest source of comfort, and my reason for not giving up. Leibniz may have been wrong that this is the best of all possible worlds, but my world is immeasurably better for having her in it.

Contents

Chapter 1

Introduction

The mind-body problem is the old philosophical problem of explaining how this lumpy, squishy, fleshy stuff relates to our private, first-person experience of an inner life. While the sciences have made extraordinary progress in explaining bodily processes, relatively little progress has been made on the mind-body problem. In fact, proposed solutions to the problem still fall into two main categories: *physicalism* and *anti-physicalism*. The physicalist, impressed with the sciences' successes in explaining other previously mysterious phenomena, proposes that everything is physical. Something is physical if "it is (approximately accurately) treated by current or future (at the end of inquiry, ideal) physics, and is not fundamentally mental."[1] The anti-physicalist, on the other hand, denies that the sciences will be able to solve the mind-body problem. Consciousness is different from the other mysteries that the physical sciences have solved because consciousness is not physical.

The most powerful argument for physicalism is *the causal argument* also known as *the exclusion problem*:

1. Every physical effect has a sufficient physical cause.

2. Mental events cause physical events.

3. Mental causes do not overdetermine their effects.

[1] Jessica Wilson, "Supervenience-Based Formulations of Physicalism," *Noûs* 39 (2005): p. 428. This is a working definition. See section 2.2.

C1. Mental events are physical events.[2]

Premise 1 is a statement of *physical causal closure*. Premise 2 denies that our mental states are epiphenomenal, that they have no causal efficacy. This is supported by the everyday experience of our minds seeming to be able to cause changes to our bodies and, through the body, make changes to other parts of the external world. The *feeling* of thirst will seemingly cause us to reach for a glass of water. The *desire* for a new book seemingly causes us to walk into a bookstore. The *fear* of being hit by the oncoming truck seemingly causes us to move out of the way. We feel as if we are "at the wheel" when it comes to our bodies and not merely a helpless observer. Premise 3 denies causal overdetermination. An effect is overdetermined if it has more than one sufficient cause. Why deny over-determination? Imagine two children, A and B, are both throwing rocks at a window. Each child throws a rock and both rocks hit the window at the same time. Either rock hitting the window would be sufficient to break the window. If A had not thrown her rock, then B's rock would have broken the window and *vice versa*. Now, further imagine that *every* mental effect was overdetermined in this way. Your reaching for the water glass is sufficiently determined by your feeling of thirst, but also some other cause. Had you not felt thirsty, you still would have reached for the glass of water. This would be very strange! Why would there be such a proliferation of redundant causes in the world? Considerations of parsimony push us to reject such a view. The physicalist seeks to show that if mental states are to be efficacious in the world, then they must be physical states.

In contrast, the anti-physicalist wields *the conceivability argument*:

4. P&˜Q is conceivable.

5. If P&˜Q is conceivable, then it is metaphysically possible.

6. If P&˜Q is metaphysically possible, then physicalism is false.

C2. Physicalism is false.[3]

[2]For more on the causal argument or exclusion problem, see section 2.6 and subsection 7.2.3.

[3]This version of the argument is paraphrased from Chalmers (2015, 249). For more on the conceivability argument, see subsection 7.1.2.

P is the conjunction of all physical truths. Q is some truth about consciousness. Premise 4 states the conceivability of a zombie world, physically identical to ours but without consciousness. Our zombie counterparts behave exactly as we do, but lack a mental life. When my zombie counterpart stubs his toe, he swears and holds his food just as I would, but he does not experience what-it-is-like to feel pain. Broadly and with exceptions, a proposition is conceivable if its negation is not *a priori*. Premise 5 states that if the zombie world is conceivable, then it is possible. Some state of affairs is possible if it could have obtained. Premise 6 states that if the zombie world is possible, then physicalism is false. If the zombie world could have obtained, then the physical facts do not necessitate the facts about consciousness. If the physical facts do not necessitate the facts about consciousness, then physicalism is false.

Denying a premise from either argument is a costly endeavor. The anti-physicalist is forced either to deny causal closure (deny premise 1) or to embrace epiphenomenalism (deny premise 2). Causal closure is a widely assumed view in physics and philosophy, though one with few arguments, and epiphenomenalism does away with our self-conception as agents. The physicalist is forced to either embrace a form of analytic functionalism (deny premise 4) or endorse a brute necessary identity between some physical states and mental states (deny premise 5). Analytic functionalism is the view that consciousness can be defined in a wholly functional way. If analytic functionalism is true, then P&˜Q is inconceivable because the conscious truths are a subset of the physical truths. Zombie Forrest's cursing and holding his foot *is* his being in pain. Alternatively, an unexplained, brute relation between physical states and conscious states would be strange in our would given the history of successful scientific reductive explanations.

Not many good options and, worse still, the debate has been at an impasse between the causal argument and the conceivability argument for nearly four hundred years! The conceivability arguments defended today can be traced back to similar arguments at least as far back as René Descartes's arguments in his *Meditations on First Philosophy* to which Princess Elisabeth of Bohemia responded with a causal argument during their correspondence.

What is a mind to do?

The way forward, I think, is *Russellian monism*: the view that science studies and describes only the causal structure of the world, that there must be something serving as the basis for

that structure, and that basis explains consciousness. Most promising, I think, are panpsychist forms of Russellian monism where the basis of physical role-playing properties are phenomenal properties of the same type we experience, though perhaps different in token. Specifically, I wish to defend *constitutive micropsychism*: ordinary subjects, like people or animals, are grounded in the aggregation of microsubjects. Constitutive micropsychism avoids the causal exclusion worries for anti-physicalism. Ordinary subjects inherit the causal efficacy of their constituting microsubjects. Because those microsubjects are the categorical basis for physical role-playing properties, zombie worlds are inconceivable. P&˜Q seems conceivable because we are not duplicating the categorical basis for the physical properties. Constitutive micropsychism attempts to make progress on the mind-body problem by understanding the mind and body to be two ways of describing the same thing as the mind is the categorical basis for the body.

Constitutive micropsychism and other forms of Russellian monism have attracted much attention recently as a promising way forward on the mind-body problem. However, in addition to the incredulous stares, defenders of constitutive micropsychism also face the combination problem. Briefly, the core of the problem is explaining how microsubjects produce new subjects, such as human minds.

Over the years, the combination problem has been refined by numerous philosophers into almost as many unique formulations. Here is a sampling. The problems of the construction and deconstruction of subjects. Subjects are not like Lego; they do not obviously and easily join together to make something new (*the subject-summing problem*) nor do they come apart into discrete units (*subject decomposition problem*). The problem of explaining the rich array of experiences had by ordinary subjects using only the much more limited experiences of microsubjects (*the palette problem*).

Despite the definitive article, the combination problem is a many-headed hydra. The aim of this dissertation is to give the constitutive micropsychist a blade that can strike off all of the hydra's heads at once. It is rusted with age, but was forged in the fires of rationalism's greatest philosopher, G.W. Leibniz. I suggest the constitutive micropsychist adopt a form the *confused ideas thesis*. An idea *i* is confused for a subject *s* if *i* is a mental state of *s*'s and *i* has a structure or content that is not accessible to *s* in introspection. In this dissertation, I scrape the rust off,

develop, and argue for a new form of the confused ideas thesis. I will show how it undermines key premises in the combination problems and leaves combination an open possibility.

1.1 Chapter overview

The dissertation will proceed as follows. Chapter 2 considers materialist forms of constitutive micropsychism. On behalf of the view, I develop and argue for a new solution to two forms of the combination problem: the subject-summing and subject decomposition problem. However, I show that a widely accepted analysis of what it is to be material is insufficient to distinguish between material and ideal entities. Attempts to supplement this analysis both show little promise and render constitutive micropsychism less theoretically attractive. As such, I suggest that materialist constitutive micropsychists consider other views.

Chapters 3 and 4 concern a major alternative to constitutive micropsychism: illusionism about phenomenal consciousness, the view that introspection misrepresents mental states as having qualitative characters they lack. In chapter 3, I evaluate a common argument for illusionism that I call "the arugment from error." I argue that it is a form of skeptical argument and that illusionism is a form of skepticism. I end the chapter by considering anti-skeptical responses to the illusionist modeled on Moorean arguments and dogmatism about perceptual justification. In chapter 4, I provide a novel interpretation of Leibniz's views about secondary quality ideas. I argue that during the period surrounding his writing the *New Essays Concerning Human Understanding*, Leibniz thought our secondary quality ideas were a result of confusion and that he denied their reality in the same way that today's illusionists deny the reality of the qualitative aspects of phenomenal properties.

In chapter 5, I argue for a non-illusionist form of the confused ideas thesis. I argue the thesis is entailed by *constitutivism*, the view that the mind is grounded in, realized by, or identical to material entities. If I am correct, then many views in the metaphysics of consciousness will be committed to the confused ideas thesis. I then show how the confused ideas thesis solves the structural mismatch problem, or the problem accounting for the apparent discrepancy between the structure of mental states and the structure of their plausible material correlates.

In chapter 6, I argue that the conjunction of constitutivism, the confused ideas thesis, and realism about the qualitative character of phenomenal properties entails the disjunction of panpsychism or panprotopsychism, i.e., "*pan(proto)psychism*." While other philosophers have argued from pan(proto)psychism to the confused ideas thesis, I believe this dissertation is unique in attempting to argue from the confused ideas thesis to pan(proto)psychism.

Chapter 7, I argue that the constitutivist, including the constitutive micropsychist, armed with the confused ideas thesis is able to solve most combination problems, including, but not limited to, the subject-summing, subject decomposition, palette, and structural mismatch problems. After surveying and responding to various combination problems, I evaluate how confusion fares in comparison to other proposed solutions.

Chapter 2

Materialist Russellian Monism and Two Combination Problems

Intuitively there is a substantial difference between physicalism and idealism, despite the historical difficulty of formulating a definition of what it is for something to be physical or material.[1] Russellian monism has seen a resurgence in popularity recently as a theory of consciousness. One of the attractive features of Russellian monism is that it attempts to utilize the strengths of physicalism, dualism, and idealism while avoiding their traditional problems. The result has blurred what was once clear distinctions. Despite avoiding traditional problems for other theories of consciousness, Russellian monism faces an equally difficult problem: the many forms of the combination problem.

I think maintaining the distinction between materialism and idealism is important and can do real theoretical work. In this chapter, I begin by giving a brief overview of Russellian monism. I then show how a popular analysis of the physical, developed by Jessica Wilson[2] and adopted by prominent defenders of Russellian monism, such as Philip Goff,[3] fails to distinguish between material and ideal worlds. I will suggest an addition to that analysis which maintains

[1] An earlier and very different version of this chapter appeared in my MA thesis. The original version of this chapter argued, in a different manner, for a similar response to the subject-summing problem for the materialist constitutive micropsychist. The current version of the chapter presents a similar argument, however, despite maintaining that the materialist constitutive micropsychist should still respond to the subject-summing problem in this way, ultimately argues that the view is inherently unstable due to there not being a satisfactory characterization of material nature.

[2] Jessica Wilson, "Supervenience-Based Formulations of Physicalism," *Noûs* 39 (2005).

[3] Philip Goff, *Consciousness and Fundamental Reality* (New York: Oxford University Press, 2017a).

the material-ideal distinction and show how materialistic forms of Russellian monism can use this distinction to avoid two prominent forms of the combination problem: the subject-summing problem and the subject irreducibility problem. I conclude by presenting some concerns for the prospects of a materialist Russellian monism.

2.1 What is Russellian monism?

Russellian monism is a family of theories about the mind that range from materialist to emergentist to panpsychist to idealist to neutral monist. These theories share some core commitments, but differ greatly in the details. Because responding to Goff's attack on constitutive Russellian monism is the target of this chapter, I use his formulation of the relevant notions here.[4]

Russellian monism has negative and positive theses. According to the negative thesis, *structuralism about physics*, the physical sciences give us no access to the non-structural, categorical nature of its objects of study.[5] Physics can only grant insight to the world's causal structure. It can discover much about what matter does, but nothing about what matter is intrinsically. For example, what is an electron? Put simply, physics describes electrons as negatively charged particles. What is negative charge? Put simply, physics describes negative charge as the tendency to attract positively charged things and repel negatively charged things. Physics describes what an electron does, the causal roles it plays, the kinds of interactions it has. What physics cannot do is describe what an electron is in itself.[6]

The Russellian monist argues there must be a categorical basis to the causal structure studied by physics. A full defense of this view, *categoricalism*, is beyond the scope of this chapter, but the basic motivation is something like the following.[7] Consider what would have to be the case if categoricalism were false. The world would have only relations without relata. Why

[4]For alternative formulations of the view, see Alter & Nagasawa (2012), Alter & Pereboom (2023; Fall 2023), Chalmers (2015), and Pereboom (2011, chs. 5-6; 2015; 2019). For its historical antecedent from which it derives its name, see Russell (1927), though the view arguably originates with the work of G.W. Leibniz (Alter and Pereboom, Fall 2023).

[5]Goff, *Fundamental Reality*, p. 143.

[6]Alter and Pereboom (2023) argue that the Russellian monist need not accept this structuralist view of physics.

[7]Categoricalism is not a thesis unique to the Russellian monist. Arguments for categoricalism can be found in Armstrong (1968), Chalmers (1996; 2015), Goff (2017a), and Robinson (1982).

might this be unlikely? Consider if there were only brotherhood, but no brothers to stand in the brotherhood relation.[8] It is sufficient for my purposes to capture the intuitive strangeness of denying categoricalism. This notion of "categorical" is sometimes a slippery one, but it will be sufficient to think of it as a nature that is not able to be exhaustively characterized in terms of extrinsic, structural, dispositional, or role-playing properties.

The positive thesis of Russellian monism suggests an answer to what the categorical or deep nature of matter could be. Specifically, the deep nature of matter explains consciousness or phenomenal properties.[9] The knowledge argument, the conceivability argument, and the explanatory gap all seem to indicate that the physical truths do not explain the truths about consciousness.[10] According to the Russellian monist, this is because the sciences only tell us about the causal structure of the world, not about the categorical basis for that structure. If we had access to that deep nature, then this problem would dissolve.

How the deep nature of matter explains phenomenal truths depends on the answer to three questions and here we see where the theories diverge. First, what is the deep nature of matter? There are two main answers. *Panprotopsychist Russellian monists* hold that the deep material facts are protophenomenal properties. Protophenomenal properties can be thought of as the "building blocks" of phenomenal properties such that someone who understood their nature could explain *a priori* the phenomenal truths.[11] Though more needs to be said to distinguish this view from physicalism.[12] In contrast, *panpsychist Russellian monists* hold that deep material facts are phenomenal facts. On this view, the deep material nature of fundamental entities are conscious—there is something-it-is-like to be them—and explain *o-consciousness*, the ordinary conscious states we attribute to *o-subjects*, conscious subjects like you or I that we are pre-theoretically committed to.[13] While panpsychists hold that fundamental entities are conscious, they need not hold that they are conscious in exactly the same way that o-subjects are

[8]For a defense of the view that there are relations without relata, structural realism, see James Ladyman and Don Ross (2007).

[9]Goff, *Fundamental Reality*, p. 143.

[10]See Jackson (1982) and Robinson (1982), Chalmers (2009), and Levine (1983) respectively.

[11]I will not be discussing emergentist varieties of Russellian monism in this chapter. Emergentism will be discussed in chapter 5.

[12]Defenders of this view include Stoljar (2001) and Pereboom (2011, Chs. 5-6; 2015). Pereboom considers the view to be a form of physicalism.

[13]Goff, *Fundamental Reality*, p. 144.

conscious. What-it-is-like to be a fundamental particle is likely much simpler in experience and content (if any) than what-it-is-like for us as humans. For the remainder of this chapter, the discussion will be largely centered on panpsychist formulations of Russellian monism.[14]

Second, we must ask: how do our conscious states depend on that deep nature? Because the focus of this chapter is defending a form of *constitutive Russellian monism*, I will set aside emergentist varieties.[15] According to the constitutive Russellian monist, the deep material facts ground the o-phenomenal facts. This is the "ontological free lunch" notion of grounding, where the existence of the deep material facts is sufficient for the existence of o-phenomenal facts.[16]

Third and finally, the question arises what the fundamental things in the world are.[17] One might answer that human subjects are irreducible and fundamental or one might deny that there is a fundamental level at all. However, my discussion below will primarily deal with philosophers who take the fundamental entities to be micro-physical entities.[18]

Constitutive micropsychism is the view that microsubjects, the kind of subjects fundamental particles might be, ground o-subjects. This version of Russellian monism has many advantages. O-subjects will inherit their causal efficacy from the microsubjects that ground them, avoiding worries of causal exclusion which bother non-reductive physicalists and other anti-physicalist positions.[19] Further, it allows for realism about phenomenal consciousness without relying on emergentism or independent mental substances.

[14]Goff (2017a) argues for panpsychist over panprotopsychist Russellian monism on grounds of parsimony, to which I am sympathetic. According to the argument, the simpler view is that the intrinsic nature of all matter is a property that we know exists and are familiar with, i.e., phenomenal properties, rather than that only some matter (such as the matter found in brains) is intrinsically phenomenal or that phenomenality arises from some unknown protophenomenal property. Taylor (2019) argues against the view that panpsychism is more simple.

[15]Defenders of emergentist Russellian monism include Brüntrup (2017), Goff (2015), Mørch (2014), and Seager (2010; 2017).

[16]See Schaffer (2009).

[17]This question presupposes there are such fundamental things. Cf. Nagasawa (2012) as well as Nagasawa & Wager (2017). These papers articulate a form of Russellian monism free from this assumption.

[18]Some Russellian monists also accept priority monism and consider the fundamental thing to be the cosmos as a whole, leading to a view known as cosmopsychism. See Goff (2017a) and Nagasawa & Wager (2017).

[19]See Jaegwon Kim (1993) and Karen Bennett (2003; 2008).

2.2 Russellian monism, materialism, and idealism

Depending on the way the theorist describes the relationship between (proto)phenomenal properties, the intrinsic (categorical) nature of things, and the extrinsic (relational, role-playing, or structural) nature of things, the resulting formulation of Russellian monism can turn out to be a version of materialism, idealism, dualism, emergentism, or neutral monism. Indeed, examples of each have been articulated and defended.

Of course, because "material," "ideal," and others are all terms of art, it matters how one defines those terms as well. One distinction that will be important is between materialist-leaning and idealist-leaning versions of Russellian monism. Defining a sharp difference is difficult, as Russellian monism blurs traditional boundaries in the metaphysics of mind. But, a rough and ready distinction will serve my purposes.

Defining "physicalism" and "physical" is a difficult task. An initial inclination is to attempt to define what is physical in terms of physics, but then one must navigate the horns of Hempel's Dilemma. If physicalism is characterized in terms of our current physics, it will surely be false as our current physics is incomplete and, plausibly, incorrect in some way (the first horn of the dilemma). If it is characterized in terms of some future, completed physics, then physicalism could be compatible with entities that are, intuitively, unacceptable to the "spirit" of physicalism (the second horn).[20] I will follow Wilson in attempting to avoid the second horn of the dilemma by requiring that physical entities not be fundamentally mental.[21] As such, "an entity is physical just in case it is (approximately accurately) treated by current or future (at the end of inquiry, ideal) physics, and is not fundamentally mental."[22] However, I differ from Wilson in taking this to be merely a necessary condition on the physical and not sufficient.

A world in which the fundamental entities studied by physics also involve consciousness might seem to be something near enough to physicalism, even though this violates the spirit of the view. Goff agrees and calls such entities that are (approximately accurately) treated

[20] See Hempel (1980) for the original formulation of the dilemma. Wilson (2005) covers alternative versions of the second horn of the dilemma (such as physicalism being vacuous or indeterminant) and argues that the version of the second horn presented above is the only one worthy of concern.

[21] Goff (2017a) has a similar requirement, but also requires the physical to have no fundamental teleology.

[22] Wilson, "Supervenience-Based Formulations of Physicalism," p. 428.

by current or future physics whether or not they violate the constraint against fundamental mentality, "material."[23] Thus, all physical entities will be material entities, but not all material entities are physical entities. If one adopts this view of materiality, then panpsychist Russellian monism, as presented above, is the view that the world is composed of material entities, as the intrinsic nature of fundamental entities studied by physics involves mentality.[24]

As with the account of physical entities, I take this to be merely a necessary condition for materiality, but not sufficient. Recall that according to the Russellian analysis, physics yields only an account of the extrinsic properties of its objects of study, empirical inquiry gives us no insight into the deep nature of entities. As such, this account of materiality is unable to distinguish between material and ideal worlds. For example, consider a Leibnizian world. Phenomena in that world will be approximately accurately treated by the physical sciences, Leibniz himself thought so, but it strikes me as odd to say that Leibnizian phenomena are "material" in a sense materialists would accept.

Goff attempts to avoid these concerns by stipulating that matter or material entities must also be mind-independent "in the sense of not being dependent on being perceived" and noting that the panpsychist world still meets this criterion.[25] However, this is not sufficient to rule out all forms of idealism. This rules out the Berkeleyan idealist, but not the Leibnizian idealist.

I think it is important to maintain the intuitive distinction between idealism and materialism, but the account of matter presented above does not allow for this. What needs to be added to this account to make it possible to distinguish material from ideal entities? I suggest that the materialistic Russellian monist posits that the deep nature of matter either (i) involves mentality, but is *not exhaustively characterize*d by it as these mental properties are identical to some archetypical material property or (ii) the deep nature of matter involves some archetypal material property and does not involve mentality. Call materialism of the (i) variety "materialistic panpsychist Russellian monism" or "materialistic panpsychism" and of the (ii) variety "materialistic panprotopsychist Russellian monism" or "materialistic panprotopsychism." The notion of "archetypally material" is doing a lot of work here and it is difficult to suggest candi-

[23] Goff, *Fundamental Reality*, p. 41.

[24] Later, in section 2.5, I will argue that materialist panpsychist Russellian monists are committed to the relation being that of identity with deeply material properties.

[25] Goff, *Fundamental Reality*, p. 41n24.

date properties. Derk Pereboom initially suggests two possibilities, Lockean solidity and Aristotelian prime matter, later also suggesting the possibility of non-Humean causal powers.[26] In contrast, I propose that the idealist Russellian monist posits that the deep nature of the entities studied by physics is *exhaustively characterized* by mentality, that is, there is nothing about that nature that is archetypically material and cannot be completely characterized by mental or conscious predicates. For example, the deep nature of some entity might be exhausted by some phenomenal feel and that phenomenal property lacks any nature recognizably or paradigmatically material. Again, this account relies on these terms having some intuitive pull.[27]

Differentiating between the material and ideal is not a mere verbal dispute, it allows us to do real theoretical work and make progress on philosophical problems. In the remainder of this chapter, I will show how the materialistic Russellian monist might use this distinction to make progress on the combination problem.

2.3 Full revelation to combination problems

Many anti-physicalist positions take that we stand in a special epistemological relationship to our conscious states as a starting point for their philosophical theorizing. Historically, one could point to Descartes not being able to doubt that he is a thinking thing.[28] Modern defenders of Russellian monism have a similar starting point. Goff is particularly explicit in this commitment, making such a relationship his epistemological bedrock—the super justification thesis. Imagine a case where you stub your toe on a coffee table, then immediately introspect, forming a phenomenal concept whose content is based entirely on your attending to your conscious state, i.e., a *direct phenomenal concept*.[29] Introspecting, you find that your conscious state has a qualitative feel to it or a what-it-is-like component—specifically, painfulness. Now, further imagine you are asked in that moment, "are you sure that you are feeling pain right now?"

An appropriate response to this inquiry, I think, would be incredulity. Of course you are

[26]Pereboom's original two suggestions are found in his 2011 (85). The non-Humean suggestion can be found in Pereboom (2016; 2019).

[27]Alter & Nagasawa (2012) make a similar distinction.

[28]CSM II, p. 28.

[29]Goff, *Fundamental Reality*, p. 107.

feeling pain! Why else would you be hopping around, cursing, and holding your foot in that way? This is not to suppose that your pain is the behavior, but that it is the best explanation for the behavior. That you are feeling pain at that moment and that you know it through introspection does not seem to be the kind of thing you could be wrong about. If you are inclined to accept this way of thinking, then you may be inclined to accept super-justification.[30]

According to the *super-justification* thesis, we are entitled to be confident in our introspective judgments to roughly the same degree as we are in basic mathematical judgments, like $2+2=4$.[31] Note that the thesis only applies to introspective judgments, i.e., judgments formed by reflecting on your experience. Super-justification makes no claims about perceptual judgments. Consider again the case of stubbing your toe. You are super-justified in believing that you are experiencing pain, the view says nothing about your belief that there is a coffee table out there in the world.

Goff takes super-justification to be his epistemological bedrock,a datum requiring explanation rather than a conclusion requiring an argument, a point where I am in agreement. Where I disagree, however, is what Goff takes to be the explanation, *full revelation*:

> In having a direct phenomenal concept of token conscious state C, C is directly presented to the concept user, in such a way that (i) the complete nature of the type to which C belongs is apparent to the concept user, and (ii) the concept user knows with rational certainty (or something close to it) that C exists.[32]

In other words, full revelation entails that direct phenomenal concepts are *transparent*, revealing the full nature of the states they denote. I understand full revelation to explain super-justification in this way: by supposing that our conscious states are how introspection presents them, we get a good explanation for the bedrock assumption of super-justification.[33] So, it is a kind of inference to the best explanation.

Alternatively, there is *partial revelation*:

[30]Not everyone takes it to be the case that our direct phenomenal concepts are veridical in their representation of the phenomenal natures of our conscious states. For denials of this view, see Dennett (2017), Frankish (2017), and Pereboom (2011). Note that the illusionist will not deny that we can be reasonably confident that we are in pain in such a case, but "pain" refers to something like the normal cause of the mental state and not the way the mental state is qualitatively represented in introspection.

[31]Goff, *Fundamental Reality*, p. 111.

[32]*Ibid.*, p. 120.

[33]*Ibid.*, p. 112.

> In having a direct phenomenal concept of conscious state C, some aspect A of C is directly presented to the concept user, in such a way that (i) the complete nature of the type to which A belongs is apparent to the concept user, and (ii) the concept user knows with rational certainty (or something close to it) that A exists.[34]

In contrast the transparency of full revelation, partial revelation entails direct phenomenal concepts are *translucent*, only revealing part of the essential nature of the conscious states they denote.[35]

It is not clear to me that full revelation is a better explanation of super-justification than partial revelation. Super-justification posits that we are entitled to have a high degree of confidence in the content of our introspective judgments. However, our introspective judgments rarely have a form such as "has phenomenal feel p and that is all." Given that our introspective judgments are usually not of a form that specifies the totality of the nature of our conscious states, full and partial revelation seem to serve equally well in explaining super-justification. That is, while introspection might reveal the full phenomenal nature of a conscious state, it is not obvious that it does reveal the full essential nature of that state without the further argument that what our introspective capacities reliably reveal is the whole essential nature.[36] Nor is it clear that our being entitled to having a high degree of confidence in those judgments would entitle us to have confidence in the further judgment that the nature of that state is exhausted by what is revealed to us in introspection. Thus, it seems that full revelation claims too much as an explanation of super-justification without a further argument for the view.

Because super justification only claims that we are highly justified in our introspective judgments, it is an open possibility whether what is revealed is the full or partial essential nature of the state. Super justification only requires that our judgments about what is revealed in introspection, whatever that may be. are highly justified.

A defender of full revelation might respond that considerations of parsimony make it a better explanation. Of the two candidate explanations for super-justification, full revelation posits merely the existence of what is accessible through introspection. In contrast, partial revelation commits us to the essence of the mental state being more than what is revealed in

[34]Goff, *Fundamental Reality*, p. 120.

[35]For the remainder of this disussion, I will occasionally omit reference to the "type" conditions in referring to full and partial revelation for simplicity.

[36]I argue in section 2.5 that some materialistic Russellian monists are committed to this not being the case.

introspection. So, strictly in terms of what full revelation and partial revelation commit us to, full revelation is the simpler and preferable explanation. In rebuttal, one reason to think partial revelation is to be preferred is that it explains super-justification while claiming less than full revelation does. As an epistemically weaker position that explains super-justification just as well, partial revelation, is all we are entitled to accept as an explanation.

A consequence of full revelation is that it leads to the subject-summing and subject irreducibility problems. These arguments pose serious problems for constitutive micropsychism, the "most important" and theoretically attractive form of Russellian monism.[37] According to the constitutive micropsychist, microsubjects intelligibly produce, or ground, o-subjects. That is, if some agent had a transparent understanding of the nature of microsubjects, she could *a priori* come to understand the nature of o-subjects. Further, because it is a grounding relation and grounding relations involve necessity—grounds necessitate the grounded—it seems inconceivable that there be the microsubjects and not the o-subject. However, such a situation appears conceivable, as it seems that a group of subjects never entails any distinct, additional subject. Consider three conscious subjects: Jose, Amy, and Paul. Arrange them however you would like, surely it is conceivable that there is no subject other than Jose, Amy, and Paul. This seems true for any group of subjects we can imagine; whether Jose, Amy, and Paul are humans, dogs, or fundamental particles. Call this the *conceivable isolation of subjects* (CIS): for any group of subjects, S1, S2, ... Sn, it is conceivable that there is no subject S* such that S* is not identical with any S1, S2, ... Sn.[38]

According to the defender of full revelation, CIS is true because subjecthood is a determinable of which any conscious state is a determinant.[39] Because our direct phenomenal concepts yield a transparent understanding of the conscious states they denote, we must also have a transparent understanding of subjecthood. Therefore, if we cannot conceive how subjects could combine, it is because there is nothing in the nature of subjects that would permit their combination. Notice that full revelation is the crucial step in this argument, as it is the step that permits

[37] David J. Chalmers, "Panpsychism and Panprotopsychism" in *Consciousness in the Physical World*, eds. Torin Alter and Yujin Nagasawa (New York: Oxford University Press, 2015), p. 255.

[38] Goff, *Fundamental Reality*, p. 174.

[39] *Ibid.*, pp. 174–5.

the move from "subjects combining is not conceivable" to "subjects cannot combine".[40] This is the subject-summing problem for constitutive micropsychism.

In response to the subject-summing problem, the constitutive micropsychist can appeal to our lack of access to the nature of microphenomenal properties and microsubjects to attempt to bring the argument to a standstill. We have introspective access to macrophenomenal properties in introspection, but not to the microphenomenal states which constitute them. Though, the defender of the subject-summing problem will defend CIS as seemingly generalizable to any possible subject, it is possible that microsubjects have combinatorial properties that o-subjects like us lack.

More problematic for the constitutive micropsychist is the subject irreducibility problem. While the subject-summing problem arises from the difficulty of imagining how microsubjects could "build up" into o-subjects, the subject irreducibility problem starts with o-subjects and notes the difficulty of imagining how those subjects could be reduced or "broken down" into microsubjects. Specifically, subject irreducibility is the thesis that we are unable to give an analysis of any conscious subject S using facts that do not involve S, i.e., a deflationary analysis.[41]

The constitutive micropsychist is committed to there being a deflationary analysis of o-subjects because they hold that o-subjects are grounded in microsubjects, none of which are identical to the o-subject. Goff presents a schema for a micropsychist analysis of subjecthood:

> For it to be the case that there is a conscious subject X is for it to be the case that there are micro-subjects S_1, S_2 ... S_n, none of which is identical with X, standing in n-place relation R.[42]

The question naturally becomes, what relation satisfies R? If one is committed to full revelation, then one is committed to the nature of subjecthood being transparent and, therefore, *a priori*

[40] In order to defeat the view that o-subjects are grounded in microsubjects, proponents of the subject-summing argument need only the weaker claim that "there is at least one world where there are microsubjects and no o-subject". However, what is usually defended is the stronger thesis that subjects, by their nature, are not the kinds of things that could combine.

[41] Goff, *Fundamental Reality*, p. 209.

[42] *Ibid.*, pp. 211–2.

The way this account is presented, Goff seems to be offering this as the micropsychist's analysis of subjecthood full stop. But, it is evident that this cannot be the analysis of subjecthood for fundamental micro-subjects, because they would be identical to themselves unless, as he notes, the micropsychists is to be committed to an infinite regress

accessible. If the nature of o-subjecthood is *a priori* accessible, then the relation which satisfies the above schema should be *a priori* accessible.[43]

The constitutive micropsychist's inability to suggest an R leads to the subject irreducibility problem. First, there is the argument for the principle of subject irreducibility itself. Here is Goff's presentation of the argument,

1. If the analysis of subjecthood is *a priori*, then it is deflationary only if analytic functionalism is true.

2. Analytic functionalism is false.

3. The analysis of subjecthood is *a priori* (implied by Phenomenal Transparency).

C. The analysis of subjecthood is not deflationary (i.e., Subject Irreducibility).[44]

Analytic functionalism is the view that mental concepts are functional concepts. The constitutive micropsychist is commited to this view being false in virtue of her Russellian commitments. Recall that the Russellian monist finds a place for consciousness as the categorical basis for material role-playing properties. If mental concepts were merely functional concepts, then there would be no need for this Russellian move. But, because mental concepts have intrinsic, qualitative aspects that go beyond their functional properties, analytic functionalism must be false. The notion of subjecthood we get *a priori* is something like the Cartesian notion of the mind being simple or the Leibnizian view that minds (monads) are windowless. Having no parts or obvious methods of interacting, it is unclear how subjects would combine. Once subject irreducibility is established, the subject irreducibility argument against constitutive micropsychism is as follows:

4. That Jane exists as a conscious subject is grounded by analysis in the micro-level facts only if the micro-level facts logically entail what is essentially required for Jane to exist as a conscious subject. (This premise follows from the definition of grounding by analysis.)

5. (Subject Irreducibility) What is essentially required for there to be an x such that x is a conscious individual cannot be analyzed into facts not involving x.

[43] Goff, *Fundamental Reality*, p. 210.
[44] *Ibid.*, p. 216

C1. That Jane exists as a conscious subject can be grounded by analysis in the micro-level facts only if the micro-level facts logically entail the existence of Jane.

6. Jane is a macro-level entity, and hence her non-existence is logically consistent with the complete micro-level facts.

C2. The fact that Jane exists is not grounded by analysis in the micro-level facts.[45]

Notice, again, that the crucial premise, premise 5 (Subject Irreducibility), relies on full revelation for its support, because the argument for 5 relies on the truth of phenomenal transparency.

The partial revelation thesis is necessary to defend constitutive micropsychism in the face of these problems. If full and partial revelation are equally good candidate explanations of super-justification but full revelation rules out an important view for consideration, then that is a reason to prefer the partial revelation thesis.

2.4 The strategy

Before I present my argument against the full revelation thesis and response to the combination problems, I wanted to make a brief comment about how my strategy will differ from other approaches. Many approaches to the combination problem attempt to answer the question of how subjects combine or are reduced by giving a positive characterization of the features of subjecthood which would admit combining or of the process which would allow for such combinations or reductions. In this way, they attempt to dissolve some of the mystery behind how it happens.[46]

While dispelling the mystery with a positive account might be an ideal to strive for, I take a less ambitious path. Both the materialistic account given in this chapter and the idealist account given in a later chapter does not attempt to demystify how subjects might combine or be reduced. Instead, I attempt to explain why we find the combination process mysterious and show how this leaves subject combination and decomposition an open possibility. Explanations

[45]This argument is taken from Goff (2017a, 217).

[46]For defenses of this kind, see, for example, Goff (2017b) and Seager (2017)

of this kind are, I think, important for defenders of Russellian monism to pursue as it is possible that discovering or understanding the features which admit combination or reduction is out of our epistemic reach.[47] It is plausible, I think, that we might be in an epistemic place where this is the best we can do.

In this way, I am influenced by a dialectical move used by the British emergentists and, later, contemporary emergentists about the mind. In their layered view of the world, lower layers were related to higher, emergent layers by brute laws. There was no explanation, it was thought, for why a chemical reaction took place beyond invoking a brute law of nature that must be accepted with natural piety. The British emergentists did not seek to give an explanation of the mystery of chemical reactions. Rather, they attempted to explain why it could not be explained. Today, some emergentists about the mind use a similar strategy; they do not explain how the mind depends on the brain, but they explain why they cannot explain it.[48] I will make a similar move with regard to the mystery of how subjects combine and are reduced. I attempt to explain why it is a mystery—because full revelation is false and our direct phenomenal concepts are translucent and not transparent.

2.5 Translucency and materialist Russellian monism

In this section, I provide two objections to the above combination problems on behalf of the materialist panpsychist Russellian monists. The first objection, initially appears promising. Despite being ultimately unsuccessful, I argue, it does hint at the successful strategy used in the second objection. I will address materialist panprotopsychism in a later section.

First objection: a direct phenomenal concept of some conscious state, C, does not reveal that C is grounded in, realized by, or emergent from some state of the nervous system (which is part of the complete nature of C).[49] Therefore, the full revelation thesis is false because the direct phenomenal concept of C does not reveal the full nature of the conscious state.

[47] More on this in section 2.5.

[48] For a history and critical commentary of the British emergentists, see McLaughlin (2008).

[49] The constitutive micropsychist is committed to o-conscious states being grounded by analysis (Goff 2017a) or realized (Chalmers 2015). The cosmopsychist is committed to o-subjects being grounding by subsumption (Goff 2017a). The emergentist Russellian monist is committed to their being emergent (Brüntrup 2017). Therefore, this list is exhaustive of the possibilities.

Why does the above objection fail? Recall that full revelation requires direct phenomenal concepts reveal the complete nature of the type to which the introspected state belongs. Assume C is some pain state and is grounded in, realized by, or emergent from c-fibers firing in humans. But, pain, the type to which C belongs, is not essentially c-fibers firing. Possibly, some alien species experiences of pain arise from, say, z-fibers firing. Thus, the multiple realizability of conscious states means full revelation does not entail that direct phenomenal concepts reveal the realizing structure (e.g., c-fibers firing) because C being grounded in, realized by, or emergent from *that* realizer is not part of the essential nature of C's type. Therefore, given multiple realizability, this objection fails.

However, the materialistic panpsychist is not committed to an unqualified view of multiple realizability. Instead, she is committed to the qualified view that conscious states are only realizable by material entities, with "entities" being inclusive of the options without taking a stand on whether these are properties, substances, or states. I call this position "*material multiple realizability*." The materialistic panpsychist Russellian monist is committed to this position because she identifies the deep nature of matter with phenomenal properties that ground, realize, or otherwise intelligibly cause o-subjects. If the deep nature of matter is phenomenal properties, then there could not be conscious states realized by non-material on the assumption of the necessity of identity. As such, the materialistic panpsychist ought to deny what I call "*non-material multiple realizability*," that conscious states can be realized by non-material entities, like Cartesian souls.[50] The materialist Russellian monist, on my view, can allow for the possibility of a conscious Lieutenant Commander Data, but not for Casper the Cartesian Ghost.[51]

Why should the materialist panpsychist accept the identity claim? That is, why should she accept that the deep nature of matter is identical to phenomenal properties and not some weaker claim, such as there is a contingent relationship between these phenomenal properties and the scientific role-playing properties they are the categorical basis of. Not only does such a view leave an inexplicable connection at the very foundations of our world — why is phenomenal property p the categorical base or mass and not spin or some other fundamental role-playing

[50]This does not hold for materialistic panprotopsychists. That view will be considered in section 2.6.

[51]Of course, Casper is not truly a Cartesian spirit, as he is depicted as being spatially located and extended.

property — but, as Robert Howell argues, the view can no longer avoid worries about mental causation. Specifically, "even if phenomenal properties cause things on the Russellian Monism picture, they do not cause things in virtue of their phenomenal nature."[52]

If phenomenal properties are identical to the deep nature of matter and these theorists wish to maintain a robustly materialist form of Russellian monism (see section 2.3), then two important things follow. First, by the necessity of identity, phenomenal properties will necessarily be identical to deeply material properties — it would not be possible to have one without the other. This necessary identity is what supports material multiple realizability. Second, full revelation is false. Introspection reveals the phenomenal, what-it-is-like aspect and the propositional content of our conscious states. However, introspection does not reveal that the phenomenal properties introspected upon are also, necessarily, deep material properties or archetypally material properties. This aspect of a phenomenal property's essential nature is not revealed through introspection and it is essential to the type of state that it is (by material multiple realizability), therefore full revelation is false.

Given the commitments of materialist panpsychists, they should deny full revelation and accept partial revelation. Without full revelation, the crucial premises of the subject-summing and subject irreducibility problems are unsupported. Introspection does not reveal the full essential nature of our mental states, so from "one cannot conceive of subjects combining or decomposing" it does not follow that "subjects cannot combine or decompose." Aspects of the nature of mental states hidden from introspection may play a role in the combination or decomposition of subjects. As such, constitutive materialist panpsychism remains a strong contender for a theory of consciousness in the face of the combination problem. However, I think there is a deeper theoretical problem with materialistic Russellian monism that I hinted at earlier.

What qualifies these phenomenal properties as material? This is a difficult question for the materialist Russellian monist to answer. The properties we think of as archetypally physical or material, those described by physics, are merely structural or nomic properties in the Rus-

[52]Robert Howell, "The Russellian Monist's Problems with Mental Causation," *Philosophical Quarterly* 65(2015): p. 28.

sellian analysis.[53] If all that is meant by "material" is that these properties are approximately accurately characterized by physics, then it is unclear how materialist Russellian monism is to differ from idealist Russellian monism. After all, the idealist will not deny the existence of this structure, but will deny it is material. To maintain the distinction, we need a more substantial notion of "material." At this point, early in the Russellian monist's project, we might not be able to say much about this material nature.[54] The Russellian view of matter seems to imply that empirical investigation is unable to unveil these deeper features. Further, as argued above, they are not revealed through introspection. But, a commitment to phenomenal properties being identical to deep material properties (and the nature of these properties being material in some sense worthy of the word) serves, I think, as a way of distinguishing idealist forms of Russellian monism from formulations of the view that lean more materialistic.[55]

The advantage of holding materialist Russellian monism, one might think, is that it is more commonsensical, insofar as it maintains the intuition that there is a material, non-ideal world. However, idealist critics will say, not entirely unfairly, that common sense was abandoned when we fell down the rabbit hole and landed in the Wonderland of panpsychism. So, why cling to it now? In response, I hold that materialist Russellian monism, in addition to maintaining the commonsense intuition of a material world, is one of two forms of Russellian monism that can reject full revelation and thus preserve constitutive micropsychism in the face of the combination problem. Because constitutive micropsychism is seen as the "most important"[56] form of Russellian monism, this is no mean accomplishment. It is a form of Russellian monism that avoids the problems of mental causation for anti-materialist and non-reductive materialist views as well as the conceivability arguments against materialism.

Neutral monist forms of Russellian monism may be able to utilize this response. This will depend on how the "neutral" deep nature of things is to be characterized. Some neutral monists take "neutral" to mean the fundamental nature of the world is *neither* mental *nor* physical, but gives rise to mental and physical things.[57] The problem is that this characterization is

[53] Cf. Alter and Pereboom (2023).

[54] Pereboom (2011) theorizes two possible candidates which will be discussed in the next section.

[55] More on this in section 2.6.

[56] Chalmers, "Panpsychism and Panprotopsychism," p. 255.

[57] See Stubenberg (2018) for a survey of neutral monist views.

completely negative—it is neither mental nor physical. There is no positive characterization of the neutral nature and some have questioned the coherence of such a view.[58] While the lack of a positive characterization for a third kind of fundamental, non-dependent, concrete property does not conclusively show that such a property is impossible, I think it gives us some reason to think that the mental-physical distinction might be exhaustive of the possibilities—at least until we have a coherent, positively characterized third option. In other words, I take the position of Antoine Arnauld when he said, "I am only acquainted with two sorts of substances—bodies and minds. And it is up to those who would maintain that there are others to show them to us".[59] Thus, while the neutral monist may be able to utilize a similar response to what I presented above, their notion of "neutral" appears to me to be deeply problematic.

Idealist Russellian monism, where phenomenal properties exhaust the deep nature of role-playing properties in physics, cannot adopt the above way of rejecting full revelation because it rests on phenomenal properties having an opaque material nature. As such, she must either accept full revelation or press her support for partial revelation in some other way.[60]

To summarize the objection, if conscious states are multiply realizable and materialistic panpsychist Russellian monism is true, then the deep nature of matter is identical to phenomenal properties. If the deep nature of matter is identical to phenomenal properties, then material multiple realizability is true.[61] If material multiple realizability is true, then it is part of the essential nature of conscious states that they are material. If it is part of the essential nature of conscious states that they are material and full revelation is true, then a direct phenomenal concept would reveal that the conscious state is material. However, our direct phenomenal concepts do not reveal that our conscious states are material. Thus, either conscious states are not essentially material or full revelation is false. The materialistic panpsychist Russellian monist, I have argued, is committed to conscious states being essentially material. Therefore, if materialistic panpsychist Russellian monism is true, then full revelation is false.

[58]Galen Strawson, "Realistic Monism: Why Physicalism Entails Panpsychism" in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, ed. Anthony Freeman (Exeter: Imprint Academic, 2006), p. 23.

[59]LA Letter 2, p. 219.

[60]See the final three chapters for a defense of this sort.

[61]If you deny multiple realizability, then the first objection considered at the beginning of this section succeeds and full revelation is false.

2.6 Worries about materialistic Russellian monism

In section 2.2 I made a promissory note to elaborate on the difficulties of fleshing out the material aspect of materialistic Russellian monism in a "archetypally material" way. While maintaining a non-idealist realism about materiality might seem to make Russellian monism less counter-intuitive than idealistic forms, in this section I will argue the costs of such a project are too high. This stems from the fact that Russellian monist framework seems to leave little room for a notion of material beyond the structure and dynamics studied by physics. A purely structural account of materiality will is insufficient to distinguish material entities from ideal ones. However, if the materialistic Russellian monist attempts to characterize materiality beyond mere structure and dynamics, then, I argue, she diminishes the simplicity of the view.

If the material is merely the structural and relational properties of the world studied by physics, then there will be no distinction between idealism and materialism, these views would be identical. This is probably not a state of affairs either party would be happy with. If the materialist Russellian monist wishes to differentiate herself from the idealist, then they must look to the categorical properties which underlie the structure and dynamics studied by physics for their material natures. It seems there are three options: i) the categorical nature at lower-levels is exhausted by some archetypally material nature but mental properties arise at higher levels (*materialistic panprotopsychism*), ii) there is a mix of co-occurring but distinct categorical natures with some being archetypally material and some being mental (*dualistic pan[proto]psychism*), iii) the categorical nature is some archetypally material property that is identical to some mental property *(materialistic panpsychism*).

Each of these options will be examined in turn, but first, we need a candidate for this archetypally material nature. It is unclear what epistemic access we have to these categorical natures aside from introspection and, if what I argued above is correct, introspection does not give us access to any deep material nature or, at least, not representing it as being material. Therefore, it should be unsurprising there is a paucity of suggested materialistic natures in the current literature. Derk Pereboom does offer two candidates: Aristotelian prime matter and

Lockean absolute solidity, so these are what we will work with.[62]

Materialistic panprotopsychism is the view that there are only archetypally material categorical natures at lower levels, but composites of those material natures constitute mental properties at higher levels. This view gives up almost all the advantages of panpsychist Russellian monism over materialism. The two primary challenges facing materialism are the hard problem, why there is something-it-is-like accompanying processes in the brain, and the explanatory gap, how to explain phenomenal consciousness in material terms. One of the advantages of Russellian monism over materialism is having categorical natures as an additional theoretical resource to explain consciousness. The panpsychist Russellian monist side-steps the hard problem and the explanatory gap by identifying fundamental categorical properties with phenomenal properties—the consciousness is baked right in. However, if Russellian monists reject the panpsychist route, they face the major challenge of attempting to explain consciousness in terms of non-consciousness. Why do these composites of non-mental things give rise to mental things? This difficulty is compounded by limiting the intrinsic nature of things to something archetypically material, like absolute solidity. If Russellian monism is to avoid the problems of materialism, the base needs more than material properties.

Dualistic pan(proto)psychism could be of two varieties, dualistic panpsychism and emergent panprotopsychism. First, there could be *dualistic panpsychism*, where distinct mental and material categorical natures co-occur at all levels. The question must then be asked, which nature is the categorical basis for the structure and dynamics studied by physics? If it is only the material nature, then the view becomes epiphenomenalist as the mental nature does not causally contribute. If it is only the mental nature, then what is the material nature adding? Alternatively, it could be that both are necessary to serve as the basis for some fundamental scientific role-playing property and that individually they serve merely as the basis for mental-material combinations. Further, the co-occurrence of these distinct properties seems to cry out for an explanation. These complications do not rule out the view, but, surely, we should prefer panpsychist Russellian monism on grounds of parsimony. Second, there could be *emergent panprotopsychism*. If at lower-levels there are only material natures and mental natures only

[62]Derk Pereboom, *Consciousness and the Prospects for Physicalism* (New York: Oxford University Press, 2011), p. 86.

co-occur with material natures at higher-levels, then in order to differentiate the view from materialistic panprotopsychism, the mental natures would need to be strongly emergent from the lower-level material natures. While there are defenders of strong emergence, I am of the opinion that ontological danglers, brute explanations of higher-level features of the world, are to be avoided.[63] Further, strongly emergent phenomenal properties would face the exclusion problem for dualist positions. Dualistic panprotopsychism faces many of the difficulties of dualistic panpsychism as well.

Materialistic panpsychism is the view that the categorical basis for the role-playing properties studied by physics is a single entity that is both mental and material. So, the categorical basis of an electron might be Lockean-solidity-and-phenomenal-red (hyphens used to convey it is one continuous nature) while the categorical basis of an up quark might be Lockean-solidity-and-phenomenal-blue. The problem for this view is similar to the problems for dualistic panprotopsychism: what role is the material aspect (in the examples given, Lockean solidity) playing? The differences in the structural properties (playing the role of an electron versus playing the role of an up quark) supervenes on the mental nature (in the examples given, phenomenal redness or phenomenal blueness). It is not clear what contribution the material nature is making; it is almost epiphenomenal as it does not seem to be making a difference. Perhaps this issue could be avoided by having a plurality of material natures (Lockean solidity for the quarks, Aristotelian prime matter for the bosons, and so on), but it is not as if we have many candidates for such material natures and, as mentioned above, we have no clear epistemic access to them. Further, this proliferation of distinct material natures harms the simplicity of Russellian monism, an often claimed advantage of the view.

An alternative for the materialistic panpsychist would be to argue that, in the example above, there is no aspect of the categorical nature that is distinct from the mental aspect. In the deep nature of things, there is no difference between the material and mental natures. So, the deep nature of all matter would have to be P-and-M, where "P" is a placeholder for some material nature and "M" is a placeholder for some mental nature. This avoids any issues about difference-making, but the cure is certainly worse than the disease. This would make the palette problem,

[63] See Smart (1959) and, for more reasons to reject emergence, chapter 5.

the problem of accounting for the incredible diversity of phenomenal properties we experience in terms of the sparse palette of phenomenal properties at the fundamental level, much more difficult to solve as there would only be one phenomenal nature at the fundamental level, P-and-M. Further, the difficulty raised by Howard Robinson of how the mental or phenomenal natures of fundamental material role-playing properties relate to the material roles they are the fundamental natures of now seems completely intractable.[64] If electrons and up-quarks have the same deep nature, why do they behave differently? It seems no answer can be given.

If one is going to be a materialistic Russellian monist and accept constitutive panpsychism, then I think their best response to the subject-summing and subject decomposition problems is the one I outlined in section 2.5, deny the full revelation thesis because introspection does not reveal the essential material nature of our mental states. If full revelation is false, then critical premises about how subjects might combine or decompose are unsupported and materialistic constitutive micropsychism is still an open possibility. However, it is unclear what is a sufficient characterization of "material." There is a paucity of candidate material natures and,as I have argued, even if a plausible candidate were proposed, the resulting view seems to be epiphenomenalist.

[64]Howard Robinson. *From the Knowledge Argument to Mental Substance: Resurrecting the Mind* (Cambridge: Cambridge University Press, 2016), pp. 117–18.

Chapter 3

Illusionism as Skepticism

You live in a world of illusion
Where everything's peaches and cream.

"Jungle Love"
Steve Miller Band (1977)

"The possibility of first-person scepticism should raise an issue that might have been raised *a priori*: what entitles us to any *confidence* we might have in our belief in our own s-consciousness? ... with scpeticism about [causality, morality, knowledge, free will, and personal identity] widely taken very seriously. Why should consciousness be any different?"
- Georges Rey (2016, 203, emphasis in original)

Illusionism about phenomenal consciousness has garnered much attention largely due to defenses by Daniel C. Dennett, Keith Frankish, and Derk Pereboom.[1] Briefly, the illusionist holds that introspection represents conscious states as having phenomenal or qualitative properties they actually lack, i.e., illusory. In contrast, the realist holds that our conscious states instantiate some of the phenomenal or qualitative properties introspection represents them as having. In this paper, I will explicate the illusionist's position, their claimed advantages over realism, and a common argument for their view that I call "the argument from error." Then I will mount a defense of realism about phenomenal consciousness by showing that the argument from error is an instance of a common skeptical argument. If the argument from error is a

[1] See, especially, Dennett (1991; 2016; 2019), Frankish (2016, 2019), and Pereboom (2011).

skeptical argument, then the illusionist is committed to a broader form of skepticism. Further, if we are not gripped by other skeptical arguments, we should not be gripped by the argument from error.

3.1 What is illusionism?

Illusionism shares a starting point with many realist positions: phenomenal consciousness, the qualitative, what-it-is-like aspect of experiences, is inexplicable by our best physical theories.[2] From this starting point, Frankish describes three paths forward. The *radical realist* argues for the necessity of radical theoretical innovation to account for phenomenal properties as introspection represents them. The *conservative realist* is optimistic that phenomenal properties will one day be explained by the physical sciences. Finally, the *illusionist* agrees with the radical realist that conservative advances in the physical sciences will not be able to account for phenomenal properties. However, rather than advocating for radical theoretical change, the illusionist argues that the inability of phenomenal properties to be explained by the physical sciences is evidence that phenomenal properties as introspection represents them are illusory.[3]

The illusion of phenomenal consciousness is the result of how introspection represents our conscious states, what Pereboom calls the introspective mode of presentation (IMOP). The IMOP represents conscious states as having qualitative natures they lack, which results in our forming the inaccurate judgment that there is a what-it-is-like nature to the mental state.[4] The research program proposed by the illusionist is to discover empirically how these judgments are formed, as opposed to armchair speculations about how phenomenal properties might find a place in the physical world. However, in the place of phenomenal consciousness, the illusionist accepts a state of consciousness characterized by *quasi-phenomenal properties*, call this "*quasi-phenomenal consciousness*."[5] A quasi-phenomenal property is a physical property or a complex of physical properties that introspection misrepresents as having a qualitative

[2]Throughout this chapter, I will use "phenomenal", "qualitative", and "what-it-is-like" interchangeably.

[3]This typology comes from Frankish (2016, 13–14).

[4]Derk Pereboom, *Consciousness and the Prospects for Physicalism* (New York: Oxford University Press, 2011), pp. 12–13.

[5]Keith Frankish, "Illusionism as a Theory of Consciousness," *Journal of Consciousness Studies* 23 (2016): p. 21.

character.[6] This is what we refer to when we make claims about phenomenal experiences: functional states realized by the brain. Phenomenal terms like "pain" have their reference fixed, on this view, not by the qualitative feeling of pain (as there is no such thing), but by the quasi-phenomenal property which typically produces the inaccurate representation. So, certain physical states are consistently misrepresented as pain while distinct physical states are consistently misrepresented as hunger and the physical states misrepresented as pain are not misrepresented as hunger and vice versa (usually, see Pereboom's examples covered in section 3.3 of this chapter).[7]

If illusionism is correct, then there is no hard problem of consciousness; there is no problem of explaining why physical processes in the brain are accompanied by subjective, what-it-is-like experiences.[8] Instead, there is what Frankish calls "the illusion problem" or the problem of explaining why the illusion of phenomenal consciousness seems so real and vivid.[9] This can be thought of as a sub-problem of the meta-problem of consciousness: the problem of explaining why there seems to be a hard problem of consciousness.[10]

It is important to understand how strong the illusionist's claim is, but also to understand that it is not as strong as it is sometimes caricatured as being, namely, as holding to eliminativism about consciousness *tout court*. To ensure there is no over-stating the view, I will let Frankish speak for himself:

> Illusionism makes a very strong claim: it claims that phenomenal consciousness is illusory; experiences do not really have qualitative, 'what-it's-like' properties, whether physical or non-physical. This should be distinguished from a weaker view according to which some of the supposed features of phenomenal consciousness are illusory. ... Strong illusionism, by contrast, denies that the properties to which introspection is sensitive are qualitative: it is an illusion to think there are phenomenal properties at all.[11]

Thus, if we are to say that the illusionist is an eliminativist, we must be clear that they are only an eliminativist concerning phenomenal consciousness. We should not make the cari-

[6] Frankish, "Illusionism," p. 15.
[7] See also Pereboom (2011, 15 and 35).
[8] David J. Chalmers, "Facing Up to the Problem of Consciousness," *Journal of Consciousness Studies* 2 (1995).
[9] See Frankish (2016; 2019).
[10] David J. Chalmers, "The Meta-Problem of Consciousness," *Journal of Consciousness Studies* 25 (2018).
[11] Frankish, "Illusionism," p. 15.

catured claim that they are eliminativists about consciousness full stop or that they deny that consciousness exists. One of many ways of doing this is to use Ned Block's terminology, the illusionist denies phenomenal consciousness but accepts access consciousness.[12]

Illusionists will object here and say they are not eliminativists about phenomenal consciousness, but only about the qualitative aspects introspection represents phenomenally conscious states as having. Pereboom is explicit about this, drawing an analogy to primitivism about color:

> [the qualitative inaccuracy hypothesis] is no more eliminativist about phenomenal properties than most contemporary theories of secondary qualities are about color. In each case, primitivism about the properties at issue, for example, is denied, but denying primitivism does not amount to eliminativism about phenomenal properties, only about primitive versions of these properties.[13]

Pereboom makes a strong case for this, but I think there is an important difference between the primitive color case and the phenomenal property case. When we revised our conception of what colors are, we revised our conception of something, on most accounts, in the mind-independent real world. In the case of phenomenal properties, this is not the case and a common position in the philosophy of mind has been that the essential nature of these properties is how they feel in experience.

3.2 The argument from error

In this section, I will present and evaluate the relative merits of three versions of what I call the argument from error, a common illusionist argument. I then present a general form of the argument.

First, consider the argument from error as it is presented by Frankish:

> If a property resists explanation in physical terms or is detectable only from a certain perspective, then the simplest explanation is that it is illusory. ... Given the force of these considerations, if there is even a remote possibility that we are mistaken about the existence of phenomenal consciousness, then there is a strong

[12] Ned Block, "On a Confusion about a function of Consciousness," *Brain and Behavioral Sciences* 18 (1995).
[13] Pereboom, *Prospects,* p. 45.

> abductive inference to the conclusion that we are in fact mistaken about it. And there is reason to think that we could be mistaken about it. ... we have no introspective way of checking the accuracy of our introspective representations, and so cannot rule out the possibility that they are non-veridical.[14]

If it is possible to be mistaken about the existence of our phenomenally conscious states, then the best, simplest explanation is that those states are illusory. As phenomenal consciousness resists physical explanation and is only first-person detectable via introspection. It is always possible to be mistaken in this way because there is no way of checking the accuracy of how introspection presents conscious states to us. Therefore, the best explanation is that phenomenal consciousness is illusory.

Frankish, I think, gives the strongest version of the argument. The realist should not deny either that phenomenal consciousness resists physical explanation (that is why they are a radical realist, after all) nor should they deny that introspection is the only way to access phenomenal consciousness. On a critical note, Frankish has not shown that introspection could be mistaken or that there is a possibility of error by showing that there are no ways of checking whether the results of introspection are accurate. Rather, all Frankish has shown that if there were an error, we would not know. That we would not know if there were an error does not entail that error is possible.

Pereboom gives a similar presentation of the argument from error. Again, the basis of the argument is the possibility of error in introspection:

> introspective representations—those of inner sense—are caused by the mental states they represent and are wholly distinct from them. Introspective representations thus mediate the subject's awareness of those mental states, rendering this awareness in a sense indirect. As a consequence, the subject may represent a mental state as being a certain way, even though it is not really that way.[15]

Pereboom goes on to argue that because we have no independent way of checking the accuracy of introspection, we cannot know if the representations are accurate.[16] In contrast to Frankish, Pereboom argues in Kantian fashion that introspection might be unreliable because

[14] Frankish, "Illusionism," pp. 27–8.
[15] Pereboom, *Prospects*, p. 10.
[16] Pereboom, *Prospects*, p. 24.

it represents, but is distinct from, mental states introspected upon and representations can misrepresent in virtue of being distinct from the thing represented. Pereboom's representational view of introspection is to be contrasted with an acquaintance view where introspection directly reveals or acquaints us with our mental states. However, there need not be deep conflict. Pereboom and Frankish agree on the core issue, that the results of introspection are not able to be independently verified. The argument from error need not depend on a specific view about how introspection accesses conscious states, only that it is the way of accessing conscious states.[17] Further, Pereboom provides some possible cases of error in introspection, discussed below.

Finally, here is the argument as it appears in Dennett. In considering the realist's objection that

> We have "privileged access" to the causes or sources of our introspective convictions. No logical room for any tricksters intervening here! We couldn't be victimized by any illusions here! You might be a zombie, unwittingly taking yourself to have real consciousness with real qualia, but I know that I am not a zombie![18]

Dennett responds,

> No, you don't. The only support for that conviction is the vehemence of the conviction itself, as soon as you allow the theoretical possibility that there could be zombies, you have to give up your papal authority about your own nonzombiehood. I cannot prove this yet, but I can encourage would-be consciousness theorizers to recognize the chasm created by this move and recognize they can't have it both ways.[19]

Like Frankish and Pereboom, Dennett argues that introspection alone is insufficient to justify belief in phenomenally conscious states. He further argues that because phenomenally conscious states are not needed to explain behavior and introspection does not justify belief in them, they are unnecessary theoretical posits. However, a notable difference between Dennett's presentation of the argument from error in contrast to Frankish and Pereboom is that Dennett sees this problem as, perhaps only, arising for the realist when they accept the possibility of zombies, creatures who are narrowly physically identical to us but lacking phenomenal

[17] Pereboom (2011, 19–21) argues forcefully that the illusionist's worry about introspection applies just as well to non-causal accounts of introspection and phenomenal concepts.

[18] Daniel C. Dennett, *From Bacteria to Bach and Back: The Evolution of Minds* (New York: W.W. Norton & Co., 2017), p. 263, emphasis in original.

[19] *Ibid*., emphasis in original.

consciousness.[20] Where "narrowly physical properties" are the structural properties studied by physics, but leaving out any categorical basis for that structure. Dennett is not explicit ("I cannot prove this yet") about why the problem is dependent on accepting the possibility of zombies, but I suppose the reasoning is something like the following. If zombies are possible, then your evidence cannot differentiate between whether you are phenomenally conscious or a zombie. If zombies are narrow physical duplicates of us, then phenomenal consciousness does not explain our behavior. "Behavior" here is taken broadly to mean bodily movements, verbal reports, and beliefs. If phenomenal consciousness does not explain behavior and introspection is not sufficient to justify belief in it, then phenomenal consciousness is an unnecessary theoretical posit.

I think Dennett's predicating the argument from error on the possibility of zombies is unnecessary and makes the argument too strong. First, not all realists accept the possibility of zombies.[21] Second, implicit in Dennett's formulation of the argument is that only behavior needs to be accounted for, which will not be convincing to realists who hold that phenomenal experiences are also data that needs to be explained and not merely something posited to explain behavior. Dennett's argument would be stronger were he to focus on introspection itself as the only way of accessing phenomenal consciousness, like Frankish and Pereboom, rather than taking on the controversial assumptions about zombies and behavior. If, as I suggested above, Dennett is using the possibility of zombies to merely show that one's evidence cannot distinguish between the zombie world and the phenomenal realist world being actual, then Dennett is making the same point in an unnecessarily stronger way.

Abstracting from the three arguments presented above, we find the following argument, (most explicit in Frankish and Pereboom, and to a lesser extent in Dennett):

1. There is no way independent of introspection to know if introspection accurately represents conscious states.

[20] I specify that zombies are narrowly physically identical rather than simply physically identical in order not to beg the question against the Russellian monist. According to the Russellian monist, if a zombie was broadly physically identical to me, that is, it was not merely identical in its structural properties but also identical in terms of the categorical basis of that structure, then it would have the same conscious states that I have because consciousness depends on that categorical basis.

[21] For example, Balog (2011).

2. If there is no way independent of introspection to know that introspection accurately represents conscious states, then introspection is not sufficient evidence for phenomenal consciousness.

3. Introspection is not sufficient evidence for phenomenal consciousness.

With the additional premise that introspection is our only evidence for believing in phenomenal consciousness, the illusionist can show that if introspection is undermined, then all our evidence for phenomenal consciousness is undermined. By "undermined," I mean that the reasons given for realism about qualitative phenomenal properties are inadequate to justify the belief in them. Premise 1 should be uncontroversial. Our capacity for introspection is not like our capacity to do mathematics, for example. If I attempt to solve a mathematical problem, I can pass my work to another to check that my results are correct. This is not to say that our agreement on a solution to a mathematical problem shows that we know the metaphysical reality of mathematical entities, only that I can check my ability to do mathematics against an external agreed-upon standard by checking my work against another's without solely relying on my ability to do mathematics. In contrast, the only way to see if an introspective representation is accurate is to introspect again to check if the results agree. Premise 2 is supported if this kind of bootstrapping is not acceptable. It would be like testing the reliability of my ability to do mathematics by solving equations and asking myself if I got it right.[22] As such, it is possible that introspection may be inaccurate such that it distorts the nature of the conscious states it provides access to. The relevant notion of "possible" here is epistemic possibility. This is especially clear in Pereboom where he is explicit that the QIH is an open possibility, i.e., "possible given what we human beings now rationally believe."[23]

If the argument from error is sound, then the illusionist has shown that the realist is not entitled to the confidence they have in accepting phenomenal consciousness. Our epistemic situation with respect to phenomenal consciousness is not what we ordinarily take it to be. Further, the intuitions underlying all antiphysicalist arguments are undermined because these intuitions rely on the accuracy of introspection.

[22] Compare with the case of bootstrapping the reliability of color vision in Cohen (2010: 142–43).

[23] Pereboom, *Prospects*, p. 15. Note, also, that the illusionist still has not shown that error is possible, only that we would not know if there were errors.

3.3 The skeptic and the illusionist

Might illusionism be a form of skepticism? It is clear that illusionism is not a form of global skepticism nor external world skepticism. The illusionist does not deny the possibility of knowledge. The illusionist should not be seen as a skeptic about other minds or even about minds generally as they do not deny that there are minds. When I suggest the illusionist is a skeptic, I do not mean that they are committed to a form of skepticism traditionally discussed.

I also do not suggest they are a skeptic in the trivial sense that they deny knowledge of phenomenal consciousness because they deny that there is phenomenal consciousness. Using "skeptic" in this way would render any position which denies the existence of some supposed phenomenon "skeptical". That might be (trivially) true in some sense, but I am suggesting a deeper connection between illusionism and traditional forms of skepticism.

Illusionism is a form of skepticism, I argue, in this way: the argument from error against phenomenal consciousness makes the same move as some traditional skeptical arguments. A traditional form of skeptical argument notes that there is no way to confirm the accuracy or reliability of some capacity without using that capacity. Because of this, if that capacity were in error, we would not know (because we would have to utilize the very capacity in question to check). If there is no way to know if the capacity in question is in error, then that capacity does not justify any beliefs. Simply put, the schema for the most genereal skeptical argument from error is:

4. There is no way independent of capacity c to know if c accurately represents phenomenon p.

5. If there is no way independent of capacity c to know if c accurately represents phenomenon p, then c is not sufficient evidence for p.

6. Capacity c is not sufficient evidence for p.

The above schema should look familiar. Replace "c" with "introspection" and "p" with "phenomenal consciousness" and we have the general form of the illusionist's argument. Here is how the external world skeptic would advance an instance of the argument from error:

7. There is no way independent of perception to know if perception accurately represents the external world.

8. If there is no way independent of perception to know if perception accurately represents the external world, then perceptual experiences are not sufficient evidence for the common-sense external world.

9. Perceptual experiences are not sufficient evidence for the common-sense external world.

According to the skeptic, just as our only evidence for introspection's veracity is other introspective judgments, the only evidence we have for the veracity of perception is perception itself. But, bootstrapping perceptual justification will not be acceptable to the skeptic.

4 - 6 and 7 - 9 are somewhat old-fashioned ways of putting the skeptical argument. Today, it is common to present skeptical arguments using a form of the closure principle. Epistemic closure states that if you know p and you know that p entails q, then you are in a position to know that q. Put in terms of closure, a skeptical argument is as follows:

10. You do not have sufficient evidence that you are not a brain in a vat.

11. If you do not have sufficient evidence that you are not a brain in a vat, then you do not have sufficient evidence that you have hands.

12. You do not have sufficient evidence that you have hands.

Premise 11 here is the closure premise; having hands entails that you are not a handless brain in a vat. While 10 - 12 is the contemporary way of presenting the skeptical argument, the reasons the skeptic will give for 10 are usually arguments like 4 - 6 or 7 - 9.

Generally, the skeptic raises the possibility of error with respect to some capacity (e.g., perception or introspection), then notes that if our beliefs formed by that capacity could be wrong, then we do not know the content of those beliefs. Consider the skeptic about the external world. She argues that you cannot know that you have hands because you could be wrong. It could be a dream or a hallucination or, dare I say, an illusion. We are inclined to cite our evidence for our having hands in response. We can see the hands, we can feel the hands, we can hear the hands when we clap, and so on. The skeptic will not be moved by these reasons

because you are presupposing the accuracy of the very faculties that she is calling into question. In the same way, the illusionist will not be moved by the realist's appeals to their qualitative experiences which they access through introspection—the faculty called into question. As Frankish explicitly expresses the doubt in his formulation of the argument from error, "we have no introspective way of checking the accuracy of our introspective representations, and so cannot rule out the possibility that they are non-veridical."[24] As does Pereboom,

> In the case of external sensory representation, we have readily available ways of checking the entity represented that are independent of the representation under scrutiny, while for introspection such ways of checking are at best very limited. ... decisive ways of checking introspected phenomenal properties are not available to us.[25]

An immediate difference between the skeptic and the illusionist should be apparent. The skeptic is usually taken to be making an epistemological claim. They do not attempt to reach the conclusion "you are a brain in a vat," merely that "you do not know you are not a brain in a vat." In contrast, the illusionist makes a metaphysical or ontological claim: "phenomenal consciousness does not exist." However, this difference will not be problematic for my argument. The epistemological claim is weaker than the metaphysical claim. The illusionist reaches their metaphysical claim as an abductive inference from an epistemological claim. The best explanation for beliefs about phenomenal consciousness, given that introspection is an unreliable guide and the success of other scientific explanations, is that it is illusory. So, the realist should respond to the illusionist as they respond to the skeptic; objecting at the epistemological level of the argument from error, which supports the illusionist's metaphysical conclusion.

If I am correct in arguing that the argument from error is a form of skeptical argument, then the illusionist who wields the argument is committed to any form of skepticism motivated by the same move from a capacity's possibility of error to denying that that capacity can sufficiently justify beliefs or knowledge. As mentioned above, this includes perception. Denying that perception produces justified beliefs or knowledge about the external world leads to global skepticism. Thus, the illusionist who wields the argument from error is committed to global

[24]Frankish, "Illusionism," p. 28.
[25]Pereboom, *Prospects*, p. 23.

skepticism. As I have shown that this is a common form of argument amongst prominent illusionists, this is not a minor problem for their view.

Can the illusionist can avoid this commitment? Illusionists could provide a reason why introspection is epistemically different from other capacities targeted by the skeptic, such as perception, such that the move from the possibility of error to denying the capacity produces sufficiently justified beliefs is acceptable for introspection but not, for example, perception. If they could do this, then it would show that we should be skeptics about phenomenal properties but not about the external world. This route is equivalent to refuting the external world skeptic and would prove quite difficult as it would require justifying perception without relying on perception.

Illusionists could also point to other factors that might differentiate the illusionist from other forms of skepticism. One way of doing this might be to object that the skeptic merely points to the possibility of error, while some illusionists try to give actual cases of error. Consider this case from Pereboom, describing a dentist giving his daughter a Novocain shot:

> the dentist hid the needle from her and told her that he would be dropping bits of cold water into her mouth. She didn't flinch. When I asked her afterward whether the experience was unpleasant, she said that she didn't like the drops of water much, but they didn't hurt.[26]

Taking this example at face value, it appears to show a case where introspection rendered an inaccurate judgment. Of course, for the illusionist, it is not inaccurate in the sense that the daughter was really feeling phenomenal pain and not phenomenal cold, but that she represented the quasi-phenomenal property, the cause, of pain as coldness. However, the illusionist is not unique among skeptics in presenting actual cases of error. Consider Descartes' presentation of the skeptic:

> But from time to time I have found that the senses deceive, and it is prudent never to trust completely those who have deceived us even once.
>
> Yet although the senses occasionally deceive us with respect to objects which are very small or in the distance…[27]

[26] Pereboom, *Prospects*, p. 24.
[27] CSM II, p. 12.

Here, Descartes motivates skepticism by pointing to actual cases where our senses have deceived us about some features of the external world. Consider, too, this passage:

> How often, asleep at night, am I convinced of just such familiar events – that I am here in my dressing-gown, sitting by the fire – when in fact I am lying undressed in bed! ... I see plainly that there are never any sure signs by means of which being awake can be distinguished from being asleep.[28]

Descartes plays the skeptic here by pointing to actual cases where we have formed inaccurate judgments about whether we are awake as a way of establishing that we cannot reliably judge whether we are awake or asleep. So, the illusionist cannot avoid the charge of being a skeptic on the grounds that they do not merely point to the possibility of error, as the skeptic supports their case with actual cases of error as well.

Similarly, the illusionist might attempt to differentiate themselves from the skeptic by arguing that illusionists have independent reasons to reject phenomenal consciousness in addition to the possibility of error in introspection, specifically the inability of physical sciences to explain phenomenal consciousness despite success in other fields. In contrast, the external world skeptic or the skeptic about other minds lacks any such independent reason in addition to the possibility of error.[29] On this view, the general form of the illusionist's argument from error should be:

13. There is no way independent of capacity c to know if c accurately represents phenomena p.

14. There are independent reasons to reject p.

15. If there is no way independent of capacity c to know if c accurately represents phenomena p and there are independent reasons to reject c, then c is not sufficient evidence for p.

16. Capacity c is not sufficient evidence for p.

Where the parallel argument for the illusionist is:

[28] CSM II, p. 13.

[29] I am grateful to Michael V. Griffin for this objection.

17. There is no way independent of introspection to know if introspection accurately represents phenomenally conscious states.

18. There are independent reasons to reject phenomenal consciousness.

19. If there is no way independent of introspection to know that introspection accurately represents conscious states and there are independent reasons to reject phenomenal consciousness, then introspection is not sufficient evidence for phenomenal consciousness.

20. Introspection is not sufficient evidence for phenomenal consciousness.

And the parallel argument, for example, for the external world skeptic is:

21. There is no way independent of perception to know if perception accurately represents the common-sense external world.

22. There are independent reasons to reject the common-sense external world.

23. If there is no way independent of perception to know if perception accurately represents the common-sense external world and there is independent reason to reject the common-sense external world, then perception is not sufficient evidence for the common-sense external world.

24. Perception is not sufficient evidence for the common-sense external world.

According to this objection, 18 is justified while 22 is not.

Amending the argument from error in this way successfully distinguishes the illusionist from the skeptic. However, there are two important things to note. First, this form of the argument is distinct from the arguments presented by prominent illusionists (see section 2). As such, it does not refute my claim that a prominent form of illusionist argument is also a form of skeptical argument. Second, this new form of argument is dialectically much weaker than my generalized reconstruction of Dennett, Frankish, and Pereboom's arguments. If, as premise 17 claims, there is no way to confirm the accuracy of introspection besides introspection, then it should come as no surprise that the physical sciences have been unable to make progress on explaining phenomenal consciousness as the physical sciences are based on perception and, *ex*

hypothesi, perception cannot confirm the accuracy of introspection. So, if the motivation for 18 is based on perception, then 17 and 18 cannot both be true. That is, it cannot be the case that only introspection can confirm the accuracy of introspection (17) and that perception gives us reason to think that introspection is inaccurate (18). Realists about phenomenal consciousness might accept 17, but should reject 18 for this reason.

3.4 Skeptical vs. debunking arguments

David J. Chalmers evaluates another form of argument common in the illusionist literature: genealogical debunking arguments.[30] Debunking arguments attempt to undermine our justification in a belief by showing that the genealogy of that belief — that is, how that belief was formed — is not truth-conductive, does not reliably produce true beliefs, or only accidentally results in true beliefs. For example, consider Sharon Street's Darwinian dilemma against non-natural moral realism.[31] Street argues that belief in non-natural moral facts was affected by evolutionary pressures. Because evolutionary pressures are select for fitness and not the truth, our beliefs about non-natural morality were not formed in a truth-responsive way. Therefore, we lack justification for belief in non-natural moral facts. Alvin Plantinga gives a similar genealogical debunking argument against naturalism.[32]

The similarities between skeptical arguments and debunking arguments should be clear. Both attempt to undermine our justification for a belief by attacking the source of the belief. However, the manner in which the source is attacked differs. Genealogical debunking arguments, as I understand them, attempt to establish a positive thesis: that the way used to form the belief is not truth-conductive, does not reliably produce true beliefs, or only accidentally results in true beliefs. In contrast, skeptical arguments attempt to establish a negative thesis: that we do not know whether some way of forming beliefs is justified or that we cannot differentiate cases where the way used to form the belief fails to track the truth. The genealogical debunker assumes we are justified with respect to and know some propositions and uses that

[30] See Chalmers (2018; 2020a).

[31] Sharon Street, "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies* 127(2006).

[32] Alvin Plantinga, *Where the Conflict Really Lies: Science, Religion, and Naturalism* (New York: Oxford University Press, 2011).

to undermine some other belief while the skeptic questions the possibility of knowledge or justification.

It is also important to note that arguments Chalmers notes as genealogical debunkers are not the argument from error which I identified as skeptical. Here is an example of an illusionist debunking argument given by Chalmers,

25. There is a correct explanation of our beliefs about consciousness that is independent of consciousness.

26. If there is a correct explanation of our beliefs about consciousness that is independent of consciousness, those beliefs are not justified.

27. Our beliefs about consciousness are not justified.[33]

By "consciousness" in the argument above, Chalmers means "phenomenal consciousness." Premise 25 would be supported by some illusionist thesis, such as Pereboom's qualitative inaccuracy hypothesis, where introspection misrepresents the nature of conscious states as having qualitative natures they lack. An illusionist view like this gives a physical explanation for why we form beliefs like "I'm experiencing pain." This is the positive thesis the illusionist must defend. Premise 26 is the debunking premise; if our beliefs about phenomenal consciousness were formed via a process not involving phenomenal consciousness, then beliefs about phenomenal consciousness are not justified. This debunking argument is distinct from the argument from error, discussed above, which depends only on the negative thesis that we cannot know if introspection is a reliable or accurate way of forming beliefs about conscious states.

3.5 Two kinds of anti-skeptical projects

I take it that I have established that illusionism, as supported by the argument from error, is a form of skepticism. How is the realist to proceed? How can the realist learn from how we respond to the skeptic? Some philosophers respond to the skeptic simply by ignoring them, by

[33]David J. Chalmers, "The Meta-Problem of Consciousness," *Journal of Consciousness Studies* 25 (2018): p. 45.

not taking them seriously. After all, how many of us are kept awake at night by the possibility that we are brains in vats? If you're inclined to ignore the external world skeptic, you may be inclined to ignore the illusionist and need not read further. I think that skepticism deserves a response, whether of the external world or the illusionist variety.

James Pryor identifies two kinds of responses to the skeptic.[34] First, there is the *ambitious anti-skeptical project*, where one attempts to refute the skeptic using only premises the skeptic would accept. Descartes' *Meditations* is an ambitious project; he attempts to reason from indubitable premises to the existence of the external world.[35] Ambitious anti-skeptical projects are incredibly difficult and, as Pryor puts it, "The prospects for this ambitious anti-skeptical project seem somewhat dim."[36] This is because the ambitious project's aims are Herculean: to defend the reliability of a capacity (perception, introspection, etc.) without appealing to that capacity. This is not to say that ambitious projects are not attempted, but they are not considered to be very successful.

I agree with Pryor that a successful ambitious project is unlikely, but is it impossible? I think so. Skepticism, in the forms discussed in this chapter, operates by questioning whether we can form sufficiently justified beliefs through faculties that cannot be independently verified, e.g., perception and introspection. Because these faculties cannot be independently verified, we cannot defend their use to the skeptic without presupposing the reliability of the very faculty we are attempting to defend. However, the skeptic will not accept arguments of this sort. Therefore, there can be no successful ambitious anti-skeptical projects, i.e., responses to the skeptic which rely on only premises the skeptic will accept.[37] So, if we are to save our knowledge from the skeptic and the ambitious project is not viable, we must allow for modest projects.

The second kind of anti-skeptical project, the *modest anti-skeptical project*, aims not to satisfy the skeptic, but to satisfy yourself that you still know some or most of the things you took yourself to know before the skeptic cast doubt on them. Important here is that when undertaking

[34] James Pryor, "The Skeptic and the Dogmatist," *Noûs* 24 (2000).

[35] This is not to say that Descartes' project in the *Meditations* was not revisionary. Only that Descartes attempts to prove that we know many of the things we took ourselves to know, despite the skeptic's claims.

[36] Pryor, "Skeptic and the Dogmatist," p. 517.

[37] Of course, this argument assumes there will be no successful neo-Cartesian argument that non-circularly gets from some *a priori* truth to the external world and our senses working when we do not let our will overreach our intellect.

the modest anti-skeptical project, you do not need to restrict yourself to only premises that the skeptic would accept.[38] For the remainder of this chapter, I will consider several modest anti-skeptical projects and evaluate how they might be appropriated towards responding to the illusionist's argument from error.

3.5.1 The Moorean facts response

One kind of premise that modest projects can avail themselves of is what David Lewis called "*Moorean facts*," that is "things that we know better than we know the premises of any philosophical argument to the contrary."[39] Alternatively, we can think of Moorean facts in terms of justification rather than knowledge. They are the things we are justified to believe in better than we are justified in accepting any philosophical hypothesis to the contrary. Sure, these facts beg the question against the skeptic, but we have already given up the ambitious project, we do not have to play by the skeptic's rules or use only premises they would accept.

Moorean facts take their name from a classic example of a modest anti-skeptical project: G.E. Moore's (1939) "Proof of an External World".

28. Here is a hand.

29. Hands are external objects.

30. External objects exist.

The Moorean fact against the external world skeptic here is premise 28, that we have hands. A Moorean fact against the phenomenal consciousness skeptic (i.e., the illusionist) is that we sometimes feel pains, pleasures, and the whole rainbow of phenomenal experiences. So, I suggest a parallel argument.

31. Here is a pain.

32. Pains are phenomenally conscious states.

[38] Pryor, "Skeptic and the Dogmatist," p. 517.
[39] David Lewis, "Elusive Knowledge," *Australasian Journal of Philosophy* 74 (1996): p. 549.

33. Phenomenally conscious states exist.[40]

Important to note is that by "pain" I mean the ordinary, phenomenal sense of the word, i.e., pain as qualitatively accurately represented by introspection. The experience we have when we introspect after stubbing a toe or stepping on a Lego. I do not mean the quasi-phenomenal state "pain" picks out for the illusionist. The illusionist may raise the argument from error to premise 31, but this should have the same force as the skeptic raising the possibility of error when you look at your hand.

Moorean facts do not entail that the capacities which give us access to them are infallible. Perception sometimes gets it wrong, Pereboom's cases seem to show that introspection does as well. But, just as the possibility of error and actual cases of error do not leave us unjustified in believing we have hands, we are not unjustified in accepting we have phenomenal consciousness.

The main drawback to this Moorean anti-skeptical response is that even though it need not convince the skeptic, it seems rather table-pounding. While it does seem to me that I am more justified in believing I have hands than I am a brain in a vat or that I experience phenomenal pain than I do not, it would be nice to have a response to the skeptic that has a little more philosophical meat on its bones than just asserting what's meant to be established. So, the next modest anti-skeptical project I consider is, hopefully, more philosophically robust. But, it might be the case that as evolved creatures with imperfect cognitive capacities, Moorean projects are the best response we have to the skeptic, whether or not it is a satisfying one.

3.5.2 The dogmatic response

Pryor defends dogmatism, a modest anti-skeptical project, against the external world skeptic. According to the dogmatist, if it perceptually seems to subject s that p, then s has immediate, *prima facie* justification in believing that p.[41] Justification to believe that p is immediate when it is not based on her justification for other propositions.[42] Justification to believe that p is

[40] Chalmers (2018; 2020a) proposes a similar Moorean argument against the illusionist, though he does not draw the parallels between illusionist and skeptical arguments.

[41] Pryor, "Skeptic and the Dogmatist," p. 519.

[42] *Ibid*., p. 532.

prima facie when it is defeasible by additional evidence, that is, being *prima facie* justified in believing that p is consistent with the belief being false.[43] This justification only applies to the perceptually basic proposition representing the objects of our experience. Pryor leaves open which propositions are perceptually basic as a matter to be settled by cognitive scientists.[44] However, such propositions will likely represent features such as color, shape, distance, and motion.[45]

Adapting the dogmatic response to the illusionist is fairly straightforward. It seeming to one that she is in a phenomenally conscious state p gives us immediate, *prima facie* justification in believing p. Our beliefs in phenomenally conscious states seem to be good candidates for being immediately justified, as we do not normally take them to be the kind of things whose justification depends on our justification for believing something else — an error the illusionist attempts to force us to make by questioning our justification for accepting the reliability of introspection. Further, given that dogmatism was originally a position for perceptual justification, the dogmatic response to Pereboom should be particularly effective given his representational view of introspection.[46] If perception is a kind of representation and introspection is a kind of representation, then dogmatism might be a unified response to both skeptics about the external world (who undermine perceptual justification) and illusionists.

A difficulty might arise for panpsychist realists with the requirement that the justification is with respect to basic features. According to the illusionist view, "introspection delivers a partial, distorted view of our experiences, misrepresenting complex physical features as simple phenomenal ones."[47] Most realists would agree that phenomenal properties are simple properties. However, the constitutive panpsychist holds that the macro-phenomenal properties of the experiences of creatures like people and other animals are grounded in the micro-phenomenal properties of more fundamental entities. As such, macro-phenomenal properties might not be basic, but misrepresented as basic by introspection. This is similar to a view held by G.W. Leibniz that much of our mental life is confused perceptions; "confused" because we can not

[43] Pryor, "Skeptic and the Dogmatist," p. 543.
[44] *Ibid.*, p. 539.
[45] Steven L. Reynolds, "Effective Sceptical Hypotheses," *Theoria* 79 (2012): p. 269.
[46] See section 3.3.
[47] Frankish, "Illusionism," p. 18.

pick out the individual constituents of a mental state and represents it as simple. Though, perhaps the problem is not so bad for the constitutive panpsychist. After all, they do not hold that introspection justifies the belief that the macro-phenomenal property, e.g., pain, is constituted by many micro-phenomenal properties. The justification given by introspection is meant only to cover that one is experiencing that macro-phenomenal property. So, perhaps it will be sufficient for the constitutive panpsychist dogmatist to hold that the justification is of features represented as basic, but not that they *are* basic.

3.6 Conclusion

How dialectically effective are these modest arguments? Against the illusionist who raises the argument, I doubt it will be very effective. Dennett, Frankish, and Pereboom are unlikely to slap their foreheads and say "of course!" But, these are not ambitious anti-skeptical projects. The aim was not to convince the illusionist. The aim was for the realist to convince or reassure themselves that the illusionist's skeptical doubts about phenomenal consciousness are misplaced. But, the realist must still give responses to the illusionist's non-skeptical arguments. These modest responses will only be dialectically acceptable against illusionists who use this skeptical form of argumentation. Other forms of illusionist non-skeptical arguments are available. But, prominent illusionists today defend this form of skeptical argument as exemplified by the argument from error.

However, it is important to note that the illusionist who rejects these modest projects does so at her peril. If we deny the acceptability of modest anti-skeptical projects, we should be more than illusionists. It would be necessary for us to become skeptics about other minds, the external world, and many other things. Thankfully, I am skeptical that this will be necessary.

Chapter 4

Was Leibniz an Illusionist about Phenomenal Qualities?

Walter Ott contrasts an important problem for early modern philosophers with a problem for today's philosophers of mind, Joseph Levine's explanatory gap.[1] Levine's explanatory gap concerns explaining the qualitative, subjective, or what-it's-like aspects of consciousness in terms of physical properties and processes. Similarly, the early moderns were concerned with explaining why the primary qualities of external bodies caused some secondary quality ideas in us and not others. For example, when I perceive the flowers of a rose bush, whether I perceive them as red and not as orange or blue is underdetermined by the primary qualities of the rose bush and their interaction with my body. Call this "the early modern gap."[2]

As a staunch adherent to the principle of sufficient reason (PSR), G.W. Leibniz found the positions of his contemporaries on this question unacceptable. However, I do not think it is fully appreciated how radical Leibniz's solution to the problem is. In this chapter, I will argue that Leibniz thought that secondary quality ideas were illusory, that the properties they represent obtain neither in the objects themselves nor in the mind as veridical phenomenal experiences. The chapter will proceed as follows. In section 1, I lay out the early modern explanatory gap.

[1]This chapter was presented to CEU's Little Metaphysics Club. I am grateful for comments from its members: Dávid Bartha, Sara Boljević, Michael V. Griffin, and Ruben Noorloos. I am also grateful for comments from Jeffrey Watson.

[2]Walter Ott, "Leibniz on Sensation and the Limits of Reason," *History of Philosophy Quarterly* 33 (2016).
Joseph Levine, "Materialism and Qualia: The Explanatory Gap," *Pacific Philosophical Quarterly* 64 (1983).

In section 2, I explicate Leibniz's confused ideas thesis and explain how that alone does not close the early modern gap. In section 3, I will present Leibniz's radical solution to the early modern gap found in his *New Essays on Human Understanding* (*New Essays*). In section 4, I argue that Leibniz's position on the early modern gap is similar to a contemporary position in the philosophy of mind on the explanatory gap, illusionism, prominently defended by Daniel C. Dennett, Keith Frankish, and Derk Pereboom.[3] I will conclude by fielding some objections to this interpretation of Leibniz.

Of course, there are difficulties with attributing a contemporary position to a historical philosopher. While it is important to understand the context of the debates historical philosophers were engaged in and how they differ from our own, I think it is a mistake to think that our debates today are so different that they would find our questions and problems unrecognizable or that they would have nothing to contribute. Further, illusionism or, at least, the eliminativist aspect of it, is a new name for an old position in philosophy.

4.1 The early modern gap and the Cartesian solution

Before Galileo and Descartes, the early modern gap did not exist. The prevailing Aristotelian view was (broadly) that we perceived objects in the world as they were.[4] What accounted for the redness in my perception of the rose was the redness of the rose itself. Of course, there were similar problems in antiquity about why things might appear differently to different observers, but (broadly) there was not a question about whether sensible qualities like redness were qualities of the objects themselves.

Note, I will use "sensible qualities" to refer strictly to the qualities represented by secondary quality ideas such as phenomenal redness, the feeling of warmth, and so on. This usage follows Leibniz's usage in various places. We will see that, on Leibniz's view, this usage of "sensible qualities" picks out any quality we receive a notion of through a single sense and that notion must be confused. Sensible qualities, for Leibniz, refer to what we would call today "phenom-

[3] See especially Dennett (1991; 2016; 2019), Frankish (2016, 2019), and Pereboom (2011).

[4] Cf. Brown and Normore (2019), who argue that the Cartesian view replaced an Aristotelian view that was at least as strange.

enal properties." I do not think this comparison is an anachronism or reading our contemporary philosophical interests into historical philosophers. The epistemically problematic properties Leibniz groups here are all "felt" properties. Using "sensible" in this restricted way might sound odd. Are shapes and sizes not sensible? Of course, they are, but Leibniz is drawing our attention here to the qualitative aspects of things; not shape and size but the feel of those surfaces. For this reason, I will use sensible and phenomenal properties interchangeably, as our contemporary term is clearer. The other obvious candidate for terminology is "secondary qualities," which I do occasionally use. However, this term also admits of ambiguity and requires diligence to distinguish between talk of secondary qualities and our ideas of them.

The early modern gap arises with the advent of modern science and philosophy, specifically with Galileo and Descartes taking phenomenal qualities, e.g., color and smell, out of bodies and restricting them to being experiences in the mind or soul. This view became so widespread during the period that in his *Treatise of Human Nature*, David Hume noted,

> the **fundamental principle** of [early modern] philosophy is the opinion concerning colors, sounds, tastes, smells, heat and cold; which it asserts to be nothing but impressions in the mind, derived from the operation of external objects, and without any resemblance to the qualities of the objects.[5]

With sensible qualities taken out of the world, the stage is set for the early modern gap. Here is how Leibniz presents the problem in a letter to Queen Sophie Charlotte of Prussia,

> We use the external senses as a blind man uses his stick… Senses allow us to know their particular objects, which are colors, sounds, odors, flavors, and the tactile qualities. But they do not allow us to know what these sensible qualities are, nor what they consist in, for example, whether red is the rotation of certain small globes which, it is claimed, make up light, whether heat is a vortex of very fine dust, whether sound is produced in air as circles are in water when a stone is tossed in, as some philosophers claim. We do not see these things, and **we cannot even understand why this rotation, these vortices, and these circles, if they are real, should bring about exactly the perceptions we have of red, heat, and noise.**[6]

So, the early modern gap is the lack of an explanation of why bodies and their motions cause ideas in us that bears no resemblance to the bodies themselves, such as small globes rotating at a certain speed causing ideas of red in observers.

[5] Hume, *Treatise*, I.iv.4, emphasis added.

[6] 1702 G VI 499/AG 186, emphasis added. I was directed to this passage by Ott's (2016) paper.

Leibniz ascribes a solution to the Cartesians where the only explanation to be had lies in God's free, but arbitrary choice. Of course, a staunch proponent of the PSR like Leibniz could never accept such a view, but Leibniz seems to think that if sensations were simple, e.g., phenomenal redness had no parts, aspects, or representational complexity, then what he considers the Cartesian position, that the relationship between states of the soul and states of the body in sensations are arbitrarily annexed by God, would be unavoidable.[7] This is due to Leibniz's commitment to the PSR and his understanding of the relationship between causes and effects. Effects, for Leibniz, must correspond to their causes, "[i]t is thoroughly reasonable that the effect should correspond to the cause; and how could one ever be sure that it does not".[8] If sensible qualities or secondary quality ideas were simple and effects must correspond to their causes, then we should expect the cause to be simple as well. Otherwise, there would not be a correspondence between the simple effect and its complex cause. But, the causes are not simple, they are, proximately, complex motions in the brain or, distally, complex motions on the surfaces of external objects.[9] If the nature of cause and effect is to correspond and simple secondary quality ideas do not correspond to their complex causes, then no explanation can be found in those objects. So, if secondary quality ideas are simple as they appear to be, then the only explanation for why a body causes some secondary quality idea in us and not another is God's arbitrary choice.

The question is, why does perceiving a body cause in the mind the color sensation it does as opposed to some other color sensation? Let's consider a specific case, why does perceiving a rose cause a sensation of phenomenal redness in the mind as opposed to a sensation of phenomenal blueness? The Cartesian answer, according to Leibniz, is that it is because God arbitrarily chose to actualize roses causing color sensations of red as opposed to blue. Leibniz's objection to this view, put simply, is two-fold. First, by the PSR, the choice cannot be arbitrary. Second, if color sensations (as well as other sensible qualities like heat and sounds) were simple, i.e., had no parts, aspects, or representational complexity, then the connection between

[7] 1704 A VI.6 131/NE II.viii.13. See also, Wilson (1999a, 326).

[8] 1704 A VI.6 131-2/NE II.viii.15.

[9] Speaking with metaphysical rigor, the only causes of our ideas are the previous states of our soul. But, it is easier to speak of the external things causing my ideas of them. Further, Leibniz is happy to talk this way on many occasions, including in the texts considered here. Indeed, he even goes so far as to say that there is a sense in which it's true that our ideas are caused by external things. (1686(?) A VI.4 1571–72/AG 59)

them and their causes in the world would have to be arbitrary due to the nature of causes. Take the case of perceiving a rose causing an experience of phenomenal redness in the soul. Light reflecting off the surface of the petals of the rose, being a body and having many parts with many motions, interacts with the body of the perceiver causing complex motions in the eye and brain. The cause, therefore, is complex. Because the effect must correspond to the cause and the cause is complex, the effect must be complex as well. So, we should expect the sensation of phenomenal redness to be complex as well. But, if phenomenal redness is simple, as it appears to be, then the effect does not correspond to the cause. If the effect does not correspond to the cause, then there can be no explanation other than God's arbitrary choice. But, by the PSR, an arbitrary choice of God cannot be the explanation. In fact, that the effect follows from the cause is stated as an equivalent to the PSR in Leibniz's "Primary Truths."[10] If there were not some correspondence between the cause and effect (if, say, a complex cause had a simple effect), then, for Leibniz, that cause would not serve as an explanation for that effect except by invoking an arbitrary choice of God. But, by Leibniz's understanding of the PSR, namely that effects follow from their causes, this cannot be the case. Therefore, phenomenal redness and the other secondary quality ideas are not simple as they appear. Denying that secondary quality ideas are as simple as they appear to be is the first step in Leibniz's radical solution.

4.2 Leibniz's confused ideas thesis

Leibniz's solution to the early modern explanatory gap lies in his theory of confused ideas or notions. Similar to many of his early modern contemporaries, Leibniz utilizes the terms "clear," "distinct," "idea," "adequate," and other related terms. Unlike many of his contemporaries, Leibniz defines these terms in his works, such as his "Meditations on Knowledge, Truth, and Ideas" ("MKTI"). In general, these terms are used in "MKTI" to denote various degrees of conceptual discrimination where the lowest level, obscurity, consists of being unable to discriminate a notion from anything else, and the highest level, intuitiveness, consists of immediately being able to discriminate all the notions involved in an idea at once. What constitutes an "idea," as we will see in the next section, is more complicated than we might

[10] 1686(?) A VI.4 1442/AG 31.

expect. For now, it is sufficient to note that all ideas are notions, but not all notions are ideas.

Confusion, for Leibniz, is contrasted with distinctness where both of these categories fall under clarity. There is a low bar for clarity, a notion "is *clear* when I have the means for recognizing the thing represented."[11] A notion is *distinct* when "connected with marks and tests sufficient to distinguish a thing from all other similar bodies...in a word, those that pertain to everything for which we have a *nominal definition* (which is nothing but an enumeration of sufficient marks)."[12] A notion is confused when it is not distinct, that is, "when I cannot enumerate one by one the marks sufficient for differentiating a thing from others, even though the thing does indeed have such marks and requisites into which its notion can be resolved."[13]

Leibniz gives the following example of a distinct idea with a nominal definition,

> But a *distinct notion* is like the notion an assayer has of gold... heaviness, color, solubility in *aqua fortis*, and others, which are among the marks of gold, such that gold may be distinct, yet *inadequate*.[14]

The notion of gold the assayer has is inadequate, as opposed to adequate, because all the marks involved in the definition are not themselves distinct. Specifically, the assayer's nominal definition of gold involves the color of gold, and notions about color are confused.

Notions of colors, like all notions gained through a single sense, are necessarily confused,

> [a]nd so we recognize colors, smells, tastes, and other particular objects of the senses clearly enough, and we distinguish them from one another, but only through the simple testimony of the senses, not by way of explicit marks.[15]

Leibniz is not explicit in MKTI that this confusion is due to these notions being through a single sense. But, just a few sentences later, he is explicit that notions gained through several senses are capable of being distinct,

> Notions common to several senses, like the notions of number, magnitude, and shape are usually of such a kind, as are those pertaining to many states of mind, such as hope or fear.[16]

[11] 1684 A VI.4 586/AG 24, emphasis in original.
[12] 1684 A VI.4 587/AG 24, emphasis in original.
[13] 1684 A VI.4 586/AG 24.
[14] 1684 A VI.4 586–87/AG 24, emphasis in original.
[15] 1684 A VI.4 586/AG 24.
[16] 1684 A VI.4 586–87/AG 24.

Note a contrast here. On the side of confusion, Leibniz groups colors, smells, and tastes, the troublesome sensory qualities. On the side of distinctness, Leibniz groups the notions of number, magnitude, and shape which are among the primary qualities. If notions gained through several senses are capable of being distinct but colors, smells, tastes, and other notions related to secondary quality ideas are necessarily confused, i.e., not recognized by way of explicit marks, then the notions involved in secondary quality ideas must be gained through a single sense. If the above contrast is not enough, Leibniz is later explicit about notions gained through a single sense being necessarily confused in a letter to Queen Sophie Charlotte,

> the notions *of the particular senses*, which are *clear but confused*, and the *notions of the common sense*, which are *clear and distinct*.[17]

But why should notions involved in secondary quality ideas be necessarily confused? My notion of blue is clear; I can recognize blue when it is presented to me. One might think the notion is also distinct and suggest that a candidate sufficient mark for differentiating blue from everything else is the phenomenal experience of blue. This appears to work as nothing that is not blue will be accompanied by the phenomenal experience of blue when perceiving it (under "normal" conditions).

A clue to why these notions of sensible qualities are necessarily confused is found, I think, again in a letter from Leibniz to Queen Sophie Charlotte,

> assayers have marks by which they distinguish gold from any other metal and even if a person had never seen gold, he can be taught these infallible marks for recognizing it, should he encounter it one day. But it is not the same with these sensible qualities. For example, one cannot give marks for recognizing blue, if one has not seen it. Hence blue is its own mark, and in order for someone to know what blue is, we must necessarily show it to him.[18]

Nominal definitions allow a person who has never experienced a thing to identify it. As in the example, if I had never seen nor held gold, the assayer could still teach me to identify it without ever requiring that I handle the metal.[19] But I could not give a test for blueness (or other colors)

[17] 1702 G VI 499/AG 187–88, emphasis in original.

[18] 1702 G VI 499/AG 187.

[19] This seems to indicate that the color of gold is not a necessary feature of the assayer's nominal definition. If it were, then she could not teach others who had not experienced gold to identify the metal. So, the color of gold is a contingent feature of the assayer's nominal definition, included on the basis of her actually having perceived gold. Or the student must have other experience with that color.

that could be used in the same way. That is, one could not teach someone to recognize blue without showing them blue. You can only learn to identify blue by experiencing blue. It is easy to see how this case might generalize to other sensible qualities, like smells and tastes. Hence, sensible qualities are necessarily confused because we are incapable of providing a nominal definition.

If secondary quality ideas are confused, they must be confusions of a complex of other notions. After all, what differentiates a confused notion from a distinct notion is not that the distinct notion has marks where the confused does not, after all, "the [confused notion] does indeed have such marks and requisites into which its notion can be resolved", but whether or not those marks are accessible to introspection.[20] Further, Leibniz explicitly states that he categorizes secondary quality ideas as composite notions and not as primitives,

> Thus we cannot explain what red is to a blind man, nor can we make such things clear to others except by leading them into the presence of the thing and making them see, smell, or taste the same thing we do, or, at very least, by reminding them of some past perception that is similar. This is so even though it is certain that the notions of these qualities are composite and can be resolved because, of course, they do have causes.[21]

Here, Leibniz indicates that the causes of a confused notion, like red, necessitate that the confused notion can be resolved into further notions. As established earlier, Leibniz thought that the causes of these notions were complex, so they must resolve into a complex of notions. But, notions of what? As the preceding passage indicates, they are the notions of the causes of the confused notion. To make this more apparent, consider that while the typology presented in Leibniz's "MKTI" appears to be concerned with conceptual capacities, specifically with our ability to discriminate the relations between various notions in our mind, Leibniz also speaks of confusion with regard to our perceptions. A favorite example of Leibniz's is a perception of the color green. Even if one knows that green is the composite of blue and yellow, it is not presented in perception as a complex of these two, but as a simple green experience.[22] Consider also Leibniz's example of the spinning cog,

[20] 1684 A VI.4 586/AG 24
[21] *Ibid.*
[22] 1702 G 499/AG 187; 1704 A VI.6 120/NE II.ii.1.

> when the swift rotation of a cog-wheel makes us perceive an artificial transparency, as I have noticed on visits to clock-makers, we are not able to discern the idea of the cause of this, i.e., the idea of the teeth of the wheel. The wheel's rotation makes the teeth disappear and an imaginary continuous transparent [ring] appear in their place; it is made up of successive appearances of the teeth and of gaps between them.[23]

In these examples of confused perceptions, the perceptions are confused because they represent a complex phenomenon as a simple, unified phenomenon. They are literally *con fused*, as Leibniz puts it in the New Essays, "a confusion — a running together".[24] The complex of blue and yellow is represented as simple greenness. The individual teeth and the spaces between them are represented as a homogenous blur. When a perception represents a multitude as a unity, it is confused.

It may not be immediately apparent how Leibniz's confused perceptions and confused notions relate to each other. Some philosophers, such as Margaret Wilson[25] and Nicholas Jolley,[26] argue that Leibniz has two distinct notions of confusion. Conceptual confusion, where one is unable to give a nominal definition of an idea or notion, and perceptual confusion, where one represents a multitude as a unity. However, I agree with Stephen Puryear who effectively argues that there are not two notions of confusion in Leibniz, conceptual and perceptual, but a single notion of confusion at work.[27] Briefly stated, Puryear argues that perceptual and conceptual confusion both consist in being unable to distinguish the implicit content, whether it be of a notion or a perception:

> Leibnizian perceptions are confused in the sense that they have implicit complexity or content. But this is precisely the sense in which ideas are said to be confused too. Leibniz holds that to call an idea clear but confused is to say that though the idea itself is accessible to the mind, its ingredients are, at least for the most part, not. In general terms, his view is that "confusion is when several things are present, but there is no way of distinguishing one from another" (MP 146). [Puryear 102-3]

[23] 1704 A VI.6 403/NE IV.vi.7.

[24] 1704 A VI.6 132/NE II.viii.15.

[25] Margaret Dauler Wilson, "Confused vs. Distinct Perception in Leibniz: Consciousness, Representation, and God's Mind," in her *Ideas and Mechanism: Essays on Early Modern Philosophy* (Princeton: Princeton University Press, 1999b).

[26] Nicholas Jolley, *Leibniz and Locke: A Study of the New Essays on Human Understanding* (New York: Oxford University Press, 1984).

[27] Stephen M. Puryear, "Was Leibniz Confused about Confusion?" *The Leibniz Review* 15 (2005).

In what follows, I will assume Puryear is correct and that Leibniz is working with a singular notion of confusion when he discusses ideas and perceptions.

Given that there is a single theory of confusion at work in Leibniz and that what is implicit in a confused perception are the causes of the perception, this lends credence to the view that what is implicit in a confused notion are the causes of the notion. So, the implicit notions in confused secondary quality ideas are the causes of those secondary quality ideas and the implicit complexity of a confused perception is the cause of that perception. An example, let "qualitative" or "phenomenal red[ness]" refer to the confused notion of red we receive through perception that is accessible to introspection. Let "red" refer to the real cause of phenomenal redness, what phenomenal redness is a confused notion of, or the implicit notions involved in our confused notion of phenomenal red. Our notion of a rose's qualitative redness is the result representing a complex of rotating small globes as a unity. This results in the representationally simple sensation. A distinct notion of red would be one where introspecting on the idea of qualitative redness could be resolved into ideas of rotating small globes. The idea of rotating small globes is present in our idea of qualitative redness as *petites perceptions* — unconscious perceptions in the mind that sometimes, through aggregation or intensity, become conscious perceptions. "Unconscious" here means that they are not accessible in introspection. These *petites perceptions* of rotating small globes aggregate and are represented by the mind as the conscious perception of qualitative redness.

Leibniz's understanding of the early modern explanatory gap can now be seen to have three components, i) the thing in the world that is perceived, ii) the petites perceptions of the thing, and iii) the confused conscious perception of the thing. There is no gap between the thing itself and the *petites perceptions*, as the *petites perceptions* represent the causes of the perception of the object. Further, the petites perceptions are the implicit content of the confused conscious perception.

4.3 Confusion and illusion

It might seem like Leibniz's work here is done. He has closed the gap between the object in the world and the perception or notion of the object, at least at the level of the *petites perceptions*. However, this is not enough to truly close the early modern gap. Recall Hume's statement in his *Treatise* that secondary quality ideas are "impressions in the mind, derived from the operation of external objects, and without any resemblance to the [primary] qualities of the objects."[28] If there is no resemblance between the qualities of the objects in the world which correspond to our petites perceptions and the qualitative nature of our confused perceptions, then how do our confused perceptions take on the character of secondary quality ideas they have? Leibniz seems to have moved the problem from being a gap between the world and our mind to a gap between our perceptions and their constituents.

Could these qualitative properties be present at the level of the petites perceptions? Consider Theophilus' (Leibniz's mouthpiece in the *New Essays*) response to Philalethes' (Locke's stand-in) presentation of the early modern explanatory gap:

> if we did know the internal constitution of such a body, we would only find such primary (or, as you call them, manifest) qualities as might depend upon it — i.e. come to know what sizes, shapes and moving forces depend on it. **But we would never know what connection they might have with the secondary or confused qualities, i.e. sensible qualities such as colours, tastes and so on.**[29]

Leibniz thinks that even if we knew the internal constitution of bodies, the things represented by those primary qualities and our *petites perceptions*, we still would not understand how our ideas of sensible qualities come about. So, even if I understood everything about the inner constitution of a rose, including the speed and size of the small globes that cause my sensation of phenomenal redness, I would not be able to explain the connection between the small globes and phenomenal redness. The petites perceptions represent the world as it is, but the aggregated, confused representation has this occult, phenomenal quality. So, there is no gap between the world and the *petites perceptions*, but between the *petites perceptions* and the confused perception.

[28] Hume, *Treatise*, I.iv.4.

[29] 1704 A VI.6 402-3/NE IV.vi.7, emphasis added.

But, is the inability to understand this connection a merely epistemic limitation? Theophilus continues his above remarks,

> But men often look for 'a knot in a bulrush' [Plautus] and give themselves problems where none exist, by asking the impossible and then bewailing their helplessness and the limits of their insight.[30]

Leibniz suggests here that attempting to solve the early modern gap creates a problem where none exists. But, how could the early modern gap be an artificial problem? Here is where Leibniz makes his radical move, dissolving the early modern gap:

> 'sensory images', like colours and tastes and so on ... these ought to be called 'images' [*fantôme*] rather than 'qualities' or even 'ideas'. ... It would be enough for all our purposes if we understood them as well as we do that artificial transparency: it would be neither reasonable nor possible to profess to know more; for it is self-contradictory to want these confused images [*fantôme*] to persist while wanting their components to be discerned by the imagination [*fantaisie*] itself. It is like wanting to enjoy being deceived by some charming perspective and wanting to see through the deception at the same time — which would spoil the effect.[31]

Here, Leibniz describes the troublesome sensory qualities as "*fantôme*" and denies them the label of "idea" and "quality." In short, I understand Leibniz to deny the reality of the qualitative aspect of secondary quality ideas. The phenomenal redness of the rose is, for Leibniz, no more real than the blur created by a spinning cog. When we try to explain the blur in terms of the cog's teeth and their movement, we do not attempt to preserve the reality of some transparent object. Rather, we accept that the transparency is an artifact of our confused notion of the cog. In the same way, for Leibniz, we should accept that sensory qualities are a mere artifact of our confused notion of objects. For the remainder of this section, I will elaborate and support this reading by discussing Leibniz's use of "*fantôme*", his theory of ideas, and the role of the imagination.

Notice how the language Leibniz uses conveys the illusoriness or irreality of these sensory images in multiple ways. Specifically, his use of "*fantôme*." It's worth quoting, I think, Peter Remnant and Jonathan Bennett's editorial footnote on this passage, "In this paragraph, 'image'

[30] 1704 A VI.6 404/NE IV.vi.7.
[31] 1704 A VI.6 404/NE IV.vi.7.

always renders '*fantôme*', not '*image*' as elsewhere. '*Fantôme*' has connotations of illusoriness which '*image*' lacks." Leibniz's choice of language is meant to imply the illusoriness of these sensory qualities. Leibniz, I will argue, dissolves the early modern gap by denying the reality of these sensory images, what today we would call "phenomenal properties" like colors, tastes, odors, and all confused notions of a single sense.

The connotations of illusoriness given by "*fantôme*" are echoed in other parts of this text: "rather than 'qualities; or even 'ideas;'" comparison to the "arteficial transparency;" "deception." It is "self-contradictory to want these confused images [*fantôme*] to persist while wanting their components to be discerned by the imagination [*fantaisie*] itself."[32] The natural reading here, I think, is that one should avoid this contradiction by giving up the confused images (the role of the imagination will be discussed below). Leibniz further compares this *fantôme* to deception and making the notion distinct, i.e., to access the implicit notions which underlie the *fantôme*, to "see[ing] through the deception".[33] If these sensory qualities are deceptions, then they cannot be veridical.

I take it to be established that a surface-level reading of the text supports the view that Leibniz takes (at least some) sensory qualities, secondary quality ideas, or phenomenal properties to be illusory. That is, these confused notions appear to have qualitative characters that they actually lack. Put another way, I think Leibniz thinks that when we look at a red rose and form a judgment that we are having an experience of phenomenal redness (something like the contemporary idea that there is something it's like to see red) that judgment is mistaken. The appearance of phenomenal redness is merely a confused result of how perception receives the notion.

To bolster this case, I will evaluate why Leibniz denies these notions the status of "idea" — "these ought to be called 'images' [*fantôme*] rather than 'qualities' or even 'ideas'."[34] This might sound odd if one has, like Locke, a very permissive definition of "idea,"

> Every Man being conscious to himself, That he thinks, and that which his mind is employ'd about whilst thinking, being the Ideas, that are there.[35]

[32] 1704 A VI.6 404/NE IV.vi.7.
[33] 1704 A VI.6 404/NE IV.vi.7.
[34] 1704 A VI.6 404/NE IV.vi.7.
[35] Locke, *Essay*, II.I.1. See also, II.VIII.8.

For Locke, an idea is whatever is present before the mind when thinking; it is the direct object of thought. This view is far too permissive for Leibniz. What Locke terms "ideas" is analogous to Leibniz's technical term "notion." Not all of our thoughts are of ideas,

> we don't perceive ideas of even those things we know distinctly, unless we make use of intuitive thinking. And, indeed, it happens that we often mistakenly believe that we have *ideas* of things in mind when we mistakenly suppose that we have already explained some of the terms we use.[36]

Intuitive thinking, for Leibniz, occurs when we "grasp the entire nature of a thing all at once... [and] consider all of its component notions at the same time."[37] In contrast to intuitive thinking, there is "blind" or "symbolic" thinking where "we make use of signs, whose explicit explanation we usually omit for the sake of brevity, knowing or believing that we have the ability to produce it at will."[38] A sign or symbol is anything that stands in the place of a complex notion in thought, that doesn't explicitly involve all of the notion's component complexity. A sign, then, seems to be when we treat a composite notion as a simple notion by replacing the complex notion with a sign or symbol. Leibniz thinks such thinking happens "almost everywhere" in our thinking.[39]

Why do sensory images fail to qualify as ideas? Here we seem to have a first approximation of an answer. Distinct notions fail to qualify as ideas unless we have intuitive understanding of them. Confused notions, failing to meet even the requirements for distinctness, also fail to be ideas. However, this does not explain why Leibniz further denies that these sensory images are qualities. There is more to the story. Symbolic thinking can also explain how we can mistakenly believe we have an idea. Leibniz continues,

> For, often, we do understand in one way or another the words in question individually or remember that we understood them previously. But since we are content with this blind thinking and don't pursue the resolution of notions far enough, it happens that a contradiction that might be included in a very complex notion is concealed from us.[40]

[36] 1684 A VI.4 588/AG 25, emphasis in original.
[37] 1684 A VI.4 587/AG 24-5.
[38] 1684 A VI.4 587/AG 24.
[39] *Ibid.*
[40] 1684 A VI.4 588/AG 25.

"[B]lind thinking," remember, is symbolic thinking. Because a symbolic notion can hide a contradiction, we might mistakenly take such a notion to be an idea when it is not.

> And so we also have a distinction between *nominal definitions*, which contain only marks of a thing to be distinguished from other things, and *real definitions*, from which one establishes that a thing is possible. ... an idea is true when its notion is possible and false when it includes a contradiction. Moreover, we can know the possibility of a thing either *a priori* or *a posteriori*. The possibility of a thing is known *a priori* when we resolve a notion into its requisites, that is, into other notions known to be possible, and we know that there is nothing incomplete among them. ... The possibility of a thing is known *a posteriori* when we know through experience that a thing actually exists, for what actually exists or existed is at very least possible. ... For the most part we are content to have learned the reality of certain notions through experience, from which we then compose others following the example of nature.[41]

If a thought includes a contradiction, then it is false. Leibniz contrasts false ideas with true ideas, which are possible. Given this contrast, we can conclude that false ideas are impossible. This seems right, as we should not want contradictions to be possible. If the notions involved in *fantôme* are notions of impossible things, then these *fantôme* cannot be qualities as there are, plausibly, no impossible qualities.

One reading of the above passage suggests that Leibniz thinks there are false ideas, that is, ideas of impossible things or ideas which involve a contradiction. However, earlier in the same paper, Leibniz states that "we certainly have no idea of impossible things."[42] This earlier passage suggests that Leibniz thought that if a notion contained a contradiction, what above was termed a "false idea," or a thought hid a contradiction behind blind or symbolic thinking, it is not an idea but is merely a notion. I need not take a stand here on whether there are false ideas or merely contradictory notions, though I do find it interesting that Leibniz denies confused sensory images the status of idea.[43]

But, what is an idea, then? For Leibniz, ideas seem to be either complexes of consistent notions or the primitive notions (which, being primitive, are consistent with themselves, there

[41] 1684 A VI.4 589/AG 26, emphasis in original.
[42] 1684 A VI.4 589/AG 25.
[43] 1704 A VI.6 404/NE IV.vi.7.

is no complexity to generate a contradiction).[44] Because the notions involved in an idea are consistent, they form a real definition, which shows the possibility of a thing.

I think Leibniz denies sensory images the status of idea and quality for the same reason: their notions hide a contradiction. Leibniz is explicit in the above text in stating that involving a contradiction is a disqualification for being an idea. Plausibly, there are no contradictiory qualities. Leibniz never outright states that the confused senory image involves a contradiction, only that "it is self-contradictory to want these confused images [*fantôme*] to persist while wanting their components to be discerned by the imagination [*fantaisie*] itself."[45] I will discuss the role of imagination below, but we already have the resources to show that Leibniz is committed to the notions involving a contradiction. Leibniz is committed to the PSR. According to Leibniz, the PSR follows from his predicate in notion theory of truth.[46] According to this view, all truths have a subject-predicate structure and the concept of the predicate is contained in the concept of the subject. Our sensory images are inexplicable in terms of their causes.[47] Given the predicate in notion theory, if x is inexplicable in terms of y, then y must not be contained within the concept of x.

If our notions of sensory images are neither qualities nor ideas and, following the typology of "MKTI," ideas are real definitions or consistent composites of notions, then the content represented in sensory images must not be possible because they are not ideas. That is, the confused notion we have of them must hide some contradiction. If sensory images are not possible, then they cannot be qualities of objects nor qualities in the mind.

Further, he compares phenomenal properties to the transparency of the spinning cog, "It would be enough for all our purposes if we understood them as well as we do that artificial transparency: it would be neither reasonable nor possible to profess to know more".[48] When we analyze what's happening in perception with the transparent blur resulting from the spinning cog, we don't try to preserve the reality of the blur as something real in the world. It is, as

[44] 1684 A VI.4 589/AG 26. See also Leibniz's proof of the possibility of God, discussed in section 4.5 below, which is based on the attributes being independently possible and, having no overlapping nature, jointly compossible. (DSR 90-95)

[45] 1704 A VI.6 404/NE IV.vi.7.

[46] 1680(?) A VI.4 1443/AG 31

[47] 1704 A VI.6 402-3/NE IV.vi.7.

[48] 1704 A VI.6 404/NE IV.vi.7.

Leibniz says in the text quoted above, a deception. In denying the status of ideas to sensory images, Leibniz does more than deny their reality, he denies their possibility. For the same reason, I think, Leibniz denies that sensory images are qualities.

Also, consider the following passage as a further example of Leibniz's commitment to the illusoriness of secondary quality ideas,

> the sensation of warmth or of light results from many tiny motions which, as I said earlier (viii.13), express the motions in objects, and are different from them only in appearance, and that only because we are not aware of this analysed multiplicity.[49]

Here, Leibniz again states that our notions of warmth and of light, which I argued above are confused and representationally simple. These confused notions result from a complex cause, in this case, many tiny motions. Despite being confused, our notions of warmth or of light do express the tiny motions that cause them, because effects must correspond to their causes. Indeed, our notions of warmth or of light differ from their causes only in their appearance, e.g., the feeling of the warmth or the color of the light, and this is due solely to the fact that the notion is confused, i.e., because we have not analyzed the multiplicity of notions involved in the confused notion. Were we to analyze the confused notion into its component distinct notions, this appearance of difference would not remain. Leibniz continues,

> As against this view, some contemporaries believe that our ideas of sensible qualities differ entirely from motions and from what occurs in the objects, and are something primary [*primitif*] and inexplicable and even arbitrary; as though God had made the soul sense whatever he had a whim that it should sense, rather than whatever happens in the body — which is nowhere near the right analysis of our ideas.[50]

Leibniz now contrasts his view with some of his contemporaries, specifically his understanding of Descartes and other Cartesians. Leibniz understood the Cartesians to hold that the relation between external objects and the notions of secondary qualities they cause in us were completely arbitrary and decided by God. Notice here that Leibniz contrasts his view with the view that notions like those of colors or warmth are primary, basic, or primitive as they appear. Despite their appearing primitive or lacking any internal complexity, which Leibniz grants, the notions involved are deeply complex.

[49] 1704 A VI.6 165/NE II.xx.6.
[50] 1704 A VI.6 165-66/NE II.xx.6.

But why is it "self-contradictory to want these confused images [*fantôme*] to persist while wanting their components to be discerned by the imagination [*fantaisie*] itself"?[51] Leibniz says the following about the imagination,

> since our soul compares the numbers and shapes that are in color, for example with the numbers and shapes that are in tactile qualities, there must be an *internal sense* in which these different external senses are found united. This is called the *imagination*, which contains both the notions *of the particular senses*, which are *clear but confused*, and the *notions of the common sense*, which are *clear and distinct*. ... We also see that particular sensible qualities are capable of being explained and reasoned about only insofar as they contain what is common to the objects of several external senses, and belong to the internal sense.[52]

Thus, the imagination is the faculty where notions received through external senses are compared and contrasted. It is through this process that the implicit components of these notions are made explicit and we come to have distinct notions. The imagination does this by comparing notions received through different senses and finding they contain some common notions between them. Through the imagination finding these commonalities, we come to form nominal definitions or distinct ideas. For example, I can receive the notion of extension from gazing at the table in front of me and from running my hand along its surface. By comparing the notions of extension from sight and touch, I can gain a distinct notion of extension. The same cannot be done with secondary quality ideas because their notions are gained from a single sense. Unlike extension, sesnory qualities are gained only through a single sense. So, for example, on could not compare the notion of phenomenal blue gained through sight with the notion of phenomenal blue gained through any other sense modality because that notion is not gained through any other modality. Therefore, if the imagination discerned the implicit components of a confused notion of, say, phenomenal red and phenomenal red is necessarily confused, it is contradictory to want phenomenal red to persist while its implicit notions are made distinct because the notion cannot be both confused and distinct at the same time to the same mind.

Consider, also, this text,

> But the colour yellow is a reality, all the same, like the rainbow. ... if our eyes became better equipped or more penetrating, so that some colours or other qualities

[51] 1704 A VI.6 404/NE IV.vi.7.
[52] 1702 G 499/AG 187-88, emphasis in original.

> disappeared from our view, others would appear to arise out of them, and we should need a further increase in acuity to make them disappear too; and since matter is actually divided to infinity, this process could go onto infinity also.[53]

The rainbow is, of course, Leibniz's go-to example for something that exists only as well-founded phenomena, "Bodies, which are commonly taken for substances, are nothing but real phenomena, and are no more substances than perihelia or rainbows."[54] But, colors being well-founded like the rainbow does not prevent Leibniz from holding that they are illusory. Leibniz holds that things such as "heat, colors, parhelia, beings of abstraction, time, number, motions itself and shape ... have a foundation in nature, but no final reality."[55] These notions are true phenomena, they have causes in the world, but that need not mean their confused representations are veridical or have reality. Sensory images have causes, the *petites perceptions*, but no final reality.

Leibniz thinks solving the early modern gap is "impossible" and creates a problem "where none exist[s]"[56] because he denies the reality of sensory images. It is impossible to explain the connection between, say, the rotation of small globes and phenomenal redness because there is really no such thing as phenomenal redness just as there is really no such thing as the transparent blur of the cog. The early modern gap creates a problem where none exists for the same reason. Leibniz denies the reality of sensory images because they are impossible, hence one cannot give a real definition of such notions and, therefore, they cannot qualify as true ideas (if they can qualify as ideas at all).

4.4 Leibniz and twenty-first century illusionists

I have argued that in these passages from the *New Essays* and other texts, Leibniz is defending a form of what we would call today "illusionism" or "eliminativism" about the qualitative nature of sensory qualities. Illusionism is a new name for an old position. Today, it is prominently defended by Dennett, Frankish, and Pereboom among others. For the purposes of my discussion,

[53] 1704 A VI.6 219/NE II.xxiii.12.
[54] LD 21 January 1704, 287. See also LD 30 June 1704, 303 and LA 8 December 1686, 157.
[55] LA 30 April 2687, 183.
[56] 1704 A VI.6 404/NE IV.vi.7.

I will focus on how Frankish explicates the position, as the parallels are clearest. Illusionism is the conjunction of functionalism and a distinct form of eliminativism about the qualitative nature phenomenal properties are represented by introspection as having.[57] What makes the illusionist's eliminativism distinct is their explanation of why we have the beliefs we do about phenomenal properties; why it seems like there is a hard problem. In other words, they provide an answer to David Chalmers' meta-problem of consciousness.[58] The illusionist's answer is that "introspection delivers a partial, distorted view of our experiences, misrepresenting complex physical features as simple phenomenal ones."[59] This should be sounding familiar. On this view, when we introspect on a feeling of painfulness, introspection represents a complex brain state as simple phenomenal painfulness. This misrepresents the brain state as having qualitative features it actually lacks. Specifically, introspection misrepresents our brain states as having a what-it-is-like nature. While details will differ between illusionist accounts about how this occurs, the view is something like that introspection on the brain state that plays the functional role of pain causes us to form the belief that "I am in [phenomenal] pain." Because introspection is our only way to access the nature of our conscious states, we are unable to realize this error.

By arguing that Leibniz is defending a form of illusionism, I am not arguing that Leibniz is a functionalist about the mind. Instead, I want to draw attention to the four following parallels between Leibniz's position on sensory qualities and the illusionist's. First, both Leibniz and the illusionist believe that our representations of phenomenal properties arise from misrepresenting some complex state as simple. For Leibniz the complex of the aggregate notions of *petites perceptions* involved is represented simply in introspection as a secondary quality idea; for the illusionist the complex state of the brain is represented simply in introspection as a phenomenal property idea. Arguably, in both cases, this is the result of notions received from a single sense being necessarily confused. Our confused notions of phenomenal properties are only gained through introspection, a kind of inner sense that misrepresents complex brain states as being simple phenomenal properties. Further, it is plausible that this representation is necessarily

[57]Frankish (2016, 14) is explicit about illusionism being functionalist. Cf. Pereboom (2016).

[58]David J. Chalmers, "The Meta-Problem of Consciousness," *Journal of Consciousness Studies* 25 (2018).

[59]Keith Frankish, "Illusionism as a Theory of Consciousness," *Journal of Consciousness Studies* 23 (2016): p. 18. See also Pereboom (2011).

confused for the illusionist as it is for Leibniz.

Second, the notions that result from these misrepresentations are non-veridical. Specifically, they are, in a sense, a deception or illusion. That is, for Leibniz, our confused secondary quality ideas do not accurately resemble anything in the world or our experiences; for the illusionist, the same can be said for our ideas of phenomenal properties.

> I have just pointed out how there is a resemblance, i.e., a precise relationship, in the case of secondary qualities as well as of primary. It is thoroughly reasonable that the effect should correspond to the cause; and how could one ever be sure that it does not, since we have no distinct knowledge either or the sensation of blue (for instance) or of the motions which produce it? *It is true that pain does not resemble the movement of a pin; but it might thoroughly resemble the motions which the pin causes in our body, and might represent them in the soul; and I have not the least doubt that it does.* That is why we say that the pain is in our body and not the pin, although we say that the light is in the fire; because there are motions in the fire which the senses cannot clearly detect individually, but which form a confusion — a running together — which is brought within reach of the senses and is represented to us by the idea of light.[60]

Leibniz allows that there is a resemblance at the level of the *petites perceptions*, the representations of the motions which cause the feeling of pain or the sensation of blue, but, as discussed above, there is no resemblance between the confused notion of blue or pain and its cause. Leibniz believes there is a resemblance between the notion and the cause in the implicit structure of the notion. Leibniz is committed to this due to his view on causes, effects, and their correspondence. So, if there were a complex cause, such as the internal constitution of a rose or a brain state, the effect, the sensation of red in the soul, must also be complex.[61] Similarly, the illusionist "denies that the properties to which introspection is sensitive are qualitative: it is an illusion to think there are phenomenal properties at all."[62]

Third, both Leibniz and the illusionist think that this deception would not persist if one had an adequate, or, at least, distinct, understanding of the misrepresented complex. Leibniz is explicit on this point,

[60] 1704 A VI.6 131-2/NE II.viii.15, emphasis added.

[61] Of course, Leibniz does not think, speaking with metaphysical rigor, that the rose causes changes in states of the brain or that the brain or body causes changes in states of the soul as there is no inter-substantial causation on Leibniz's considered view and the direct object of perceptions are really states of one's soul. But, it is useful to speak in this way.

[62] Frankish, "Illusionism," p. 15.

> for it is self-contradictory to want these confused images [*fantôme*] to persist while wanting their components to be discerned by the imagination [*fantaisie*] itself. It is like wanting to enjoy being deceived by some charming perspective and wanting to see through the deception at the same time — which would spoil the effect.[63]

While the illusionist is not explicit on this, they seemed to be committed to the view. Consider the famous Mary thought experiment.[64] That Mary can know all the physical facts about seeing red and not have an idea of phenomenal redness speaks in favor of this. Her coming to have the confused, what-it-is-like notion of red only occurs when she represents those facts through introspecting on her own experience of seeing something red.

Fourth and finally, Leibniz and the illusionists take a similar stance toward the explanatory gaps argued for by their contemporaries. That is, they both think that what is to be explained and resolved is not the gap itself, but why we are inclined to think there is a gap. For Leibniz, this is a rejection of the early modern explanatory gap between the primary qualities of objects and the secondary quality ideas they cause in us. Recall Leibniz's assessment of this gap,

> But men often look for 'a knot in a bulrush' [Plautus] and give themselves problems where none exist, by asking the impossible and then bewailing their helplessness and the limits of their insight.[65]

The early modern gap, for Leibniz, is not resolved by maintaining the reality of secondary quality ideas and explaining how they are caused by the primary qualities in the world. Rather, it is resolved by not "want[ing] these confused images [*fantôme*] to persist while wanting their components to be discerned by the imagination [*fantaisie*] itself."[66] The first step to doing this, for Leibniz, is recognizing that these notions of secondary quality are confused notions presented the way they are because they are the product of a single sense and that this presentation hides a contradiction, so the confused notion cannot be real.[67] Similarly, for the illusionist, the explanatory gap and the hard problem of consciousness are not solved by explaining how phenomenal properties arise from physical processes in the brain. Rather, they are resolved by addressing the meta-problem, explaining why we think there is such a problem in the first

[63] 1704 A VI.6 404/NE IV.vi.7.
[64] Frank Jackson, "Epiphenomenal Qualia," *Philosophical Quarterly* 32 (1982): p. 130.
[65] 1704 A VI.6 404/NE IV.vi.7.
[66] 1704 A VI.6 404/NE IV.vi.7.
[67] 1684 A VI.4 586–87/AG 24.

place.[68] The answer, the illusionist suggests, is that "introspection delivers a partial, distorted view of our experiences, misrepresenting complex physical features as simple phenomenal ones."[69] In both cases, the appearance of a problem is caused by some faculty misrepresenting the nature of a mental state as having simple properties it does not have.

4.5 Objections

One might object to this interpretation in the following way: if the notion of phenomenal redness is a confused representation of *petites perceptions* and petites perceptions are distinct perceptions of the rotating of small globes and phenomenal redness is not an idea because it is impossible (i.e., contradictory), then where is the contradiction? Surely (ding! [Dennett 2013]), the petites perceptions expressing the rotation of small globes are not contradictory themselves? Further, the aggregation of these expressions of rotating small globes must not be contradictory either. So, this must not be Leibniz's view.

In answer to this, I would like to draw a parallel between this issue and another issue in Leibniz surrounding his proof of God's possibility in *De Summa Rerum*. Leibniz's proof is as follows, God is the being with all the attributes. Attributes are positive, not negative. Attributes can either be composites or primitives, the composites being aggregates of the primitives. So, to show the possibility of God, Leibniz must show that all of the primitive attributes are compossible. Consider two primitive attributes, A and B.

> If they are incompatible, then the proposition "Quality A and quality B cannot be in the same subject" will be necessary, and so will be either an identical proposition or one which is demonstrable. It cannot be an identical proposition; for then "Where A is, B cannot be" would be the same as "A is A" or "A is B," and so one would express the exclusion of the other, and so one of them would be the negative of the other. But, this is contrary to the hypothesis, for we have assumed that all attributes are affirmative. It is not demonstrable; for if one were to demonstrate, "Where A is, B cannot be," that could only be done by the analysis of one or other term, or of both. This is contrary to the hypothesis; for we have assumed them to be unanalyzable. Therefore this incompatibility cannot be demonstrated. Therefore,

[68] Keith Frankish, "The Meta-Problem is the Problem of Consciousness," *Journal of Consciousness Studies* 26 (2019).

[69] Frankish, "Illusionism," p. 18. See also, Pereboom (2011).

> there is no incompatibility, and so any two affirmative qualities are compatible, and so all such qualities are compatible with all. Therefore a being which has all attributes is possible.[70]

Here, Leibniz shows that if primitive attributes were incompatible, then their incompatibility would either be an identity claim or it would be demonstrable that they were incompatible. It cannot be an identity claim, because then one attribute would include the negation of the other. But, by assumption, attributes are only characterized positively. It cannot be demonstrable, because such a demonstration would require the analysis of terms. But, by assumption, A and B are primitives and, therefore, unanalyzable into more basic concepts. Therefore, A and B must be compossible, so any primitive notions must be compossible.

Leibniz has a concern about his above argument related to the current objection under consideration. If primitive attributes are all compossible and composite attributes are composites of compossible primitives, then there is no clear reason why the composite attributes should be incompossible with other composite or primitive attributes. As such, it is unclear of what impossibility consists.[71] How does this relate to the current objection? Recall from MKTI that our primitive notions are the absolute attributes of God[72] and these are the same attributes Leibniz discusses in his proof of God's possibility, then how could there be a contradiction (given that the proof shows they must be consistent)? It seems to be a general problem with Leibniz's accounts of our ideas and possibility (the two are intimately related) about how to account for contradiction or impossibility.

I do not have an answer to the current objection, but it is notable that it is a problem in Leibniz studies generally and not just for my interpretation.

[70] DSR 93.
[71] Thanks to Mike Griffin for drawing my attention to this.
[72] 1684 A VI.4 590/AG 26.

Chapter 5

Constitutivism and Confusion

There are a group of theories of consciousness I call "constitutivism" which fall prey to the structural mismatch problem, the problem of the apparent discrepancy between the structure of our mental states as they appear to us (phenomenal experiences, thoughts, etc.) and the structure of the material correlates, e.g., brain states, of that experience. Views are constitutivist if they accept that the mind is identical to, realized by, or grounded in material states. As in previous chapters, something is physical if it is "(approximately accurately) treated by current or future (at the end of inquiry, ideal) physics, and [are] not fundamentally mental."[1] Something is material if it meets the requirements for being physical, whether or not it is fundamentally mental.[2] The structural mismatch problem arises due to a commonly accepted view about introspection accurately representing both the nature and the structure of our conscious states. Typically in the debate, there are defenses of two extremes. On the one hand, there are revelation theorists who hold that introspection presents us with the complete essence of the type the mental state belongs to.[3] On the other hand, there are those like the illusionists, who hold that introspection systematically misrepresents the nature of our mental states.[4] Solving the problem then, as I see it, lies in articulating a position weaker than revelation and stronger than illusionism. I think this can be achieved using insights from G.W. Leibniz, particularly a version of his confused ideas thesis. In the previous chapter, I argued that Leibniz's version

[1] Jessica Wilson, "Supervenience-Based Formulations of Physicalism," *Noûs* 39 (2005): p. 428.
[2] Phillip Goff, *Consciousness and Fundamental Reality* (New York: Oxford University Press, 2017a), p. 41.
[3] See, historically, Descartes (1985) and, contemporarily, Goff (2017a).
[4] See Pereboom (2011) and Frankish (2016; 2019).

of the confused ideas thesis was an illusionist view. I will articulate a version of the confused ideas thesis which maintains a realism about qualitative character introspection represents phenomenal properties as having. In this chapter, I first lay out a general form of the structural mismatch problem for constitutivism. Then, I motivate and defend a version of Leibniz's theory of confused ideas. Finally, I show how constitutivists can use the theory of confused ideas to avoid the structural mismatch problem.

5.1 Constitutivism and the structural mismatch problem

I will use "constitutivism" to refer to a family of views that posit that the mind is identical to, realized by, or grounded in material states, usually brain states. I'll use "constitute" as a shorthand for the disjunction of these three options and will use "brain" as a shorthand for whatever basal material states constitute the mental states without ruling out views like extended or embodied cognition. Tim Crane takes constitutivism to be necessary condition for any form of materialism.[5] However, it applies to a variety of views that are arguably not materialist. Views under this umbrella include, but are not limited to, constitutive panpsychism, weak emergentism, illusionism, mind-brain identity theory, neutral monism, and some forms of property dualism.[6] Because illusionism falls under the umbrella of constitutivism, it should be clear that to be a constitutivist does not commit one to realism about the qualitative nature of phenomenal consciousness. Constitutivism casts a wide net, capturing both prominently defended and minority views among philosophers of mind. For the purposes of this chapter, I will assume the truth of constitutivism.

The constitutivist must reject, for example, the view of the British emergentists and contemporary strong emergentists view that the mental is strongly emergent from the material.[7]

[5]Tim Crane, "Cosmic Hermeneutics vs. Emergence: the Challenge of the Explanatory Gap," in *Emergence in Mind*, ed. Graham Macdonald and Cynthia Macdonald (New York: Oxford University Press), p. 30.

[6]For consititutive micropsychism, see Chalmers (2015; 2017), Roelofs (2019; 2021) and chapter 2 of this dissertation. See Pereboom (2011) and Wilson (2021) for weak emergence. Perebooom (2011), Frankish (2016), and chapters 3 and 4 of this dissertation for illusionism. For the mind-brain identity theory, see Lewis (1966), Coleman (2013; 2017) and Stubenberg (2018) for neutral monism, and, finally, Chalmers (1996) for property dualism.

[7]For an overview and critique of the British emergentists, see Mclaughlin (2008). For contemporary accounts, see Chalmers (2006) and Wilson (2021).

The mental is strongly emergent from the material if it is dependent on some base material facts, but cannot be wholly explained by the base material facts. Instead, brute psycho-material laws must be invoked to explain the correlation between material states and mental states. Further, constitutivists must reject that mental substances are distinct from material substances, as Cartesian dualism proposes. This is not to imply that there is not some form of the structural mismatch problem for non-constitutivists, just to clearly delineate constitutivists from other views.

There are various forms of the structural mismatch problem, with different formulations specific to the position it is attacking.[8] Some versions of the problem are formulated noting the apparent difference between the structure of propositional thoughts and the brain states which constitute them. For this chapter, I will focus on the apparent discrepancy between the structure of phenomenally conscious states and the material states that constitute them.

Consider, first, Grover Maxwell's presentation of a general version of the structural mismatch problem,

> How is it that the occurrence of a smooth, continuous expanse of red in our visual experience can ... involve particulate, discontinuous affairs such as transfers of or interactions among large numbers of electrons, ions, or the like? Surely being smooth or continuous is a structural property, and being particulate or discontinuous is also a structural property ... incompatible with being smooth and continuous.[9]

Here, Maxwell draws attention to the mystery of how a smooth, continuous experience of red can arise from numerous particulate and discontinuous physical states of affairs which constitute that experience. Note, this is not a statement of the hard problem of consciousness, i.e., the problem of explaining why there is an experience that accompanies the material goings-on.[10] Rather, this is the problem of explaining how a mental state can have a different structure than its basal material state given that there is a tight dependence relation between the two. Why is there not a match between the structure of the mental state and the material state the way there is between other higher-level entities and their basal material constituents? For example,

[8] See Chalmers (2017), Goff (2017a), Nagasawa & Wager (2017), Roelofs (2019), and Stoljar (2001).

[9] Grover Maxwell, "Rigid Designators and Mind-Brain Identity," *Minnesota Studies in the Philosophy of Science* 9 (1978), p. 398.

[10] David J. Chalmers, "Facing Up to the Problem of Consciousness," *Journal of Consciousness Studies* 2 (1995).

consider the statue and the lump of clay. The statue depends on and is not identical to the lump of clay, but there is no structural mismatch between the statue and the lump of clay when there is the statue.

A common question at this point is "what exactly is meant by 'phenomenal structure?'" Authors writing on this problem do not provide an explicit definition. Instead, they attempt to motivate an intuitive grasp of the concept and its mismatch to physical structure through examples, recall Maxwell's "continuous" and "smooth" red experience. I will follow in this tradition. Contrast my visual field right now with a phenomenal experience of a uniform red field (either imagined or by focusing on a plain red surface in one's visual field to the exclusion of everything else). My visual field is comprised of many diverse phenomena. There is a keyboard, a monitor, a flickering candle, a coffee mug, a textured white wall, and so on. It seems that the content of my visual field has a lot of structure or complexity to it, with each of the objects in that field also having a complex structure. The multicolored keys of the keyboard are many different splotches of color arranged in rows, the light from the candle is flickering against the blue glass of its container, the mug is largely a uniform white but there is also some black and red writing in the middle. Further, all these things are arranged spatially; the mug is behind the keyboard but in front of the monitor. In contrast, the uniform red field experience lacks most of this structure. Aside from the spatial aspect to it, there is nothing to differentiate the left side from the right. Now, contrast both these structures to the plausible material correlates of these experiences, signals traveling down the optic nerve to the occipital lobe, complex patterns of neurons firing, and, at lower levels of description, vast chemical reactions and sub-atomic activity. It is unclear how these vastly different structures relate.

To further illustrate the apparently simple structure of phenomenal properties, contrast the discerning drinker of gin and tonics from the amateur drinker. The amateur drinker, one who has never had a gin and tonic before or has only had a few, might describe the taste after introspecting as "like gin" or, my favorite report, "like a pine tree," in the way that someone might say a dish "tastes like chicken." The taste of the gin and tonic to them is simple, it lacks complexity. The simple structure of the amateur drinker's experience clashes with the complex brain structure, the vast array of neurons firing which constitutes the experience. However,

the discerning drinker, who has had many gin and tonics, is able to pick out notes of specific botanicals from the gin, the acidity of the lime, and the slight bitterness of the quinine in the tonic. Even though the discerning drinker finds complexity in her experience of the gin and tonic, the various aspects of the flavor experience (the botanicals, the acidity, the bitterness) each appear to her as simple as the amateur drinker's experience of "the pine tree." That is, the flavors do not appear to be a complexes of further phenomenal properties or anything else. Introspecting on the flavor of the botanicals does not allow her to pick out more basic flavors or constituents of the experience. Further, the structure of the discerning drinker's experience also clashes with the constituting complex brain structure. The phenomenal feel of the botanicals, the acidity, and the bitterness each appear simple, unlike the brain states.

The structural mismatch problem can also be presented in a more targeted way. David Chalmers presents the following version of the structural mismatch problem against constitutive panpsychism,

1. Microphenomenal structure is isomorphic to micro[material] structure.

2. Microphenomenal structure constitutes macrophenomenal structure.

3. Micro[material] structure constitutes macro[material] structure.

4. Macrophenomenal structure is not isomorphic to macro[material] structure.[11]

Premise 1 is a statement of the Russellian view that microphenomenal properties, i.e., the phenomenal properties that the panpsychist posits fundamental microsubjects experience, serve as the basis for micromaterial role-playing properties.[12] Premise 2 is a statement of constitutive panpsychism, the view that macrosubjects and their macrophenomenal qualities (the phenomenal qualities experienced by creatures like us, dogs, and other macrosubjects) are grounded

[11]David J. Chalmers, "Panpsychism and Panprotopsychism," in *Consciousness and the Physical World*, eds. Torin Alter and Yujin Nagasawa (New York: Oxford University Press, 2015), p. 191.

[12]Is panpsychist Russellian monism constitutivism as I've defined it? A view is constitutivist if and only if it holds the mind is identical to, realized by, or grounded in the material. Something being material is consistent with it being fundamentally mental. So, panpsychist Russellian monism is a form of constitutivism. If phenomenal properties or microsubjects, minds, serve as the basis for fundamental micromaterial role-playing properties, then that seems to violate the requirement that the mind is constituted by the material. In favor of Russellian monism as a form of constitutivism, surely macrophenomenal properties are determined, in part, by their physical structure, even if the Russellian monist posits that structure has mental features for its basis.

in microsubjects and microphenomenal qualities. Premise 3 is a widely accepted view about the dependence of macromaterial properties on micromaterial properties. Premise 4 is a statement about the apparent discrepancy between the structure of macrophenomenal states and the macromaterial states they depend on. For example, a phenomenal experience of a uniform red field seems to be relatively simple — there is no complex internal structure to the experience beyond, perhaps, the spatial aspect. Such an experience seems to lack diverse parts, components, or aspects. However, the brain state which constitutes that phenomenal state is amazingly complex in a way the phenomenal experience does not seem to be. For example, the brain state is compositionally complex, it is made up of numerous neurons which are compositionally complex themselves. The red experience seems to lack parts other than its representation of spatial parts. Further, the brain state is complex in its activity, being a complex network of neural firings and activation patterns. The red experience, in contrast, is not dynamic in this way and appears static while the experience persists.

A general form of the structural mismatch problem can be stated which applies to all forms of constitutivism. The problem arises from the inconsistency of four initially and independently plausible propositions.

5. Phenomenal properties seem or appear to be simple. (phenomenal simplicity)

6. Introspection presents us with an accurate representation of phenomenal properties. (introspective reliability)

7. Brain states are complex. (material complexity)

8. The mind is grounded in, realized by, or identical to the brain. (constitutivism)

5 is a statement about how phenomenal properties seem or appear to us in introspection, lacking any internal complexity. When I introspect on the feeling of phenomenal blueness it does not appear to have an internal structure, there are no "parts" to phenomenal blueness. 6 claims that the way introspection presents phenomenal properties to us, i.e., as simple, is an accurate representation of their nature and structure. There is no systematic error in the way introspection presents our phenomenally conscious experiences to us. 7 is an empirical claim about the

complex structure of brain states. 8 expresses the core commitment of what I have dubbed "constitutivism."

Despite their being independently plausible, accepting 5 through 8 results in the structural mismatch problem. 5 and 6 imply that phenomenal properties are as they appear: simple. In conjunction with 8, it follows that this simple state is grounded in, realized by, or identical to some complex brain state. But, by 7, these brain states are complex. How can phenomenal mental states be simple, yet be closely related to complex brain states? Why is the complexity of the basal brain state not reflected in the mental state?

How might various positions resolve this problem? I am not aware of any who deny 5, even the illusionists, as phenomenal properties appearing to be simple is one reason illusionists reject their existence.[13] The illusionists reject 6 and avoid the structural mismatch problem by denying that introspection accurately represents phenomenal properties, because they deny the reality of phenomenal properties as they are qualitatively represented in introspection.[14] However, most realists about phenomenal consciousness must accept something like 6.

I will discuss two notable rejections of 7. First are the fusionists, who think the brain is an extended simple. Though fusionism is a form of strong emergentism, it accepts that the mind is grounded in the brain, even if the brain itself is emergent, so it falls under the umbrella of constitutivism. How can the fusionists claim that the structures and patterns of activity in brains are simple? Through distributional properties.[15] Imagine a polka-dotted field. There are two ways to describe such a field. You could describe the coloration at individual units of space that make up the field and explain its polka-dottiness as arising from the relation of coloration at those units. Alternatively, one could take the distributional properties approach which takes the whole field as prior to its parts and explains its polka-dottiness as a pattern distributed across that field. The fusionists see brain activity in a similar way. This solution requires one to accept strong emergence. Later, in section 5.2, I present an argument against strong emergence. Further, these responses rely on controversial views about the brain as an

[13] Keith Frankish, "Illusionism as a Theory of Consciousness," *Journal of Consciousness Studies* 23 (2016), p. 18.

[14] Frankish, "Illusionism," 15. For more on illusionism and my argument against it, see chapter 3.

[15] William Seager, "Panpsychism, Aggregation and Combinatorial Infusion," *Mind and Matter* 8 (2010). Hedda Hassel Mørch, *Panpsychism and Causation* (Ph.D. diss, Oslo, 2014).

extended simple. I will suggest a solution that does not rely on this.

Second, Lockwood's preferred solution to the structural mismatch problem involves rejecting 7 by recognizing the brain as a quantum system with merely conceptual subdivisions which need not "correspond to anything that would seem natural; nor need the subsystem in question correspond to what would ordinarily be thought of as a *part* of the brain."[16] Lockwood's solution depends on a controversial empirical thesis about the brain as a simple quantum system. I hope to present a solution that does not require such commitments.

Because rejecting 5, 6, or 7 each seem unacceptable, non-constitutivists reject 8, constitutivism. Without the tight relation between mental and material, there is less to motivate the need for their structures to match. Why is constitutivism worth saving? I think one prominent reason is that the alternative positions raise problems arguably worse than the mind-body problem. Substance dualists must contend with the problem of mental causation and emergentists must face both the problem of mental causation as well as the, I think, more pressing problem of accepting brute, high-level psycho-material laws — the dreaded ontological danglers. (Smart 1959) All theories of mind raise problems in their answer to the mind-body problem. However, the problems raised by the cure should not be worse than the disease.

I agree with 5 or, at least, that they appear to be simpler than their plausible material correlates. Constitutivism commits me to 8 and attempting to avoid the problems of strong emergence or controversial empirical views about quantum physics commits me to 7. Therefore, I must find a way to reject 6. However, unlike the illusionist who also rejects 6, I wish to maintain realism about phenomenal consciousness. While I am not an illusionist, I do think that they are correct when they argue that introspection is the only source of our beliefs about the qualitative nature of phenomenal properties.[17] A rejection of 6, therefore, threatens to undermine our justification for accepting phenomenal consciousness. Yet holding too strongly to 6, e.g., by accepting a view as strong as the full revelation thesis, makes the structural mismatch problem intractable for the non-emergentist constitutivist. The solution then, as I see it, lies in articulating a position weaker than revelation and stronger than illusionism. That is, to find a principled,

[16]Michael Lockwood, "The Grain Problem," in *Objections to Physicalism*, ed. Howard Robinson (New York: Oxford University Press, 1993), p. 287, emphasis in original.

[17]See Frankish (2016) and Pereboom (2011).

middle-ground view of introspection available to the constitutivist which avoids the structural mismatch problem, maintains the possibility for realism about phenomenal consciousness, and does not rely on emergence. The aim of the remainder of this chapter is to articulate and motivate such a position in a manner broad enough that any form of constitutivism may accept it.

5.2 What are confused ideas?

I think the constitutivist would benefit from borrowing an idea from Leibniz, namely, his theory of confused ideas. Some argue it is a bit misleading to call Leibniz's views on confused ideas a singular "theory" rather than a family of views about conceptual capacities, the nature of perception, and, as almost everything in Leibniz seems to ultimately come down to, his theory of expression.[18] It is beyond the scope of this paper to give all these ideas the attention necessary to discuss them in full.[19] However, suffice it to say that I will assume Stephen M. Puryear is correct when he argues,

> in the relevant contexts Leibnizian perceptions are confused in the sense that they have implicit complexity or content. But this is precisely the sense in which ideas are said to be confused too. Leibniz holds that to call an idea clear but confused is to say that though the idea itself is accessible to the mind, its ingredients are, at least for the most part, not. In general terms, his view is that "confusion is when several things are present, but there is no way of distinguishing one from another" (MP 146).[20]

In this section, I aim to explicate the core part of Leibniz's theory that I think the constitutivist should appropriate, I consider a couple of arguments from Leibniz for confused ideas and present one of my own.[21]

In defending the confused ideas thesis, I will use "ideas" in a broad, but somewhat archaic sense to be inclusive of a variety of mental states such as perceptions, thoughts, etc. While

[18] Margaret Dauler Wilson, "Leibniz and Materialism," in her *Ideas and Mechanism: Essays on Early Modern Philosophy*, (Princeton: Princeton University Press, 1999c).

[19] See chapter 4 for a more in-depth discussion of Leibniz's views.

[20] Puryear (2005), "Was Leibniz Confused about Confusion?" 102-3.

[21] Wilson (1999a; 1999b) are excellent discussions of Leibniz's arguments and his views on confusion generally. My discussion of them is indebted to her work.

this use of "ideas" is uncommon today, it is true to the early modern roots of the view. The core of the confused ideas thesis is this: our mental states often have an internal structure or content which is not accessible nor apparent through introspection, i.e., an "implicit" structure or content. I will also follow the early moderns by contrasting confused ideas with distinct ideas. While confused ideas have an implicit structure or content, a distinct idea is one where that structure or content is at least partially accessible to introspection, i.e., explicit. Note also that confusion and distinction admit of degrees. An idea can be more or less confused or distinct if its internal structure or content is more or less accessible to introspection.

To illustrate this, consider again the case of the discerning and amateur gin and tonic drinkers. I think this case serves as an excellent example of confusion. The non-discerning gin and tonic drinker has a more confused experience of the drink as they are unable to pick out the various flavors present. The discerning drinker has a more distinct experience because they have cultivated their palette (a nice way of saying "drank a lot") and gained the introspective capacity to pick out the various flavors. An assumption of this view is that the qualities that the discerning drinker is able to pick out (the botanicals, the bitterness, the acidity) contribute just as much to the amateur drinker's experience as they contribute to the discerning drinker's, even though she is unable to pick them out in introspection. They serve as the implicit structure of the amateur drinker's experience. This assumption seems acceptable, after all, why would they not contribute just as much to each experience?

Consider, also, this example from Leibniz,

> when the swift rotation of a cog-wheel makes us perceive an artificial transparency, as I have noticed on visits to clock-makers, we are not able to discern the idea of the cause of this, i.e., the idea of the teeth of the wheel. The wheel's rotation makes the teeth disappear and an imaginary continuous transparent [ring] appear in their place; it is made up of successive appearances of teeth and of gaps between them, but in such rapid succession that our imagination cannot distinguish them.[22]

When the cog is stationary, the observer is able to distinctly pick out each tooth in her visual experience. But, when the cog is in motion, she experiences only a homogeneous blur. But, assumes Leibniz, the individual teeth still contribute just as much to her experience when they

[22] 1704 A VI.4 403/NE IV.vi.7.

are moving as when they are stationary. Thus, they must be present in the experience in a confused way.

Now that the core of the confused ideas thesis has been laid out, I will consider two Leibnizian arguments for the view. First is the argument from the Principle of Sufficient Reason (PSR), given in response to his understanding of Descartes' view that God arbitrarily establishes the connection between brain states and mental states, e.g., a certain brain state and an experience of phenomenal redness. This arbitrariness is unacceptable to Leibniz, who thought God must always have a reason for acting. Leibniz thought if sensations or ideas were simple, then the relation would have to be arbitrary because causes must correspond to their effects.[23] Therefore, a complex cause, like the physical constitution of a rose or a brain state, must have a complex effect. Of course, Leibniz's considered position does not involve inter-substantial causation, but even he found it easier to speak this way. But, by the PSR, it cannot be the case that the relation between ideas and their causes is arbitrary. So, in order for the relation between the causes of ideas and the ideas themselves to be intelligible, the mental states must be complex.[24].

This argument will probably not be convincing to contemporary philosophers for a few reasons. Of specific concern is the reliance on the PSR, a strong metaphysical principle not in favor today, though not without its defenders.[25] The other concern being the centrality of God to the argument. Though we can see connections to debates today and the common ground of attempting to make intelligible the connection between mental states and physical states — a position that has support, indeed seems to be a common view within contemporary analytic philosophy of mind.

There are some parallels to be drawn between Leibniz's interpretation of the Cartesian position and contemporary emergentism about the mind. Like the Cartesians, contemporary strong emergentists must think that the psycho-material laws are somewhat arbitrary. Mental states are not wholly determined by the material states they depend upon, but also the brute and inexplicable, psycho-material bridge laws. In the words of the British emergentists, the bridge

[23] 1704 A VI.6 131-2/NE II.viii.15.

[24] NE II.viii.13; D.A. ed, p. 131; see also M. Wilson (1999a).

[25] Recent defenses of the PSR include Amijee (2021); Dasgupta (2014); Lin (2012).

laws must be accepted with "natural piety;" they are the ontological danglers of J.J.C. Smart's nightmares.[26] So, there is no reason to be found solely in the basal material states why they result in an experience of red as opposed to any other mental state.

Leibniz also gives an argument based on the Principle of Continuity (PC). In his *New Essays on Human Understanding* PC is presented as, "[n]othing takes place suddenly, and it is one of my great and best confirmed maxims that *nature never makes leaps*."[27] According to the PC, whenever something goes from one state to another, it does so by passing through an intermediate state. By the PC, Leibniz reasons, "For anything which is noticeable must be made up of parts which are not: nothing, whether thought or motion, can come into existence suddenly."[28] According to Leibniz, this means the mind would not be able to proceed from an idealess state, such as a dreamless sleep might be (though Leibniz denied there were ever such states of the mind), to a state with ideas. As such, he hypothesizes *petites perceptions*, perceptions of which we are unaware, but bring about conscious perceptions through aggregation. These *petites perceptions* are the implicit structure of our conscious perceptions. The structure is "implicit" because it is not accessible through introspection.

The core premise of Leibniz's second argument is the PC. An unmodified form of the PC would be controversial today. However, I think many philosophers, especially many panpsychists as well as reductive physicalists, would accept an amended form of the PC on empirical grounds (a continuity of explanation, to be discussed below) and is an expression of their rejection of strong emergentism. I will utilize this amended form of the PC in my argument for confused ideas.

For my purposes, I need not adopt the entirety of Leibniz's views on confused ideas. I merely wish to appropriate the core of his notion of "confused." Recall, an idea is confused when it has a structure or content not accessible through introspection, often because the idea is represented as simple. Note that actually simple ideas are not confused on this account. My own argument for confused ideas draws on a modified form of the PC, "PC2," based on the nature and history of successful scientific explanations.

[26] J.J.C. Smart, "Sensations and Brain Processes," *Philosophical Review* 68 (1959).
[27] A VI.4 56/NE Preface, emphasis in original.
[28] A VI.4 117/NE II.i.18.

PC2: there is an upward continuity of explanation in the natural world.

If one thinks the mind is constituted by the brain but is not a strongly emergent property or substance, then one should accept confused ideas. This is because those who accept that the mind is constituted by the brain and reject strong emergentism, i.e., constitutivists, typically do so for reasons similar to the PC2. Further, if PC2 and constitutivism are true, then, I will argue, our ideas must be confused.

What is strong emergentism and why should one reject it? The British emergentists thought reality was layered. That is, there they conceptualized the world as having different "levels," with each level being governed by its own laws. Higher levels of the world, on this view, are related to lower levels by bridge laws. Because chemical bonds were inexplicable in their time, it seemed to be a brute fact that, for example, sodium and chlorine could make table salt. Such phenomena were accepted with "natural piety" and thought to happen between every level of the world. Thus, chemistry emerged from physics, biology from chemistry, and psychology from biology. Historically, one way of making the case for emergentism was itself a case based on continuity, though it was a continuity of discontinuity. As Crane phrases the emergentist's case, "[w]hat we find when we look at levels of nature are discontinuities and downwards causationa."[29] At each "level" of reality, the need for brute laws of emergence was needed to bridge that level to higher and lower ones.

Today, the tables have turned on emergentism. Brute laws are no longer invoked to relate levels. Configurational forces, hypothesized higher-level forces that only arise when elements were in the right arrangement or configuration are not invoked to explain chemistry just as vital forces are not invoked to explain biology. This was not because the emergentist research program was incoherent, but empirically false.[30] The last bastion for strong emergence today is the mind. However, we have strong inductive reason to think that emergence with respect to the mind is just as mistaken as emergence with respect to chemistry or biology. Scientific progress has not revealed a world with layers, instead, there is a remarkable continuity of explanation

[29]Tim Crane, "The Significance of Emergence" in *Physicalism and its Discontents*, eds. Carl Gillett and Barry Loewer (Cambridge: Cambridge University Press, 2001), p. 222.

[30]Brian McLaughlin, "The Rise and Fall of British Emergentism" in *Emergence: Contemporary Readings in Philosophy and Science*, eds. Mark A Bedau and Paul Humphreys (Cambridge: MIT Press, 2008).

from physics to chemistry and chemistry to biology.[31] This is not to say it has all been worked out, rather it is to look at the big picture and make an inference at what future explanations may reveal. As Luke Roelofs puts it, "I strongly suspect that whatever gaps currently exist in our scientific reductions — of neuroscience to biochemistry, of biochemistry to particle physics, etc. — reflect our own current limitations, not facts about reality."[32] Rejecting emergence for this reason, that is, assuming mentality is dependent on brain structures just as chemical bonds are dependent on atomic structures, is to be committed to PC2. Assuming that gaps in our understanding of the world reflect gaps in our knowledge and not discontinuities in nature is to be committed to PC2.

What inference can we draw from PC2? If mental states are constituted by complex brain states, then mental states cannot be simple. The brain states that constitute mental states are not simple; they are complex networks of neurons firing in complex patterns. There is complexity at the level of the neurons, but the neurons are also complexes of different parts (which are, in turn, complexes of various chemicals which are, in turn, complexes of atoms and so on). The continuity of explanation underlying the rejection of emergentism is a continuity of complex states producing further complex states as we move "up" in the world.[33] If mental states were simple, then this would be evidence for the emergenist's view. But, if we reject strong emergence, then we must accept an extension of the continuity we find in other domains of scientific study. Thus, we should expect that our mental states are complex as well. That is, if continuity of explanation is a principled reason to reject strong emergence, as the history of scientific progress seems to show, then we should reject simplicity as a model of phenomenal properties. If phenomenal properties are not simple as they appear in introspection, then they have an implicit structure. That is, the confused ideas thesis is true.

Yet it is a widespread assumption among philosophers of mind that the phenomenal properties we experience are simple. If our mental states are as they appear in introspection and introspection does not reveal an internal complexity of some mental states, then they must be

[31] Cf. Wilson (2021) for a contemporary take of strong emergence.

[32] Luke Roelofs, "Is Panpsychism at Odds with Science?" *Journal of Consciousness Studies* 28 (2021): p. 120.

[33] This phrasing of the position is meant to allow for the world to be infinitely decomposable, where everything has proper parts. If the world is ultimately made up of simples, then we can say that all complexity is ultimately resolvable into simples.

simple. For example, the sensation of red, the painfulness after stubbing a toe, or the amateur drinker's experience of a gin and tonic. Accepting that the appearance of simplicity with respect to our mental states implies they really are simple is an error for views that accept that the mind is identical to, realized by, or grounded in the brain. Especially in the domain of panpsychism, where it has led, I think, to the many forms of the combination problem but also leading to problems for non-panpsychist views, such as the structural mismatch problem, as discussed above. However, this is the truth behind what I take to be the illusionist's strongest critique of realism about phenomenal properties, "the basic illusionist claim that introspection delivers a partial, distorted view of our experiences, misrepresenting complex physical features as simple phenomenal ones."[34] Where the illusionist goes wrong is the further claim that because phenomenal properties are represented by introspection as simple despite the basal physical state being complex, there must be no phenomenal properties as they are qualitatively represented in introspection. Instead, they should infer, as I am suggesting, following Leibniz, that phenomenal properties are not as structurally simple as they appear.

5.3 How confusion saves constitutivism

Armed with Leibniz's confused ideas thesis, the constitutivist can now avoid the structural mismatch problem. Recall the general form of the structural mismatch problem:

5. Phenomenal properties seem or appear to be simple. (phenomenal simplicity)

6. Introspection presents us with an accurate representation of phenomenal properties. (introspective reliability)

7. Brain states are complex. (material complexity)

8. The mind is grounded in, realized by, or identical to the brain. (constitutivism)

If introspection presents phenomenal properties to us in a confused way, i.e., it represents them as being simple when they are actually complex, then we need not accept an unqualified form

[34]Frankish, "Illusionism," p. 18.

of 6 (as the revelation theorist does) nor deny it outright (as the illusionist does). Rather, the constitutivist can accept a qualified form of 6 which avoids the structural mismatch problem:

9. Introspection presents us with an accurate representation of the qualitative character of phenomenal properties, but not their structure. (confused ideas thesis)

If introspection does not give us access to the implicit structure of phenomenal properties, then it is possible that there is no mismatch between the implicit structure of phenomenal consciousness and the structure of brain states which constitute those phenomenal properties. The confusion defense does not provide a positive case for the implicit structure of phenomenal consciousness being isomorphic to neural structure, but it does block the move from the appearance of structural mismatch to there being structural mismatch. Is the implicit structure of phenomenal experience isomorphic to brain structures? This might not be something we can know. If Leibniz was right about some ideas being necessarily confused, then the deep structure of mental states is not something that can be discovered through introspection. Even if we cannot introspect that the structure of phenomenal consciousness is isomorphic to brain structure, there might be a cumulative case or inference to the best explanation argument for thinking that it is. This cumulative case will look different depending on which constitutive view is being defended. But, a commonality between the cases will probably rely on something like PC2.

5.4 Alternatives to confusion

In this section, I will evaluate some alternatives to the confused ideas thesis available to the constitutivist. As I am assuming the truth of constitutivism for this chapter, I will not consider non-constitutivist alternatives. There are four ways for the constitutivist to avoid the structural mismatch problem; i) show the structure of phenomenal properties are not simple as they appear to be, ii) show how complex (material) states can give rise to simple (phenomenal) states, iii) show that the structure of underlying material states are not complex as they appear, and iv) deny the reality of phenomenal states. The confused ideas thesis a version of option (i), defended above. So, the remainder of this section will deal with the remaining options.

Options (ii) and (iii) are varieties of emergence. (ii) is the traditional kind of emergence, where a simple mental substance or properties emerges from complex basal material states. (iii) is the kind of emergence defended by fusionists, like Mørch (2014), where the brain is an extended simple emergent from more fundamental material objects.

In addition to the argument against emergence based on a continuity principle (section 5.2), I will give a further reason to avoid emergence: what Goff calls "the threat of noumenalism" or the threat "that human beings are cognitively closed to the deep nature of fundamental reality" where a mind is cognitively closed when its "concept-forming procedures ... cannot extend to a grasp of" some theory or concept.[35] So, the threat of noumenalism is the threat that we are not able to intellectually grasp, conceptualize, or understand the deep nature of fundamental reality. Goff levies this worry against panprotopsychist views, where the deep nature of fundamental physical role-playing properties are not themselves mental but gives rise to the mental. Because we cannot form an understanding of such properties, this view leaves the fundamental nature of the world utterly mysterious to us. In a similar way, I think, strong emergence entails that we are cognitively closed to understanding the nature of fundamental entities at every level of the world. It might sound odd to say "fundamental entities at every level" as, when there is talk of fundamental entities, there is usually only exist at one such level. Traditionally, discussions of fundamentality distinguish between independent fundamental entities (which depend on nothing else for their existence) and dependent entities, whose existence depends on the fundamental entities. But, according to Elizabeth Barnes, "ontological emergence is the claim that some things which are fundamental are not ontologically independent."[36] Emergence does not, I think, entail the threat of noumenalism at the fundamental and independent level of their view, that is, the basal material states emergent entities are dependent upon. However, it is with emergent entities, fundamental and dependent, where the threat of noumenalism arises. If emergent entities involve brute and inexplicable psycho-material laws, then it is part of their nature to be butre and inexplicable. As such, just as these laws would need to be accepted with natural piety and, therefore, noumenalized, the entities themselves would be noumenal as well.

(iv) is the illusionist option. There can be no mismatch between an apparently simple

[35] Goff, *Fundamental Reality*, p. 168.

[36] Elizabeth Barnes, "Emergence and Fundamentality," *Mind* 121 (2012): p. 882.

phenomenal state and a complex physical state if there is no phenomenal state. For reasons to reject the core argument of the illusionist, see chapter 3.

5.5 Objections

In the final section of this chapter, I will field objections to the confused ideas thesis. The first objection has to do with the appearance-reality distinction. The external world might seem to be one way when it is in fact another way, e.g., the Müller-Lyer illusion seems to be two line segments of differing lengths when they are in fact the same length. Things in the world might differ from how they appear, but we cannot be mistaken about the appearances themselves. That is, when I am looking at the desk in front of me, I might be mistaken in thinking that the desk is brown, but I cannot be mistaken that it appears brown to me. The strongest objection to the confused ideas thesis, I think, is that it implies that there is an appearance-reality distinction with respect to our mental states.[37] A denial of the appearance-reality distinction is necessary for the confused ideas thesis, as how the structure of our mental states appears to us may be different from how the structure really is. But, this opens the door to an appearance-reality distinction with respect to all aspects of our mental life. Because almost all of our knowledge of our mental states is wholly dependent on how they appear to us in introspection, the confused ideas thesis seems to threaten all knowledge gained through introspection. This would be especially problematic if it held with regard to our experiences of phenomenal properties, that is, if the confused ideas thesis implied that when we think we are in some phenomenal state, like feeling pain or seeing red, we are in reality either in a different phenomenal state or no phenomenal states whatsoever (the latter being the illusionist's view).

This objection is rather important, so I want to address it in parts. First I will argue that the confused ideas thesis entailing an appearance-reality distinction with regard to mental structure is not problematic. Then that the confused ideas thesis does not entail an appearance-reality distinction with respect to the qualitative character of our mental states. Next, I will address whether the mere consistency of the confused ideas thesis with an appearance-reality distinction is problematic. I conclude this section with some remarks about how the appearance-reality

[37] I'm grateful to Howard Robinson for pushing me on this point in conversation.

distinction and denying it affects our knowledge.

I do not think we should be troubled about the confused ideas thesis entailing an appearance-reality distinction with respect to the structure of our mental states. First, intuitions about the transparency of the mental seem to be much stronger when about the transparency of qualitative character and weaker when applied to the content or the structure of mental states. Second, there is a long history in both the philosophical and, to varying degrees, empirical study of the mind which presents it as having aspects that are hidden from us or inaccessible to introspection. Consider Leibniz, from whom I borrow the view, Freud, and the very notion of unconscious mental states. Third and finally, something like the confused ideas thesis must be accepted if we are to solve the structural mismatch problem. However unintuitive or uncomfortable the notion of confusion might be, the alternatives, such as strong emergentism, illusionism, or an intractable structural mismatch problem, seem much worse.

I think the confused ideas thesis does entail an appearance-reality distinction for our mental states, but not in a problematic way. The thesis implies a distinction between the appearance and reality of our mental structure. Confusion does not, however, entail an appearance-reality distinction with regard to the qualitative feel of mental states, which would be problematic. Consider again the cases of the amateur and discerning gin and tonic drinkers. The defender of the confused ideas thesis need not say that the amateur drinker is mistaken when they report their experience of the gin and tonic being a homogenous, uniform single flavor and they are really experiencing the diverse, plurality of flavors the distinguished drinker is able to pick out. The defender of the confused ideas thesis, as I am presenting it, accepts that the amateur drinker's experience is qualitatively how she describes it. The amateur drinker's experience is what-it-is-like to experience a gin and tonic to that degree of confusion while the discerning drinker's experience is what-it-is-like to experience a gin and tonic to a lesser degree of confusion (being able to distinctly pick out the botanicals and other flavors, but not able to discern any further). The defender of the confused ideas thesis is only committed to there being an appearance-reality distinction with respect to the structure of our mental states, not the qualitative character. The what-it-is-like aspect of the amateur drinker's experience is what it is like to experience the flavor of a gin and tonic to that degree of confusion.

It might be objected that the cases I have used to motivate the confused ideas thesis actually undermine the view. For example, one might argue that what the gin case shows is that the amateur drinker inaccurately represents the qualitative character of her gin experience while the discerning drinker accurately represents the experience. It cannot be that both the amateur and discerning drinkers accurately represent the qualitative character of their experiences.

However, I think it is possible they both do represent the phenomenal character accurately. The discerning drinker experiences the drink in a way that is less confused than the amateur drinker's and it seems plausible that the way we represent the structure of an experience would affect its phenomenal character. When the amateur drinker infers from the appearance of phenomenal simplicity that it is phenomenally simple, that is, she cannot detect a complex phenomenology when introspecting on her experience, so she infers there is not one. I am attempting to draw a subtle distinction, so I will restate it differently. The amateur drinker is not mistaken about how she experiences the flavor of the gin and tonic, as that is what-it-is-like to experience those botanicals, quinine, and lime together to that degree of confusion; she is only mistaken if she thinks that her experience is structured the way it appears. The defender of the confused ideas thesis denies that the appearance of simplicity with respect to our mental states entails they really are simple. So, in summary, the confused ideas thesis entails only an appearance-reality distinction with respect to the structure of our mental states, but not with respect to the phenomenal, qualitative, what-it-is-like of our mental states. Undoubtedly, the confused ideas thesis is consistent with there being a qualitative appearance-reality distinction, is this mere consistency problematic enough?

A further objection might be the following. In chapter 4, I argued that Leibniz was what we would call today an eliminativist or illusionist about phenomenal properties. My explanation of this relied in part on Leibniz's use of the confused ideas thesis. This raises the question, does confusion entail eliminativism about qualities, as Leibniz thought? If it does, this would be seen by many as a reason to reject my view.

I do not think the confused ideas thesis entails eliminativism. The eliminativism I argued Leibniz held or, at least, was committed to was the result of other commitments he held. For example, Leibniz's uniquely Rationalist commitments about causation. It is worth noting that,

although similar, the problem Leibniz was addressing, why external bodies cause the specific secondary quality ideas they do in us and not others, is distinct from the structural mismatch problem I am considering in this chapter. The confused ideas thesis on its own need not lead to eliminativism, unless one builds into their definition of "mental state" or "phenomenal property" or "qualia" that it must be simple. But this conceptualization of our mental states and phenomenal properties are not concepts worth keeping.

If we accept PC2 as a reason for rejecting emergence, does that commit us to there being a material explanation of consciousness? If PC2 commits us to a materialist explanation of consciousness and one thinks there cannot be a material explanation of consciousness because of the hard problem, the explanatory gap, the knowledge argument, or some other reason; then that would be a reason for some to reject PC2.

PC2 alone does not entail that there is a material explanation of consciousness. Recall, PC2 states that there is an upward continuity of explanation in the natural world. PC2 makes no claim about the material, it makes a claim about the continuity of explanation. PC2 in conjunction with some claim about the material being fundamental would entail that there must be a material explanation of consciousness, but that claim has not been argued for here and I do not accept it. As presented in this chapter, PC2 is compatible with anti-materialist arguments that mathematical physics alone cannot explain consciousness but needs to be supplemented by something like the Russellian view.

5.6 Conclusion

The structural mismatch problem arises from the apparent asymmetry between phenomenal structure and the structure of the plausible material correlates. I have attempted to argue that although our phenomenal experiences seem simple (phenomenal redness does not present itself in experience as having "parts" of the experience) they may not be. That is, the appearance of simplicity does not entail actual simplicity with respect to mental states, this is the heart of the confused ideas thesis. The difficulty becomes explaining how our phenomenal states could be different from how they appear to us. There is a strong intuition among philosophers of mind

that there is no appearance-reality distinction with phenomenal states. I want to preserve that in a limited way. I want to say that the what-it-is-like aspect of the experience is veridical and there is no appearance-reality gap to the qualitative feel of an experience, but the apparent structure may be misleading. The difficulty is that I think the structure and the experience are tightly connected: it is the phenomenal properties standing in certain structural relations that make up higher-level phenomenal properties on my view.

Chapter 6

From Confusion to Panpsychism

In the previous chapter, I defended a group of theories I termed "constitutivist" from the structural mismatch problem. A theory falls under the umbrella of constitutivism if it posits that mental states are constituted by, i.e., grounded in, realized by, or identical to, material states—usually brain states. As in previous chapters, I will follow Jessica Wilson in considering something physical if "it is (approximately accurately) treated by current or future (at the end of inquiry, ideal) physics, and is not fundamentally mental."[1] Something is material if it is approximately accurately characterized by current or future physics regardless of whether it is fundamentally mental. The structural mismatch problem arises due to the apparent discrepancy between the seemingly simple structure of mental states and the complex structure of their constituting material states. I defended constitutivist theories by arguing for the confused ideas thesis, that some of our ideas have an implicit structure that is not accessible to introspection.[2] I will not repeat the argument for the confused ideas thesis in this chapter. Instead, I will rely on the intuitiveness of two cases, the cog case and the gin case. If some of our ideas are confused, then a mental state seeming simple in introspection does not entail that the mental state is simple. The confused ideas thesis leaves an open possibility that mental states have a complex internal structure and this internal structure might match the constituting material structure.

In this chapter, I will argue that if non-strongly emergentist constitutivism is true and we

[1] Jessica Wilson, "Supervenience-Based Formulations of Physicalism," *Noûs* 39 (2005): p. 428.

[2] The version of the confused ideas thesis I defend is a modified form of the view defended by the great philosopher G.W. Leibniz, hence the somewhat archaic, early modern use of "idea" to broadly refer to many different mental states and not just to thoughts.

have confused ideas, then the disjunction of panpsychism or panprotopsychism, abbreviated henceforth as "pan(proto)psychism," is true. Otherwise, the structural mismatch problem will be inescapable and constitutivists will either have to accept a brute relation between mental states and material states or deny constitutivism. While some philosophers have argued from pan(proto)psychism to the confused ideas thesis,[3] I believe my argument is novel in being the first attempt to argue from confusion to pan(proto)psychism.

6.1 From confusion to pan(proto)psychism

In this chapter, I will assume the truth of constitutivism and the confused ideas thesis, that our ideas have a complex structure that is not accessible to introspection. To understand the motivations behind the confused ideas thesis, consider again the following two cases. First, the cog case from G.W. Leibniz's *New Essays on Human Understanding*,

> Somewhat similarly, when the swift rotation of a cog-wheel makes us perceive an artificial transparency, as I have noticed on visits to clock-makers, we are not able to discern the idea of the cause of this, i.e., the idea of the teeth of the wheel. The wheel's rotation makes the teeth disappear and an imaginary continuous transparent [ring] appear in their place; it is made up of successive appearances of teeth and of gaps between them, but in such rapid succession that our imagination cannot distinguish them.[4]

Here Leibniz assumes that the cog's teeth contribute just as much to the observer's perception when moving as they do when stationary. If so, the individual teeth must be present in the experience of the spinning cog in some way despite the apparently simple or undifferentiated (denoted in the text by "continuous" and "transparent") nature of the homogeneous blur experience. While Leibniz's cog case helps to motivate the confused ideas thesis, the thought experiment has a lot of moving parts, both literally and figuratively. The thing perceived, the cog, undergoes a change from stationary to moving which complicates the case.

I proposed a more simple case in the previous chapter, the gin case. Contrast two imbibers of gin and tonics, the discerning drinker and the amateur drinker. Both are presented with the

[3] Luke Roelofs, *Combining Minds: How to Think about Composite Subjectivity* (New York: Oxford University Press, 2019).

[4] A VI.6 403/NE IV.vi.7.

same gin and tonic. The discerning drinker has spent many years drinking many gin and tonics, cultivating the introspective ability to pick out the various flavors present, e.g., the specific botanicals used in distilling the gin, the bitterness of the quinine in the tonic water, and the acidity of the lime. In contrast, the amateur drinker, who has never before had a gin and tonic, might describe the taste after introspecting as "it *just* tastes like gin" in the way that someone might say a dish "tastes like chicken." The taste of the gin and tonic to them is simple, it lacks any complexity. However, the botanicals, quinine, and lime must be contributing just as much to the amateur drinker's experience, who does not notice these flavors, as they do to the discerning drinker's experience who does. After all, why would they not? Further, after tasting a few, the amateur drinker might develop the capacity, like the discerning drinker, to individuate the flavors in introspection. So, the botanicals, quinine, and lime must be present in the amateur drinker's experience despite the apparently simple ("it *just* tastes like gin") nature of the experience.

Neither the cog case nor the gin case is meant to be irrefutable proof of the confused ideas thesis. Rather, they are meant to motivate the notion that the appearance of simplicity in a phenomenal state does not entail the state actually being simple.[5]

The structural mismatch problem arises from the apparent asymmetry between the structure of a mental state and the structure of the plausible material correlates of that mental state. Take, for example, the phenomenal experience of an imagined uniform red field and the plausible material correlates of that experience. The red field appears structurally simple, it has no parts or aspects. Granted, the red seems to occupy a two dimensional plane, so there is some spatial structure to the experience. While we might differentiate between the left and right half of the red field, this seems to be a merely conceptual distinction, as the red seems continuous and homogeneous across the field, unitary and not as if it is "made up" of, say, many smaller red spaces. In contrast, we know that the plausible material correlates are immensely complex networks of neurons firing with neurons being composites of other complex material phenomena. The confused ideas thesis resolves this apparent structural discrepancy by postulating that the seemingly simple structure of the red field hides an implicit complex structure.

[5]See chapter 5 for the argument for the view.

Hypothesizing an implicit complex structure raises the question of just what the nature of this implicit structure is? The implicit structure could be either a mental or non-mental structure. Implicit mental structures are structures of other mental entities, such as phenomenal properties. For example, a seemingly simple experience of a phenomenally red field could actually be a complex or aggregate of more basic phenomenal properties. These more basic phenomenal properties i) are presented together to the subject in such a way that the subject cannot introspectively pick them out individually, i.e., are confused, and ii) this confusion results in the phenomenal feel or what-it-is-like of the confused experience. Note that, on this view, the implicit phenomenal properties need not be qualitatively different from the subject's confused phenomenal experience and need not be diverse among themselves. That is, it need not be like the gin case, where, plausibly, the botanicals, the quinine, and the lime are diverse phenomenal properties that make up the implicit structure of the amateur gin drinker's experience and, plausibly, none of them can solely be identified as being the feel or flavor the amateur drinker's experience. The case of the red field, for example, could be the confused experience of an aggregate of implicit phenomenal red. "Implicit non-mental structures" are an aggregate of non-mental properties, parts, or aspects.

Either implicit mental structure or implicit non-mental structure could be satisfied by variations other than what I will consider below, but they are variations a non-emergentist constitutivist cannot accept. For example, the implicit structure might be mental or non-mental, but the what-it-is-like of the confused experience might not be a result of the implicit structure alone, but emergent from that implicit structure. Because I am assuming what has been established in chapter 5, which included a rejection of emergentism, I will not consider such emergentist variations.

Given that the confused ideas thesis is neutral between the implicit structure being mental or non-mental, which should we prefer theoretically? Consider some ordinary confused mental state of some human subject, m_O. The implicit structure of m_O is either mental or non-mental. If it is non-mental, then it must be material.[6] Assume it is a material structure. If it is a material structure, then it is either narrowly or broadly material.

[6] As in chapter 2, I assume here that the mental-physical distinction is exhaustive.

Narrowly material properties are the structural and dynamical properties studied by the sciences. Could the structure be narrowly material? Not if the structure and dynamics argument against materialism succeeds.[7] Torin Alter extracts the following argument from Chalmers:

1. All physical truths are purely structural.

2. From purely structural truths, one can deduce only further purely structural truths.

3. Some truths about consciousness are not purely structural.

4. Therefore, there are truths about consciousness that cannot be deduced from (i.e., are not *a priori* entailed by) the complete physical truth.[8]

Premise 1 is supported by structuralism about physics, the view that physics only describes the structural and dynamical properties of the world, such as the relational, spatiotemporal, nomic, or functional properties and changes in those properties over time. [9] Chalmers supports premise 2 with an appeal to the history of scientific explanation,

> What can be inferred from this sort of description in terms of structure and dynamics? A low-level microphysical description can entail all sorts of surprising and interesting macroscopic properties, as with the emergence of chemistry from physics, of biology from chemistry, or more generally of complex emergent behaviors in complex systems theory. But in all these cases, the complex properties that are entailed are nevertheless structural and dynamic: they describe complex spatiotemporal structures and complex dynamic patterns of behavior over those structures. So these cases support the general principle that from structure and dynamics, one can infer only structure and dynamics.[10]

Premise 3 is supported by introspecting on our conscious states. For example, the nature of feeling pain cannot be fully explained by appealing only to the causal, relational, or nomic roles it plays or plays a part in. A description of pain limited to these features would leave out its phenomenal nature. What-it-is-like to feel pain has an intrinsic, experiential component to it.

[7] See Alter (2016), Chalmers (1996; 2002).

[8] Torin Alter, "The Structure and Dynamics Argument against Materialism," *Noûs* 50 (2015): p. 795.

[9] See section 2.1 for more on structuralism about physics.

[10] David J. Chalmers, "Consciousness and Its Place in Nature" in his *Philosophy of Mind: Classical and Contemporary Readings* (New York: Oxford University Press, 2002), p. 259.

Important for our purposes is that Alter's above presentation of the structure and dynamics argument attempts to establish an epistemic gap between the structural facts and the facts about consciousness. If the facts about consciousness cannot be *a priori* deduced from the structural facts, then the facts about consciousness are not necessitated by the structural facts. If the structural facts do not necessitate the facts about consciousness, then the implicit structure cannot be narrowly material because the narrowly material facts are, by definition, merely the structural and dynamical facts.

So, the implicit structure of m_O must not be narrowly material. Broadly material properties include all narrowly material properties, but also whatever serves as the categorical basis for narrowly material properties. Whatever this categorical basis is, it must be sufficient to explain the confused mental state it is the implicit structure of without being mental itself. This is simply panprotopsychist Russellian monism, the view that the categorical basis for the entities studied by fundamental physics are not mental but aggregate to explain mentality.

If, on the other hand, we assume m_O has an implicit mental structure, then we can inquire about the structure of those implicit mental states. The mental states that make up m_O's implicit mental structure are either simple or complex. If the implicit mental states are complex, then we can ask the same questions about their structure. Mental or material? Simple or complex? Thus, a complex will resolve into either simple mental states or broadly material properties. If the implicit mental states are simple, then they are either constituted by simple material states or the structural mismatch problem has merely been moved to a lower level. I take it that if the structural mismatch problem persists at any level, then constitutivism is false. This is because were constitutivism true and there were a structural mismatch between some mental state and its material basis, then there would be no explanation for the mismatched structure.[11] So, because constitutivism has been assumed, we can infer that the structural mismatch problem is not merely pushed to a lower level and the implicit mental structure, if simple, is ultimately constituted by simple material states. If m_O resolves into simple mental states constituted by simple material states, then panpsychist Russellian monism is true. Presented as an argument,

1. Non-emergent constitutivism is true.

[11] Recall in chapter 5, I argued for a principle of continuity, PC2, as a common motivation for constitutivist theorizing. If there is a higher-level phenomenon that resists lower-level explanation, then constitutivism is false.

2. The implicit structure of our ordinary mental states (m_O) is either ultimately mental or purely material.

3. If the implicit structure is ultimately material, then it is either broadly or narrowly material.

4. It is not narrowly material.

5. If the implicit structure is ultimately broadly material, then panprotopsychism is true.

6. If the implicit structure is ultimately mental, then the mental states that make up that structure are either simple or complex.

7. If the implicit mental structure is complex, then it either resolves into simple mental structures or into broadly material properties.

8. If the implicit mental structure is simple, then either the structural mismatch problem persists or panpsychism is true.

9. If constitutivism is true, then there is no structural mismatch problem.

10. If constitutivism is true, then either panpsychism or panprotopsychism is true.

Premise 1 is assumed for this paper. Premises 2 and 3 are each exhaustive of the possibilities. Panprotopsychism posits the existence of broadly material properties that are not mental but explain mentality, premise 5. The consequent of premise 6 is exhaustive of the possibilities for an implicit mental structure. The consequent of premise 8 is exhaustive of the possibilities. Simple material states being realized by mental states is panpsychism or near enough. Premise 9 is true on the assumption that mental states are made intelligible by their constituents. That is, if some Laplaceian archangel had complete knowledge of the constituents and the laws of nature, then she could infer the existence of the mental state. The argument presupposes the confused ideas thesis in the appeal to mental states having an implicit mental structure.

Two premises, 4 and 7, deserve further discussion. I will discuss 7 in the next section. Premise 4 follows from the definition of narrowly material properties and the success of the structure and dynamics argument. The alternative is to accept that the implicit structure of

mental states is made up of narrowly material properties and reject that mental states have any properties which could not be fully explained by narrowly material properties.

However, if the implicit structure is narrowly material and lacks any implicit phenomenal or protophenomenal structure, then no progress is made on the structural mismatch problem. The implicit non-mental structure view seems to merely restate the issue we originally set out to solve, that is, that there is a complex material structure (which is assumed on most views to be non-phenomenal) and a seemingly simple phenomenal structure. Of course, identifying the structure of phenomenal experience with the structure of the plausible narrowly material correlates (the obvious candidate for the non-mental structure) dissolves the structural mismatch problem, but it does so in a way that does not make the relationship any more intelligible than it was before. How could this immensely complex narrowly material structure make up a seemingly simple phenomenal experience? There must be more to the full story in order to render the appearance of structural mismatch intelligible.

A defender of the narrowly material structure alternative might respond to the above objection by arguing that the constitutivist has, at least, reduced the number of problems the view faces. Whereas before the constitutivist faced the explanatory gap, hard problem, and structural mismatch; now only the explanatory gap and hard problem remain. While this is true, I think leaving such a relationship unintelligible is an unacceptable cost that should be avoided.

Any attempt to make this structural relationship intelligible on the narrowly material structure view would require accepting either functionalism, illusionism (a sub-category of functionalism), or strong emergence.

A standard form of functionalism, which does not deny that phenomenal properties have the qualitative character introspection represents them as having, cannot make this relationship intelligible. This is because functionalism identifies mental states solely by the functional role they play and not by their composition or phenomenal character. Because of this, the functionalist is unable to explain why mental states have their distinct phenomenal characters.[12] If the functionalist cannot account for the phenomenal character of a mental state, then she cannot explain that state's qualitative structure either. Thus, if the implicit structure is narrowly ma-

[12]See Saul Kripke (1980) and Chalmers (1996).

terial, then the most natural form of the view seems to be a kind of illusionism. After all, this seems to be what Keith Frankish expresses in his defense of illusionism,

> the basic illusionist claim [is] that introspection delivers a partial, distorted view of our experiences, misrepresenting complex physical features as simple phenomenal ones.[13]

Further, it appears similar to the illusionist view that I attributed to Leibniz in chapter 4. Illusionism renders the relationship between the complex narrowly material structure and the apparently simply phenomenal structure intelligible by denying the reality of the phenomenal property's qualitative character. If mental states lack their apparently simple phenomenal character, then there is only the narrowly material structure to account for. Any appearance of a structural mismatch is merely an artefact of our distorted introspection.

I presented several reasons to reject illusionism in chapter 3. If I am correct that illusionism should be rejected, then this is also a reason to reject that the implicit structure is narrowly material.

Alternatively, if the implicit structure is narrowly material, then simple phenomenal properties arising from complex material properties seems to be a form of strong emergence, that is, the phenomenal feel of m_O is an emergent feature of the non-phenomenal structure. The phenomenal property must be emergent because the property has a what-it-is-like aspect that, by hypothesis, non-mental properties and structures do not. If the non-phenomenal structure had a what-it-is-like aspect, then in what sense is it narrowly material?[14] I provided an argument against emergence in chapter 5.

I have argued that the narrowly material implicit structure view is unacceptable because it makes no progress on the structural mismatch problem of explaining how complex material states relate to seemingly simple phenomenal ones. In order to make this relation intelligible, the defender of the view would have to commit herself to either illusionism or emergentism, two unacceptable views.

That the confused ideas thesis in conjunction with non-emergent constitutivism and realism about phenomenal properties entails pan(proto)psychism does not tell us very much about the

[13] Keith Frankish, "Illusionism as a Theory of Consciousness," *Journal of Consciousness Studies* 23 (2016): p. 18.

[14] The panprotopsychist option is not available here because such a view is not narrowly material.

resulting pan(proto)psychist view. This conclusion is neutral between various formulations of pan(proto)psychism; even varieties of idealism are compatible with the conclusion depending on which analysis of the material is accepted. One way these details might be filled in is with the Russellian monist picture, which I favor. On one version of this view, phenomenal properties are the categorical basis for fundamental material role-playing properties.[15] Though, of course, the non-emergentist commitments I have argued for rule out emergentist variations of Russellian monism.

6.2 The infinite decomposition alternative

Premise 7 was conspicuously ignored in the previous section and deserves further discussion. It might be objected that the premise neglects the possible alternative of infinitely decomposable states. Some state is infinitely decomposable if all of its proper parts have proper parts. On this view, priority monism, there are no simples nor is there a fundamental level of the world. Priority monist views have received prominent defenses in recent years by both panpsychists and non-panpsychists.[16]

I think the argument can be applied to the infinite decomposition alternative. If the world is infinitely decompossible, then either it infinitely decomposes into mental states or the mental states infinitely decompose into material states at some level or the world infinitely decomposes into alternating mental and material states. If the world infinitely decomposes into mental states, then something very much like traditional panpsychism is true. If the world eventually decomposes into infinite material states, then those states will either be broadly or narrowly material. As shown above, they cannot be narrowly material. If they are broadly material, then something very much like the standard panprotopsychist view is true. A third possibility is that the world infinitely decomposes into alternating material and mental states. Mental states decompose into material states decompose into mental states and so on. The material states must be broadly and not narrowly material for the reasons given above. Such a world seems to

[15] Cf. Nagasawa (2012) and Goff (2017a) for priority monist formulations.

[16] Examples of panpsychist priority monists include Goff (2017a) and Nagasawa (2012). Schaffer (2003; 2009; 2010) is a non-panpsychist priority monist.

be pan(proto)psychist. So, even though the argument does not overtly consider the possibility of infinite decomposition, the conclusion of the argument remains broadly unchallenged by it.

The priority monist might further object that a presupposition of the argument, PC2, is incompatible with infinite decomposition. This incompatibility is problematic because I defended PC2 in the previous chapter as a motivation common among constitutivism. Recall that PC2 states,

PC2: there is an upward continuity of explanation in the natural world.

On the priority monist picture, the cosmos as a whole is the basic explanatory entity and other entities are grounded in the cosmos. Yet, the cosmos is not considered a lower-level entity (it would be odd if humans were a higher-level entity than the whole cosmos!) but other entities are explained in virtue of being derivative proper parts of the universe. This seems to violate the upward continuity posited in PC2, if we take "upward" literally.

There are several possible responses to this. First, one could accept the objection and infer that either non-priority monist panpsychism or panprotopsychism is true. I think this is too hasty, as I think the objection can be avoided.

Second, the literal upward continuity can be preserved if certain dependence relations hold between the cosmos and non-basic entities. Let "x $\rightarrow$ y" stand for "x constitutes y" and "..." be a placeholder for intermediate entities or constituting relations, if there are any. If the dependence relation is something like $\{$the cosmos $\rightarrow \ldots \rightarrow$ particles $\rightarrow \ldots \rightarrow$ organisms$\}$ as opposed to something like $\{$the cosmos $\rightarrow \ldots \rightarrow$ organisms $\rightarrow \ldots \rightarrow$ particles$\}$, then the, literal, upward continuity is preserved except for in the first step from the cosmos to whatever is immediately dependent on the cosmos. Even this one step, though, might be too much for the literal interpretation of the PC2.

Third, reject the literal interpretation of "upwards." I did not intend the "upwards" in PC2 to be taken literally. Rather, it was used as a shorthand with the unfortunate implication of a fundamental bottom level. PC2 merely requires what Leibniz himself meant when he stated his formulation of the continuity principle in the *New Essays on Human Understanding*, "[n]othing takes place suddenly, and it is one of my great and best-confirmed maxims that nature never

makes leaps."[17] We can amend PC2 to make it more specific and not rule out priority monist views,

> PC2*: There is a continuity of explanation from more basic entities to the entities that depend on them such that the dependent entities can be explained by the more basic entities they depend on.

Here, "entities" is meant to be inclusive of objects, properties, or states of affairs. PC2* serves the same role that PC2 does, with the benefit of not ruling out priority monism but at the cost of catchiness.

6.3 Emergence and illusionism reconsidered

In the previous chapter and section, one of the reasons presented for the constitutivist to accept the confused ideas thesis is that the alternative is either emergence or illusionism. I argued that these alternatives are unacceptable.[18] However, if the conjunction of the confused ideas thesis, non-emergent constitutivism, and realism about phenomenal properties entails pan(proto)psychism, some (probably many) philosophers would take this to be a *reductio* of a premise or presupposition, an indication that a wrong turn was taken somewhere along the journey. In this section, I will consider how pan(proto)psychism fares with respect to the concerns I raised against emergence in chapter 4 and an objection on behalf of the illusionist. I will argue that the pan(proto)psychist commitments of the confused ideas thesis are theoretically more attractive than the emergentist and illusionist alternatives. However, I will argue that panpsychism fares better than panprotopsychism.

Of course, if pan(proto)psychism fares better than emergence and illusionism on the criteria I have considered, defenders of emergence and illusionism will charge that the many forms of the combination problem are insurmountable and render pan(proto)psychism a nonstarter. How the confused ideas thesis solves the combination problem will be the focus of chapter 7.[19]

[17] A VI.4 56/NE Preface, emphasis in original.

[18] See chapters 3 and 5.

[19] See also chapter 2, where I provide a response to two forms of the combination problem for materialist forms of panpsychism.

6.3.1 Emergence vs. pan(proto)psychism

The first reason I gave to reject emergentism was the revised principle of continuity, PC2*, the view that there is an upward continuity of explanation in the natural world. Whatever flaws can be claimed against pan(proto)psychism, violating a continuity principle is not one. On the contrary, a guiding motivation for pan(proto)psychism has always been continuity. Leibniz posited that the intrinsic natures of forces were something mind-like because minds were something we were acquainted with in the world and, as such, served better as a candidate than some other kind of property. Leibniz's world is a world of minds on a spectrum of confusion with bare monads being minds so confused they lack any experience, to the dominant monads of humans being minds with a mix of confused and distinct ideas, to God being a mind with only adequate ideas.[20] Thomas Nagel's arguments for panpsychism relies on the stronger claim that, without panpsychism, there would be a discontinuity in nature from the non-mental to the mental that could only be bridged by emergence.[21] Philip Goff argues that panpsychist Russellian monism should be accepted over neutral monist and panprotopsychist forms of Russellian monism because it is a simpler view.[22] The view is simpler because it posits the deep nature of matter, i.e., the categorical basis of material role-playing properties, is continuous with the part of nature we are most familiar with, i.e., mental properties. In contrast, the neutral monist proposes that reality is fundamentally neither mental nor material, but neutral, with the material and mental (at higher levels) arising from this fundamental neutrality. Similarly, the panprotopsychist proposes that the categorical basis of fundamental material role-playing properties is not mental, but aggregates to constitute mentality at higher levels. Neither panpsychism nor panprotopsychism propose a discontinuity of explanation in the world.

The second reason I gave to reject emergence was the threat of noumenalism, that the view entails we are cognitively closed to the deep nature of the world. Does pan(proto)psychism face this problem? Howard Robinson argues along similar lines that the unknowability of the

[20] See section 4.2 for an more on confusion, distinctness, and adequacy.

[21] Thomas Nagel, "Panpsychism" in his *Mortal Questions* (Cambridge: Cambridge University Press, 1979).
Thomas Nagel, *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False* (New York: Oxford University Press, 2012).

[22] Goff, *Fundamental Reality*, pp. 169-70.

mental states of microsubjects, a kind of noumenalism, is a reason to reject the view.[23]

Goff argues, and I am inclined to agree, that panpsychism is preferable to panprotopsychist forms of Russellian monism because it avoids this very issue.[24] That is, the deep nature of all matter is of the same kind as our mental states, not some unknowable protophenomenal property. Even if we cannot know the precise mental states experienced by microsubjects, we know they are of the same kind as our own mental states. If noumenalism is a legitimate concern, then that is reason to prefer the panpsychist conclusion over the panprotopsychist and emergentist contenders.

Even if panpsychism avoids the problem of noumenalism, one might object that the confused ideas thesis itself faces the problem. While the gin case seems to indicate that confused experiences can be made distinct through repeated experiences or more attentive introspection (hence how the discerning drinker is able to pick out the various flavor aspects present in her experience that the amateur drinker is not able to), it is probably beyond our introspective capacities to make some confused experiences distinct. For example, our experience of phenomenal redness seems qualitatively simple, but, because it is a quality of our experiences which are constituted by complex material states, it must not be simple. If it were, then the structural mismatch problem would be intractable. The seeming simplicity of phenomenal red, on my view, is a result of confusion and I do not think it is possible for humans to introspect their way to the distinct experiences which aggregate and are confusedly represented as red. There remains the possibility that red is simple microphenomenal property and our macrophenomenal red experience is of a confused aggregate of simple microphenomenal red experiences. If this possibility obtains, then the belief that "red is simple" is a Gettier case as we correctly and justifiably believe red is simple but do not know this because we represent most of our phenomenal properties as simple even if they are not.

That some experiences are necessarily confused for beings with cognitive capacities like humans should not be surprising, as we are creatures with evolved, fallible capacities. Further, if our experiences are constituted by fundamental, simple experiences or an infinitely decom-

[23]Howard Robinson, *From the Knowledge Argument to Mental Substance: Resurrecting the Mind* (New York: Cambridge University Press, 2016), p. 128.

[24]Goff, *Fundamental Reality*, p. 168.

possible chain, then we should not expect to be able to make those phenomenal properties distinct through introspection. These limitations are explicit in Leibniz's original presentation of the view. Specifically, he thought that ideas of received through a single sense were necessarily confused.[25] However, this seems to commit the defender of the confused ideas thesis to noumenalism about the nature of what constitutes these necessarily (for us) confused mental states. Further, if this is correct and noumenalism is a reason to reject emergence, then it is a reason to reject the confused ideas thesis as well.

However, I do not think this qualifies as noumenalism or, at least, it is not a problematic form of noumenalism. The threat of noumenalism is that fundamental aspects of the world are in-principle unknowable. The panprotopsychist faces the problem of noumenalism because the protophenomenal properties which serve as the categorical basis of fundamental material role-playing properties are unknowable. In contrast, panpsychism avoids this because while we may not know which specific phenomenal properties serve as the categorical basis for fundamental physical role-playing properties (they may be phenomenal properties we have never directly experienced), we do know that they are phenomenal properties and of the same kind as what we have experienced. So, we have some knowledge of them. Similarly, our introspective capacities might not be sufficient to diffuse the confusion of some experiences, but we know that they are phenomenal properties of the same kind that we are familiar with or they are protophenomenal properties. If they are constituted by phenomenal properties, then noumenalism is avoided. If they are constituted by protophenomenal properties, then it is not avoided. This indicates that noumenalism is not a feature of the confused ideas thesis, as the objection suggests, but only arises as a result of conjoining the view to another that does involve noumenalism.

6.3.2 Illusionism vs. pan(proto)psychism

That the confused ideas thesis, constitutivism, and realism about phenomenal properties together entail pan(proto)psychism might also be seen as a reason to reject realism about phenomenal properties and endorse illusionism. While the objections I levied against emergence were broad and appealed to more general principles, my main critique of illusionism was specific to

[25] 1684 A VI.4 586/AG 24; 1702 G / AG 187.

a common form of illusionist argument, the argument from error. As the pan(proto)psychist or defender of the confused ideas thesis does not wield an argument of that form, I cannot evaluate how my view fares against the objection I raised to the illusionist as I did with the emergentist. Instead, I will consider a possible illusionist objection.

Illusionists like Frankish see "anomalous" cases, like the discrepancy between the structure of phenomenal properties and their constituting material states, as a strong motivation for illusionism.[26] If illusionism is endorsed, then the apparent mismatch between the structure of phenomenal states and the structure of the constituting material states raises no problem as the phenomenal states are illusory.

I have three points to make in response. First, I argued in chapter 3 that the illusionist's arguments have little dialectical force. Second, it seems to me that the denial of phenomenal properties must be considered more radically counter-intuitive than the conclusion of panpsychism. If not more counter-intuitive, then it must be at least equally unintuitive that there is something-it-is-like to be an electron as that there is nothing it is like to feel pain. Third and finally, illusionism only resolves the structural mismatch problem for phenomenal properties. Illusionism alone does nothing to resolve the structural mismatch problem for other mental states, such as propositional thoughts or desires. My discussion of the structural mismatch problem has focused on phenomenal properties, but the problem can also be raised by noting the apparent structural discrepancy between a thought, such as the belief that "the cat is on the mat," and the constituting physical state.

6.4 Conclusion

In this chapter, I have explored a further consequence of the confused ideas thesis, i.e., pan-(proto)psychism. I have argued that non-emergent constitutivism, an umbrella term for many popular views in the philosophy of mind, in conjunction with the confused ideas thesis and realism about phenomenal properties entails pan(proto)psychism. If I am correct in this chapter and the previous chapter where I argued that it is necessary for the constitutivist to accept the confused ideas thesis in order to avoid the structural mismatch problem, then the conclusion

[26] Frankish, "Illusionism," p. 12.

is significant for many views in the philosophy of mind. Panpsychism and panprotopsychism, rather than being minority views, are actually entailed by many popular views.

Chapter 7

Confusion and the Combination Problems

In chapter 5, I attempted to show how constitutivists (i.e., those who hold that mental states are identical to, grounded in, or realized by material states) might avoid the structural mismatch problem by adopting a version of G.W. Leibniz's confused ideas thesis, the view that some of our ideas (taken broadly) are confused or have a structure that is not accessible to introspection. In chapter 6, I argued that the confused ideas thesis in conjunction with constitutivism entails the disjunction of panpsychism or panprotopsychism, "pan(proto)psychism" for short. However, if constitutivism and the confused ideas thesis entail pan(proto)psychism, then the view is burdened with the many faces of the seemingly insurmountable combination problem. In this chapter, I show how the confused ideas thesis presents a unified response to many forms of the combination problem and fares better than other notable responses. I will not claim that my survey of the combination problems considered here is exhaustive, but I have attempted to respond to forms of the problem most prominently discussed in the panpsychist literature. In general, the confused ideas thesis solves the combination problem by undermining the support for a key premise involving our introspective access to the essential nature of our mental states. Combination problems that rely on such a premise we can call "introspective combination problems" because they rely on the strength of our introspective justification about the structure of our mental states to refute panpsychism.

However, not all combination problems are introspective combination problems, some lack an introspective premise. Consider this combination problem from William James' *The Prin-*

ciples of Psychology,

> no possible number of entities (call them as you like, whether forces, material particles, or mental elements) can sum themselves together. Each remains, in the sum, what it always was; and the sum itself exists only for a bystander who happens to overlook the units and to apprehend the sum as such; or else it exists in the shape of some other effect on an entity external to the sum itself.[1]

James endorses a kind of idealism about aggregates. In his view, aggregates have no formal reality outside the minds of observers who abstract from the heap and treat it as a unified whole. This informs James' formulation of the combination problem, a passage seemingly mandatory to quote in any paper on the topic,

> Where the elemental units are supposed to be feelings, the case is no wise altered. Take a hundred of them, shuffle them and pack them as close together as you can (whatever that may mean); still each remains the same feeling it always was, shut in its own skin, windowless, ignorant of what the other feelings are and mean. There would be a hundred-and-first feeling there, if, when a group or series of such feelings were set up, a consciousness belonging to the group as such should emerge. And this 101st feeling would be a totally new fact; the 100 original feelings might, by a curious physical law, be a signal for its creation, when they came together, but they would have no substantial identity with it, nor with them, and one could never deduce the one from the others, or (in any intelligible sense) say that they evolved it.
>
> Take a sentence of a dozen words, and take twelve men and tell to each one word. Then stand the men in a row or jam them in a bunch, and let each think his word as intently as he will; nowhere will there be a consciousness of the whole sentence. We talk of the "spirit of the age," and the "sentiment of the people," and in various ways we hypostatize "public opinion." But we know this to be symbolic speech, and never dream that the spirit, opinion, sentiment, etc., constitute a consciousness other than, and additional to, that of the several individuals whom the words "age," "people," or "public" denote. The private minds do not agglomerate into a higher compound mind.[2]

David Chalmers extracts the following argument from James,

1. If constitutive panpsychism is true, human consciousness is an aggregate.

[1] William James, *The Principles of Psychology* (London: Dover, 1890), pp. 158–59.
[2] James, *Principles*, 160–61.

2. Aggregates do not objectively exist.

3. Human consciousness objectively exists.

4. Constitutive panpsychism is false.[3]

Rather than relying on introspection for a key premise (other than, perhaps, premise 3), the support for premise 2 relies on James' mereological nihilism. The confused ideas thesis, as I have defended it, has nothing to say about such an argument. It makes claims about the nature of consciousness and introspection, not aggregates nor mereology. Combination problems of this sort, "non-introspective combination problems," will not be considered. While the confused ideas thesis cannot solve all forms of the combination problem, that it provides a unified response to the many introspective combination problems is itself a strength.

In section 7.1, I will survey various forms of introspective combination problems and show how the confused ideas thesis disarms these problems of their dialectical strength. In section 7.2, I will evaluate other responses in the literature and consider their consistency with the confusion view. While the confused ideas thesis is capable of responding to all these problems in isolation, I argue some of the views considered would mutually benefit from the addition of the confused ideas thesis. I conclude, in section 7.3, by considering some objections.

7.1 The introspective combination problems

If our mental states are confused in the way I have tried to defend, then many forms of the combination problem for panpsychism can be avoided. This section will survey six forms of introspective combination problems and show how the panpsychist defender of confusion may avoid them. By giving a unified defense against the many forms of the combination problem, the defender of confused ideas gets much more theoretical bang for her buck.

[3]David Chalmers, "The Combination Problem for Panpsychism" in *Panpsychism: Contemporary Perspectives*, eds. Godehard Brüntrup and Jaskolla Ludwig (New York: Oxford University Press, 2017), p. 186.

7.1.1 The subject-summing problem

The most well-known version of the combination problem is the subject-summing problem.[4] This is the problem of how microsubjects, the kinds of subjects fundamental particles might be, join together to form macrosubjects like humans and other animals. A common way of presenting the problem is:

1. If constitutive panpsychism is true, then microsubjects necessitate macrosubjects.

2. Microsubjects do not necessitate macrosubjects.

3. Constitutive panpsychism is false.

Constitutive panpsychism is the view that macrosubjects are grounded in microsubjects. Because grounds necessitate their grounded entities, premise 1 is true. Premise 2 is supported by what Philip Goff calls the "conceivable isolation of subjects" (CIS), which states that for any group of conscious subjects, it is conceivable that there is no further subject.[5] Chalmers defends premise 2 in the form of a conceivability argument,

4. For any group of subjects [S] (with certain experiences), it is conceivable that those subjects exist (with their experiences) and no other subjects exist.

5. For any group of subjects [S], if it is conceivable that those subjects exist (with their experiences) and no other subjects exist, then this is possible.

6. For any groups of subjects [S] (with certain experiences), it is possible that the subjects in S exist (with their experiences) and no other subjects exist.[6]

4 is supported by cases such as those detailed in the James' quotes above. 5 follows from a general claim about conceivability entailing possibility. While 6 is not 2, 2 follows from 6.

[4]For the subject-summing problem, see Chalmers (2015; 2017), Coleman (2014; 2017), Goff (2006; 2017a; 2017b), Roelofs (2019), and Seager (2017). See also chapter 2 of this dissertation for a further treatment of the subject-summing problem and a response on behalf of a materialist panpsychist.

[5]Philip Goff, *Consciousness and Fundamental Reality* (New York: Oxford University Press, 2017a), p. 174.

[6]Chalmers, "Combination Problem," p. 187.

The intuitions which underlie CIS, premise 4, are intuitions about the transparency of our conscious states. Because introspection is the only way we have to access our conscious states, they are transparent if and only if introspection reveals their full essential nature. The subject-summing argument assumes that introspection has revealed the essential nature of subjecthood.[7] Having a transparent concept of subjecthood is what justifies the move from isolated subjects being conceivable to their being possible. That is, if we have a complete understanding of what it is to be a subject and we can conceive of a collection of microsubjects organized into a central nervous system without there being a macrosubject, then that is because it is possible and subjecthood lacks any features that would permit summing with other subjects.

However, on the confusion picture, this move is unfounded because the confused ideas thesis undermines the transparency view. If the confused ideas thesis is true, then introspection, at best, gives us translucent access to our mental states, revealing part but not all of their essential nature. If introspection is not transparent, then it is unclear whether we have a sufficient understanding of subjecthood to justify that isolated subjects are conceivable in a way that permits the move to possibility. Thus, given our epistemic position, it is an open possibility that subjects could sum.

7.1.2 The conceivability problem

Similar to the conceivability argument against materialism is the conceivability argument against constitutive panpsychism.[8] Here, PP is the conjunction of all micromaterial and microphenomenal truths, i.e., the broadly material facts, and Q is a some macrophenomenal truth,

1. PP&˜Q is conceivable.

2. If PP&˜Q is conceivable, it is metaphysically possible.

3. If PP&˜Q is metaphysically possible, constitutive panpsychism is false.

4. Constitutive panpsychism is false.[9]

[7] Goff, *Fundamental Reality*, pp. 174–75.

[8] For the conceivability argument against materialism, see Chalmers (1996; 2009; 2015) as well as chapter 1 of this dissertation.

[9] This formulation of the argument is from Chalmers (2017, 187). See also Chalmers (2015) and Goff (2006) for more on the conceivability problem for panpsychism.

Premise 1 states the conceivability of a panpsychist zombie world where all the broadly physical facts are the same as in our world and the fundamental physical entities of that world instantiate microphenomenal consciousness, but there is no macroconsciousness. Despite the ubiquity of microphenomenal consciousness in that world, the humans and animals are all philosophical zombies. Premise 2 states that if the panpsychist zombie world is conceivable, then it is metaphysically possible. Premise 3 states that if the panpsychist zombie world is metaphysically possible, then constitutive panpsychism is false.

The constitutive panpsychist who wields traditional conceivability arguments against materialism would be hard-pressed to find a principled way of denying premises 2 and 3 in the above argument due to the structural similarity with the traditional argument. However, the defender of the confused ideas thesis need not worry, as she should deny premise 1.

Recall Chalmers' argument for the CIS presented in subsection 7.1.1 above,

4. For any group of subjects [S] (with certain experiences), it is conceivable that those subjects exist (with their experiences) and no other subjects exist.

5. For any group of subjects [S], if it is conceivable that those subjects exist (with their experiences) and no other subjects exist, then this is possible.

6. For any groups of subjects [S] (with certain experiences), it is possible that the subjects in S exist (with their experiences) and no other subjects exist.[10]

Structurally, the conceivability problem and the argument for CIS are identical. To make it more clear, consider the following general form of the argument,

4* ms&˜MS is conceivable.

5* If ms&˜MS is conceivable, then ms&˜MS is possible.

6* ms&˜MS is possible.

Here, ms is the conjunction of all truths about microsubjects and MS is some arbitrary truth about some macrosubject. Ultimately, I am inclined to think that the subject-summing problem

[10]Chalmers, "Combination Problem," p. 187.

is a special case of the conceivability problem where Q is not some arbitrary macrophenomenal truth but specifically a truth about some macrosubject.

Confusion undermines the introspective support for 1 in the conceivability in the same way it undermined support for 4 in the CIS argument. Premise 1 appears conceivable due to introspecting on our own conscious life which reveals nothing about microexperiences, subjects combining, or anything that would rule out the panpsychist zombie world. However, if introspection does not give us transparent access to the nature of our mental states, then it remains an open possibility that there is something about conscious states that, had we epistemic access to it, would render such worlds inconceivable.

7.1.3 The knowledge problem

Chalmers presents the following formulation of a combination problem inspired by the knowledge argument against physicalism, simultaneously originated by Frank Jackson and Howard Robinson.[11] As in the conceivability problem (subsection 7.1.2), PP is the conjunction of all microphysical and microphenomenal truths and Q is a some macrophenomenal truth,

1. Q is not deducible from PP.

2. If Q is not deducible from PP, Q is not necesitated by PP.

3. If Q is not necessitated by PP, constitutive panpsychism is false.

4. Constitutive panpsychism is false.[12]

Premise 1 states that any macrophenomenal truth is not deducible from the broadly physical truths. Premise 2 notes that if Q is not deducible from PP, then it is not necessitated either as there is no entailment from PP to Q. If PP does not necessitate Q, then constitutive panpsychism is false as macrosubjects are not wholly grounded in microsubjects, premise 3.

[11]Frank Jackson, "Epiphenomenal Qualia," *Philosophical Quarterly* 32 (1982).
Howard Robinson, *Matter and Sense: A Critique of Contemporary Materialism* (New York: Cambridge University Press, 1982).

[12]This formulation of the knowledge problem for panpsychism is taken from Chalmers (2017, 189).

Premise 1 is supported by a something like the Mary thought experiment. Place panpsychist-Mary is her black and white room, give her complete knowledge of the microphysical and microphenomenal facts. Does panpsychist-Mary know what-it-is-like to experience Q? Plausibly, no. So, what-it-is-like to see experience Q is not deducible from the broadly physical facts facts (PP).

Much like the conceivability problem, the panpsychist must be careful that her answer not also defeat the problem for physicalism, undermining the attraction of pan(proto)psychism in the first place. Again, the defender of the confused ideas thesis should reject premise 1. The support for premise 1 is strongest if we have transparent access to our mental states. However, the confused ideas thesis entails that we have, at best, translucent access. This, in conjunction with the fact that, plausibly, we do not have introspective access to any microphenomenal properties, undermines the support for 1 and leaves it an open possibility that Q is deducible from PP.

But how can the confused ideas thesis address the Mary thought experiment? If panpsychist-Mary has complete knowledge of the microphysical and microphenomenal facts and the confused ideas thesis is true, then it is an open possibility that there are aspects of them hidden from us that permit combination into macrophenomenal properties. As such, if panpsychist-Mary really did have complete knowledge of PP, then she would be able to deduce Q, if the open possibility obtains. It seems to us that she would not be able to because she has knowledge of microphenomenal properties that are hidden from us in introspection due to confusion.

7.1.4 The palette problem

Next is the palette problem, the problem of explaining the vast array of macrophenomenal experiences in terms of the limited palette of microphenomenal experiences.

1. If constitutive panpsychism is correct, macrophenomenal qualities are constituted by microphenomenal qualities.

2. If Russellian panpsychism is correct, there are only a few microphenomenal qualities.

3. Macrophenomenal qualities are too diverse to be constituted by microphenomenal qualities.

4. Constitutive panpsychism is false.[13]

Constitutive panpsychism is the view that macrosubjects and their macrophenomenal qualities (the phenomenal qualities experienced by creatures like us, dogs, and other macrosubjects) are grounded in microsubjects and their microphenomenal qualities. Russellian panpsychism is the view that microphenomenal properties are the intrinsic, categorical bases that underlie the fundamental physical properties studied by the sciences. Because there are relatively few fundamental physical properties, e.g., spin, charge, etc., the Russellian constitutive panpsychist must be able to account for the whole rainbow of macrophenomenal experience using the this limited palette. However, it seems that the number of fundamental physical properties is not a large enough palette to account for the whole spectrum of macrophenomenal properties.

Premise 1 is true by definition and premise 2 can only be denied at the cost of rejecting the current scientific view about fundamental physical properties. So, the defender of the confused ideas thesis must find a way to reject 3. Introspection presents us with a confused picture of some of our conscious states. As such, the simple phenomenal character our mental states seem to have is, in fact, a complex of phenomenal properties. For example, the seemingly simple experience when focusing on a plain, white wall to the exclusion of everything else may be a complex of confused phenomenal properties, presented to introspection under a simple mode of presentation. Further, it seems plausible that phenomenal qualities can blend or combine in such a way that they can produce new phenomenal qualities through their interaction. Consider, for example, how smells influence taste. If this is the case, even a limited palette of microphenomenal properties could produce a large spectrum of macrophenomenal properties through both the actual combination or blending of those properties and the way macrosubjects confusedly represent complexes of microphenomenal properties.

[13] Again, this formulation of the argument is found in: Chalmers, "Combination Problem," p. 189. For more on the palette problem, see also Coleman (2017) and Roelofs (2019).

7.1.5 The revelation problem

The revelation argument, of the combination problems considered, most obviously relies on introspection:

1. The nature of consciousness is revealed to us through introspection.
2. If constitutive panpsychism is correct, consciousness is constituted by a vast array of microexperiences.
3. Whatever constitutes consciousness is part of its nature.
4. A vast array of microexperiences is not revealed to us through introspection.
5. Constitutive panpsychism is incorrect.[14]

Premise 1 merely states the revelation thesis. Premise 2 states a consequence of constitutive panpsychism, that our macroconscious states are made up of an aggregate of microconscious states. Premise 4 states what seems to be a fact about introspection. When I introspect on a feeling of pain after stubbing my toe, my macroexperience does not appear to me as a vast array microexperiences constituting that experience. If the nature of consciousness is revealed through introspection and introspection does not reveal a vast array of microexperiences, then constitutive panpsychism is false (premise 5).

Premise 1 seems too strong. Imagine a parallel argument against materialism,

6. The nature of consciousness is revealed to us through introspection.
7. If materialism is correct, consciousness is constituted by material states of the brain.
8. Whatever constitutes consciousness is part of its nature.
9. Material states of the brain are not revealed to us through introspection.
10. Materialism is incorrect.

[14]This version of the argument is taken from Chalmers, "Combination Problem," p. 190.

In this parallel argument, 6 is false because mental states are multiply realizable. It is possible that there could be conscious beings constitutionality very different from life on earth with a different substrate for their conscious experiences, such as silicon-based life. It might be objected that my parallel argument is not analogous to the revelation problem for panpsychism. Premises 7 and 9 should not include "of the brain." However, this version of the argument would still fail as introspection does not reveal material states to us. If consciousness is multiply realizable, then the constitutive panpsychist might be able to avoid the revelation argument.[15]

The defender of confused ideas thesis should reject an unqualified form of 1. If the confused ideas thesis is correct, the complete nature of conscious states, what the defender for the revelation argument needs, is not provided by introspection.

7.1.6 The structural mismatch problem

The response to the structural mismatch argument is probably the most important result of the confusion view. The structural mismatch argument is not just a problem for panpsychism, but for any constitutivist view which supposes that phenomenal experience is simple and dependent on complex processes or states (such as the brain). Following is the form of the problem for constitutive panpsychism, phrased by Chalmers as arising from the inconsistency of four propositions,

1. Microphenomenal structure is isomorphic to microphysical structure.

2. Microphenomenal structure constitutes macrophenomenal structure.

3. Microphysical structure constitutes macrophysical structure.

4. Macrophenomenal structure is not isomorphic to macrophysical structure.[16]

1 states the Russellian view that microphenomenal properties underlie microphysical role-playing properties. 2 is a statement of constitutive panpsychism. 3 is a widely accepted view

[15] I raised a similar objection to the subject-summing problem in chapter 2.

[16] Chalmers, "The Combination Problem," p. 191. For other versions of the structural mismatch problem, see my overview in chapter 5 as well as Chalmers (2017), Goff (2017a), Lockwood (1993), Maxwell (1979), Nagasawa & Wager (2017), Roelofs (2019), and Stoljar (2001).

about the dependence of macrophysical properties on microphysical properties. 4 is a statement about the apparent discrepancy between macrophenomenal states and the macrophysical states which they depend on. For example, if I have a phenomenal experience of a uniform red field, this seems to be simple. There are no apparent diverse parts, components, or aspects to that experience. However, the brain states which constitute that phenomenal state are amazingly complex.

Because this argument was the topic of chapter 5, I will briefly restate the solution presented there. While our mental states appear relatively simple in introspection, this is merely a confused presentation of the experience's real, complex structure. As such, the confused ideas thesis leaves it is an open possibility that premise 4 is false and, despite appearances, macrophenomenal structure is isomorphic to macrophysical structure.

7.2 Confusion and alternative solutions

The confused ideas thesis is not the only game in town. Chalmers divides attempts to solve the combination problem into two camps: combinatorial and non-combinatorial responses.[17] Combinatorial responses suggest a process by which microsubjects and their microphenomenal properties might combine to constitute macrosubjects and their macrophenomenal properties. In contrast, non-combinatorial responses deny that microsubjects might combine to constitute macrosubjects and provide an alternative account of how macrosubjects arise. The confused ideas thesis is a moderate combinatorial response as it does not provide a positive characterization of how microsubjects give rise to macrosubjects. Rather, the confused ideas thesis has a more modest aim in attempting to show that our introspective access to our mental states does not undermine the possibility of microsubjects combining. As such, the confused ideas thesis should be largely consistent with other combinatorial views.

For the remainder of this section, I will evaluate other solutions to the combination problem. First, other combinatorial responses will be considered, focusing on Goff's phenomenal bonding solution and Coleman's continuum hypothesis. In my opinion, these two responses are the best supplementary views for the confused ideas thesis in providing a positive charac-

[17] Chalmers, "Combination Problem," pp. 191–92.

terization of how the macrosubjects and macrophenomenal properties arise. Then, I consider non-combinatorial solutions and evaluate how consistent they are with the confused ideas thesis. Throughout, I will also assess the relative strengths of these solutions to the confused ideas thesis.

7.2.1 The phenomenal bonding solution

The phenomenal bonding solution hypothesizes some "relation such that when subjects stand in it they produce a further subject."[18] In true Russellian fashion, this relation is hypothesized to be some empirically known relation in physics, perhaps gravitational or spatial relations. Further, the intrinsic nature of that relation accounts for phenomenal bonding.

Both confusion and the phenomenal bonding solution deny premise 4 of the subject-summing problem something inaccessible to introspection which allows for subjects to combine.

4. For any group of subjects [S] (with certain experiences), it is conceivable that those subjects exist (with their experiences) and no other subjects exist.

The phenomenal bonding solution appeals to the possibility of an intrinsic aspect of relations (as this bonding of microsubjects is not apparent when introspecting). Were the intrinsic nature of this relation accessible to introspection, then premise 4 would not be conceivable in any world where phenomenal bonding obtains.

As with the subject-summing problem, the confused ideas thesis and the phenomenal bonding solutions provide similar responses to the conceivability and knowledge problems. If the phenomenal bonding solution is correct, the possibility of phenomenal bonding serves as a defeater to premise 1 of both arguments. If PP duplicates the broadly material facts of our world and our world has a phenomenal bonding relation, then PP&˜Q is inconceivable and Q would be deducible from PP by panpsychist-Mary in her room.

The phenomenal bonding solution does not straightforwardly apply as a solution to the palette problem, though it might play a role in how microphenomenal qualities blend together to

[18] Philip Goff, "The Phenomenal Bonding Solution to the Combination Problem," in *Panpsychism: Contemporary Perspectives*, eds. Godehard Brüntrup and Ludwig Jaskolla (New York: Oxford University Press, 2017b), p. 293.

produce novel macrophenomenal qualities. Perhaps when two microsubjects are phenomenally bonded, their respective microqualities are blended as well. This blending of microphenomenal qualities into novel macrophenomenal qualities would make it easier to explain the vast rainbow of macrophenomenal qualities in terms of the limited palette of microphenomenal qualities.

If phenomenal bonding obtains, then premise 1 of the revelation problem is false.

1. The nature of consciousness is revealed to us through introspection.

The phenomenal bonding relation would, in part, constitute macroconsciousness but is not revealed through introspection.

Which solution should panpsychists prefer? Reasons to accept the confused ideas thesis were given in chapter 5. It should be noted that the phenomenal bonding solution and the confused idea thesis are compatible. It could be the case that confusion is limited to introspective representations of phenomenal properties while subjecthood is accurately represented in introspection and there is nothing in the nature of subjects, considered by themselves, that permits combination. It could be the case that the phenomenal bonding of microsubjects and their microphenomenal properties is how these ideas become confused. This form of the confusion view supplemented with a phenomenal bonding relation would solve the introspective combination problems just as well.

However, the phenomenal bonding solution faces a serious difficulty in articulating exactly which relation in physics is the phenomenal bonding relation. Any chosen relation results in counterintuitive consequences. For example, some relations are instantiated by all things in the universe, e.g., spatial and gravitational relations. If such a relation is also the phenomenal bonding relation, then we have quite the proliferation of macrosubjects! Any two or more things will constitute a new subject. Every third book in a library. Every fourth book in a library. Every third book in a library, every other blade of grass outside, and Pluto. More limited relations, such as the strong nuclear force which holds protons and neutrons together in a nucleus, do not operate at the scale we pre-theoretically suppose subjects to exist at. As such, pan(proto)psychists hoping to adopt the phenomenal bonding solution to solve the subject-summing problem while committing themselves only to "commonsensical" macrosubjects (like some subset of organisms) would be better off with the confused ideas thesis, which neither

rules out nor is committed to counterintuitive subjects made up of my left shoe, my right ear, and the Voyager space probe.

That being said, I think counter-intuitive consequences and pre-theoretical common-sense should play little role in metaphysical theorizing. After all, there is nothing incoherent in the idea of the unusual macrosubjects described above. More broadly, my reasons for downplaying counter-intuitiveness and commonsense objections are threefold. First, we have little to no reason to think that our commonsense views or intuitions track the nature of ultimate reality. Arguably, such beliefs were influenced by evolutionary pressures to be fitness enhancing, not to track deep metaphysical truths.[19] Second, our commonsense intuitions do not reliably track cases in our everyday, physical reality. To illustrate this, consider a ball being carried in the hands of a walking person. She lets go of the ball. How does the ball move? If you answered, "it falls straight down," you are wrong but not alone. In a study of 99 Johns Hopkins' undergraduates, 49% thought the ball would fall straight down (40% of whom had taken at least one physics course) and 6% thought it would fall backwards. The ball will actually fall in in a parabolic arc in the direction of the walker's movement, an answer only 45% of the students gave![20] This is not some strange case about metaphysical entities we can only imagine and have no direct experience of, this is an ordinary case that could be found in our everyday life. Presumably, this is the kind of case we should expect our commonsense intuitions to get right. Finally, to borrow an argument from Locke, there does not seem to be commonsense. There is no common or universal view, feeling, or intuition about nearly any subject, let alone metaphysics.[21] So, commonsense is of dubious value in the domain of metaphysical theorizing, gets it wrong in cases that seem tailor-made for it, and is not even common. As such, we should put little weight in our commonsense convictions when considering metaphysical theories and their implications.

[19]Similar arguments, with very different conclusions from my own, about evolutionary influences on our beliefs being selected for their fitness enhancing features and not truth-responsiveness can be found in Sharon Street (2006) and Alvin Plantinga (1993; 2011).

[20]Michael McCloskey, Allyson Washburn, and Linda Felch, "Intuitive Physics: The Straight-Down Belief and Its Origin," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9 (1983).

[21]John Locke, *An Essay Concerning Human Understanding* (New York: Oxford University Press, 1689/1975), I.ii and I.iii. Locke's argument targets innate ideas, ideas not gained from the senses. I am interested in his focus on the lack of common assent to any proposition.

7.2.2 The continuum hypothesis

Sam Coleman defends the continuum hypothesis[22] in response to the incommensurability problem, what he sees as the real problem at the heart of the palette problem,

> [j]ust what set of microqualities is it that in recombination can yield now a pure blueness, now the smell of roses? ... The problem is that some macroqualia are apparently so *unlike* some other macroqualia that we can't imagine them having ingredients in common.[23]

The source of the incommensurability problem, according to Coleman, is our view that qualities are modality specific. For example, that colors are qualities specific to visual modalities, smells to olfactory modalities, and so on.[24] In place of the distinct modalities picture of qualities, Coleman advances the continuum hypothesis. He encourages us to imagine

> phenomenal qualities, of all kinds, on a continuum, in the way we think of the colors. So just as it's possible to move across the color spectrum in tiny, almost undetectable steps, it must be possible to move from tastes to sounds, sounds to colors, and so on, via equally tiny steps.[25]

If phenomenal qualities are not modality specific and exist on a continuum, then it is easier to see how the limited palette of microphenomenal qualities are able to account for the diverse array of macrophenomenal qualities.

I find the continuum hypothesis quite plausible and, like the phenomenal bonding solution, it is compatible with the confused ideas thesis. Further, I think the continuum hypothesis and the confused ideas thesis make a strong pair. Accepting a conjunction of the two allows for cases of genuine fusion or combination of qualities in experience (as Coleman notes this seems actual in the case of taste which is a mix of olfactory and gustatory qualities)[26] while also allowing for cases where the mind simply cannot distinguish between several qualities and

[22] Coleman (2017) attributes this view originally to Hartshorne (1934).

[23] Sam Coleman, "Panpsychism and Neutral Monism: How to Make Up One's Mind" in *Panpsychism: Contemporary Perspectives*, eds. Godehard Brüntrup and Ludwig Jaskolla (New York: Oxford University Press, 2017b), pp. 263–64, emphasis in original.

[24] This view, I think, is received from the early modern period. Leibniz, for example, held a view like this, though he did think that some qualities were accessible across multiple modalities. We can see this when he distinguishes "the *notions of the particular senses*, which are *clear but confused*, and the *notions of the common sense*, which are*clear and distinct*." (1702 G 499/AG 187–88, emphasis in original.)

[25] Coleman, "Panpsychism and Neutral Monism," p. 264.

[26] *Ibid.*

represents them confusedly. For example, one might hold that the flavor of some dish really is the result of a fusion of olfactory and gustatory qualities while an individual not being able to notice the notes of lime, quinine, and botanicals in a gin and tonic is merely a result of confusion.

Unlike the confused ideas thesis and the phenomenal bonding solution, the continuum hypothesis does not provide solutions to other forms of the combination problem. As such, I think defenders of the continuum hypothesis have much to gain by integrating the confused ideas thesis into their view.

7.2.3 Emergentist panpsychism

Emergent panpsychism is, perhaps, the most prominent form of panpsychism defended today.[27] Unlike constitutive panpsychists who explain macroconsciousness as derivative of microconsciousness, emergent panpsychism takes macroconsciousness as fundamental. Their views are, broadly, divided into two camps. First, those that structurally parallel the views of the British emergentists. On this view, the world is made up of levels with higher levels being dependent on, but irreducible to lower levels. For levels emergentist panpsychists, all levels of reality are equally real and coexist. Second, fusionist views hypothesize microconscious entities fuse together into a new, unified macroconscious entity. Unlike levels emergentists, microentities cease to exist when they fuse into emergent macroentity.

The chief benefit of emergent panpsychism is how easily the view deals with the introspective combination problems. Macrosubjects are emergent, so no account of subject-summing is necessary. Our lack of knowledge of the laws of emergence (and the possibility of their contingency) accounts for the conceivability of panpsychist zombie worlds (PP&˜Q, conceivability problem) as well as panpsychist-Mary not knowing what-it-is-like to experience some macrophenomenal property despite knowing the complete microphysical and microphenomenal truth (Q not being deducible from PP, knowledge problem). Because macrosubjects are not constituted by microsubjects, the limited palette of microphenomenal properties need not account for the vast array of macrophenomenal properties (palette problem) nor is there a prob-

[27] See Brüntrup (2017), Goff (2015, 2017a, 2017b, 2024), Mørch (2016, 2019), and Seager (2017).

lem of structural mismatch. Finally, there is no relevant revelation problem. All together, this is a serious benefit for the view.

However, there is a serious drawback to emergence: the problem of accounting for mental causation. Levels emergentist panpsychists face the exclusion problem.[28] It is broadly assumed that the world is physically causally closed, that is, for every effect there is a sufficient physical cause. Further, it seems plausible that this is true for the micro-level, we can give a complete causal picture of the world by appealing only to the causal powers of the world's basic entities. But, if micromaterial causal closure is true and macrosubjects are not wholly grounded in microphysical entities, then it is unclear what causal role macrosubjects could play. Macrosubjects would seem to either be epiphenomenal, lacking any causal powers at all, or, if they do have causal powers, these powers would seem to overdetermine their effects. Overdetermination occurs when an effect has more than one sufficient cause. Levels emergent macrosubjects would overdetermine their effects because there would also be a sufficient micromaterial cause. The levels emergentist must accept either that no one ever had a drink because they were thirsty or that overdetermination prevails the natural world.

Fusionists, on the other hand, avoid the exclusion problem because the basal microsubjects cease to exist when the macrosubject emerges. As such, the effects of fusionist macrosubjects would not also have a sufficient microphysical cause and, therefore, not be overdetermined. However, this requires the fusionist to deny microphysical causal closure. How problematic this turns out to be is a matter of debate. While broadly assumed, arguments for microphysical causal closure are scarce.

How does the confused ideas thesis fare in comparison? As both views are able to address the introspective combination problems, I take it they both fare equally well in that regard. However, constitutive panpsychism in conjunction with the confused ideas thesis does not face level emergentism's problem with mental causation nor fusionism's controversial views about extended simples and denying microphysical closure.[29]

[28] Jaegwon Kim, "Emergence: Core Ideas and Issues," *Synthese* 151 (2006).

[29] For more reasons to reject emergence, see chapter 5.

7.2.4 Dominant monad panpsychism

A rarely defended view, but, given that it shares a Leibnizian inspiration with the confused ideas thesis, it merits discussion.[30] Leibniz held that individuals were made up of infinitely many soul-like substances, monads, and that a single, dominant monad could be identified as the individual's soul. The dominant monad can also be said to exert control over its subordinate monads or body.[31]

Much as I did not adopt the whole of Leibniz's confused ideas thesis in appropriating it, neither does Nino Kadić adopt the entirety of Leibniz's monadological framework. Rather, Kadić adopts Leibniz's views that fundamental, possibly simple, entities can represent and have vastly complex mental experiences and content, fundamental entities are arranged in a hierarchical structure with dominant and subordinate varieties, and that these fundamental entities are individuated by their internal states.

Kadić proposes,

> *Microphenomenal Structuralism*. The phenomenal character of any given microsubject is determined by its relations to other microsubjects featuring in the same relevant causal structure.[32]

Microsubjects in themselves, without the phenomenal character gained from standing in relations with other microsubjects, have only a basic form of consciousness without any awareness, phenomenality, or content. They are a precondition for experience. The phenomenality of a microsubject "depends on what position it occupies in the relevant causal structure."[33]

Kadić identifies three possible forms of dominant monad panpsychism: static, dynamic, and global.

> *Static monadic panpsychism (SMP)*. All microsubjects featuring in the relevant structure relationally determine the phenomenal character of only one particular microsubject in a way that gives it a full human experience. That particular microsubject plays the dominant role.[34]

[30] Perhaps even a uniquely defended view, as it seems to be solely defended by Nino Kadić (2024), though Chalmers (2015; 2017) does consider a similar view.

[31] See Leibniz (2014).

[32] Nino Kadić, "Monadic Panpsychism," *Synthese* 203 (2024): p. 4.

[33] *Ibid.*, p. 5.

[34] *Ibid.*, p. 7.

Kadić, rightly, I think, dismisses SMP for raising too many problems. What would happen if the dominant microsubject were removed from the body? Given that cells die and are replaces, how can a single fundamental entity (presumably a constituent of a cell) play the dominant role for the organism's entire life?

> *Dynamic monadic panpsychism (DMP).* All microsubjects featuring in the relevant structure relationally determine the phenomenal character of only one microsubject in a way that gives it a full human experience. That microsubject plays the dominant role, but which particular microsubject that is can change at any given time.[35]

DMP avoids the problems of SMP by not requiring a singular microsubject to play the dominant role as any other microsubject could shift into that role. Kadić thinks the two major problems for DMP, which are also problems for SMP, involve the metaphysics and epistemology of which microsubject is in the dominant role. How does a microsubject become dominant and how do we know which is? The metaphysical question, Kadić argues, can be empirically addressed. As we learn which brain structures play a role in organizing and integrating conscious experience, we get a little bit closer to an answer. Further, "we could conceivably identify what the hierarchy looks like at the fundamental level and thus discover the principles of selection and change governing microsubjects by monitoring the brain over time, though this is likely unviable."[36] The epistemological question seems to only be answerable from the inside, "it's me!", and not from an external perspective.

> *Global monadic panpsychism (GMP).* All microsubjects featuring in the relevant structure relationally determine the phenomenal character of all other microsubjects that make up the structure, so that the phenomenal character of every single one of them is the full human experience.[37]

GMP is the least similar to Leibniz's original view, with each microsubject having the complete conscious experience of the individual. This renders the idea of a singular self an illusion, with each dominant microsubject an equally good candidate for "I."

[35] Kadić, "Monadic Panpsychsim," p. 8.
[36] *Ibid.*, p. 9.
[37] *Ibid.*

Dominant monad panpsychism clearly avoids the subject-summing problem, as there is no combination of microsubjects. Because the macrophenomenal facts are "baked in" to the dominant microsubject, the conceivability and knowledge problems are avoided. PP&˜Q is not conceivable, because, *ex hypothesi*, Q is a conjunct of PP which is also why Q must be deducible from PP. The palette problem is avoided as "while there might be few intrinsic phenomenal properties, there can be a plethora of relational phenomenal properties."[38]

Ironically, given their Leibnizian origins, it is unclear how consistent dominant monad panpsychism and the confused ideas thesis are. For an idea to be confused means that it has some parts or aspects which are not discernible through introspection. If our ideas are a result of our dominant monad standing is casual relations with other microsubjects, then, for our ideas there would have to be some mechanism to explain how the dominant microsubject could more or less distinctly represent certain ideas.

While dominant monad panpsychism is able to address many forms of the combination problem, I think the microphenomenal structuralism thesis poses serious problems. Similar to the phenomenal bonding solution, one problem arises from specifying which causal relations are "relevant" to determining the phenomenal character of microsubjects. If all causal relations are capable of determining phenomenal character, then various and distant objects will affect the phenomenal character of microsubjects. Distant galaxies and birds flying by will exert a minute gravitational influence on the fundamental entities in the brain and influence their phenomenal character. In a way, the resulting view is even more Leibnizian, as the dominant microsubject expresses a world in the subtleties of its phenomenal character. Further, such widespread causal interactions allow for the possibility of dominant microsubjects of non-traditional subjects, such as between my left shoe, every other book on my shelf, and Pluto or even a cosmic subject. The possibilities of such strange subjects, just as with the phenomenal bonding solution, does not entail the falsity of the view. But, attempting to avoid such counter-intuitive conclusions will prove just as difficult as it did for phenomenal bonding. Presumably, the dominant monad panpsychist wishes to appeal to causal relations within a nervous system or the brain as playing the relevant role. But, then she must specify why the intrinsic nature of

[38] Kadić, "Monadic Panpsychism," p. 13.

those relations such that they affect phenomenal character and not others? If causal structure in the brain are what is relevant for microphenomenal structuralism in virtue of their being causal, then it is unclear why any causal relation would not determine the phenomenality of some microsubjects in its relata.

7.3 Objections to the confused ideas thesis

Following Alfred North Whitehead, Seager does not think much of Leibniz's confused ideas thesis as a solution to the combination problem. Seager charges that "without an analysis of how confusion arises this is merely a verbal solution."[39] By "verbal solution" I take Seagar to mean something like a suggestion of a possibility without elaborating on a mechanism for how the process happens. Further, Seager quotes Whitehead on Leibniz and confusion,

> the integration of simple physical feelings into complex physical feeling only provides for the various actual entities of the nexus being felt as separate entities requiring each other. We have to account for the substitution of the one nexus in place of its components.[40]

As an objection raised against Leibniz's original version of the view. A such, I find the objection lacking and arising from a misunderstanding of Leibniz. Leibniz's panpsychism does not face the combination problems discussed in this chapter. The dominant monad of an individual influences and expresses its subordinate monads, but is not a result of monads summing to novel subjects. While a living individual, on Leibniz's view, can be thought of as made up of infinitely many monads, she is strictly identical only to her dominant monad. As such, Leibniz faces none of the combination problems we are concerned with in this chapter.

To avoid the charge of a verbal solution against my version of the view, the best I can offer is merely a sketch. Confusion seems to arise when the mind is unable to discriminate between the myriad contents accessible to it. How this occurs is a task for empirical study. One mechanism that might play an important role is "covert attention" or "the deep processing of some select signals at the expense of other signals . . . a selective signal enhancement caused by

[39]William Seager, "Panpsychist Infusion" in *Panpsychism: Contemporary Perspectives*, eds. Godehard Brüntrup and Ludwig Jaskolla (New York: Oxford University Press, 2017), pp. 237–38.

[40]Alfred North Whitehead, *Process and Reality* (New York: Free Press, 1929/1969), p. 293.

competition among signals in the brain."[41] According to Michael Graziano's attention schema theory (AST),

> the brain controls attention with the help of a model of attention. That model, or 'attention schema', is a constantly updating set of information that describes the current state of attention and predicts how attention may transition into future states. Without the model, the brain can't steer attention.[42]

However, like all models our brains build, the attention schema will be an "imperfect model, lacking details."[43]

I suggest that attention plays a role in making some ideas more distinct by allowing "the limited resources of the brain to process selected items in great depth"[44] and making some ideas more confused by not bringing these resources to bear. This is not meant to be an exhaustive account. There might be other cognitive, perceptual, or metaphysical mechanisms by which ideas are made distinct or confused. For example, as I suggested in subsection 7.2.1, the phenomenal bonding of microsubjects and their microphenomenal properties might account for the confusion of microphenomenal properties for macrosubjects. If that is the case, then phenomenal bonding might explain microphenomenal confusion while attention explains macrophenomenal attention.

[41] Michael S.A. Graziano, "Consciousness Engineered," *Journal of Consciousness Studies* 23 (2016): p. 101.
[42] Michael S.A. Graziano, "Understanding Consciousness," *Brain* 144 (2021): p. 1283.
[43] *Ibid.*
[44] Graziano, "Understanding Consciousness," p. 1282.

Chapter 8

Conclusion

Introspection in the extremes! In chapter 2, I evaluated the full revelation thesis, that introspection provides the complete essential nature of our conscious states. I argued that the view is too strong and that partial revelation, the view that introspection only provides access to part of the essential nature of our conscious states. In chapter 3, I evaluated illusionism, the view that introspection distorts the mental states introspected upon, representing them as having qualitative characters they actually lack. I argued that the full revelation thesis and illusionism were both problematic models of introspection. Full revelation is too strong and leads to the combination problems. Illusionism is too weak and motivated by skepticism. A better model is needed.

Despite arguing in chapter 4 that G.W. Leibniz utilized the view to argue for illusionism, I believe the *confused ideas thesis* is that model. A subject's idea is confused if it has a structure or content that is inaccessible in introspection. In chapter 5, I showed how the confused ideas thesis is entailed by *constitutivism* or any view where the mind is grounded in, realized by, or identical to basal material states. Then, in chapter 6, I showed how constitutivism, the confused ideas thesis, and realism about phenomenal properties entail the disjunction of panpsychism or panprotopsychism, "pan(proto)psychism." If I am correct, then many widely accepted views in the metaphysics of consciousness are committed not just to the confused ideas thesis, but pan(proto)psychism as well! Pan(proto)psychism faces the many forms of the combination problem. In chapter 7, I showed how the confused ideas thesis is able to solve most forms of the combination problem. An incredible amount of bang for our theoretical buck.

But, there is much work left to do! While the confused ideas thesis does solve the combination problem, it does not provide a positive characterization of *how* microsubjects combine. Instead, I showed how what we know about our mental states through introspection leaves an open possibility for their combination.

Is a positive characterization possible? I'm pessimistic about the prospects for such work.

The first reason has to do with my Russellian commitments. The sciences only describe the causal structure of the world, telling us nothing about how things are intrinsically. The Russellian monist finds a place for a mind as the categorical basis for this causal structure. If the physical sciences only describes the causal structure of the world and what allows for combination is an intrinsic feature of subjecthood, then the sciences cannot aide us in discovering a positive characterization. That there seems to us to be combination problems and their refusal to be solved seems to indicate that introspection cannot illuminate the process.

The second reason has to do with the nature of combination more generally. Specifically, that we lack an answer to the special composition question, the problem in ontology of explaining when some set of objects aggregate to a new object as opposed to being merely a heap. If we cannot answer when atoms constitute a table, then I am unsure we can answer when microsubjects constitute a macrosubject.

Is the lack of a positive characterization a problem? It depends.

It would certainly make the view and panpsychism, generally, more attractive to have a positive characterization. The panpsychist's inability to address the combination problems is considered a serious problem for her view. However, many philosophers accept the existence of composite objects, including many in the metaphysics of mind who think that human consciousness is the result of some composite object, like the brain. But, it is not considered a serious problem for these views that there is not an answer to the special composition question. So, if it is a problem, it is not a disqualifying one.

Abbreviations and Conventions

A = *Sämtliche Schriften und Briefe.* Deutsche Akademie der Wissenschaft edn. Darmstadt, Leipzig, Berlin: Akademie Verlag, 1923–.

AG = *Philosophical Essays*. Translated by Roger Ariew and Daniel Garber. Indianapolis: Hackett, 1989.

CSM = *The Philosophical Writings of Descartes.* 3 vols. J. Cottingham, R. Stoothoff, and D. Murdoch, eds. New York: Cambridge University Press, 1985.

DSR = *De Summa Rerum: Metaphysical Papers, 1675-1676*. G.H.R. Parkinson, trans. New Haven: Yale University Press, 1992.

G = *Die philosophischen Schriften*. 7 vols. C.I. Gerhardt, ed. Berlin, 1875-90.

LA = *The Leibniz-Arnauld Correspondence with Selections from the Correspondence with Ernst, Landgrave of Hessen-Rheinfels*. Translated by Stephen Voss. Yale University Press. 2016.

LD = *The Leibniz-De Volder Correspondence with Selections from the Correspondence Between Leibniz and Johann Bernoulli*. Translated by Paul Lodge. Yale University Press. 2013.

MP = *Leibniz: Philosophical Writings*. M. Morris and G.H.R. Parkinson, eds. London: J. M. Dent & Sons, 1973.

NE = *New Essays on Human Understanding*. Translated by Peter Remnant and Jonathan Bennett, New York: Cambridge University Press, 1704/1996.

References to Leibniz's work will generally take the following form: [(year, if known)(original language citation)/(translation citation)]. For example, 1680(?) A VI.4 1443/AG 31 references "Primary Truths" in both the Akademie edition and *Philosophical Essays*. If the year is unknown, it will be replaces with four question marks. If the year is uncertain, as in the above example, it will be followed with a question mark.

References to correspondence collections will take the following form: (Name of correspondence abbreviated) (date of letter or draft). For example, LD 21 January 1704, 287 references the letter dated 21 January 1704 from *The Leibniz-De Volder Correspondence*..

References to David Hume's *A Treatise of Human Nature*, G.W. Leibniz's *New Essays on Human Understanding*, and John Locke's *Essay Concerning Human Understanding* will have the following structure: [(name of work) (book).(chapter).(section)] using capital Roman numerals for the book, lower case Roman numerals for the part, and Arabic numerals for the section. For example, "*Treatise* I.ii.3" is a reference to book one, part two, section three of Hume's *Treatise*.

Bibliography

Alter, T. (2015). The structure and dynamics argument against materialism. *Noûs*, *50*(4), 794–815.

Alter, T., & Nagasawa, Y. (2012). What is russellian monism? *Journal of Consciousness Studies*, *19*(9-10), 67–95.

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

Alter, T., & Pereboom, D. (2023). Russellian monism and structuralism about physics. *Erkenntnis*, *88*(4), 1409–28.

Alter, T., & Pereboom, D. (Fall 2023). Russellian monism. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/fall2023/entries/russellian-monism/.

Amijee, F. (2021). Explaining contingent facts. *Philosophical Studies*, *178*(4), 1163–81.

Armstrong, D. M. (1968). *A Materialist Theory of the Mind*. New York: Routledge.

Balog, K. (2011). In defense of the phenomenal concept strategy. *Philosophy and Phenomenological Research*, *84*(1), 1–23.

Barnes, E. (2012). Emergence and fundamentality. *Mind*, *121*(484), 873–901.

Bennett, K. (2003). Why the exclusion problem seems intractable and how, just maybe, to tract it. *Noûs*, *37*(3), 471–97.

Bennett, K. (2008). Exclusion again. In J. Hohwy, & J. Kallestrup (Eds.) *Being Reduced: New Essays on Reduction, Explanation, and Causation*, (pp. 280–307). New York: Oxford University Press.

Block, N. (1995). On a confusion about a function of consciousness. *Brain and Behavioral Sciences*, *18*(2), 227–47.

Brown, D. J., & Normore, C. G. (2019). *Descartes and the Ontology of Everyday Life*. New York: Oxford University Press.

Brüntrup, G. (2017). Emergent panpsychism. In G. Brüntrup, & L. Jaskolla (Eds.) *Panpsychism: Contemporary Perspectives*, (pp. 48–71). New York: Oxford University Press.

Brüntrup, G., & Jaskolla, L. (Eds.) (2017). *Panpsychism: Contemporary Perspectives*. New York: Oxford University Press.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–19.

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

Chalmers, D. J. (2006). Strong and weak emergence. In P. Clayton, & P. Davies (Eds.) *The Re-Emergence of Emergence: the Emergentist Hypothesis from Science to Religion*. Oxford University Press.

Chalmers, D. J. (2009). The two-dimensional argument against materialism. In A. Beckermann, B. P. McLaughlin, & S. Walter (Eds.) *The Oxford Handbook of Philosophy of Mind*, (pp. 313–37). New York: Oxford University Press.

Chalmers, D. J. (2015). Panpsychism and panprotopsychism. In T. Alter, & Y. Nagasawa (Eds.) *Consciousness in the Physical World*, (pp. 246–76). New York: Oxford University Press.

Chalmers, D. J. (2017). The combination problem for panpsychism. In G. Brüntrup, & J. Ludwig (Eds.) *Panpsychism: Contemporary Perspectives*, (pp. 179–214). New York: Oxford University Press.

Chalmers, D. J. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies*, *25*(9-10), 6–61.

Chalmers, D. J. (2020a). Debunking arguments for illusionism about consciousness. *Journal of Consciousness Studies*, *27*(5-6), 258–81.

Chalmers, D. J. (2020b). How can we solve the meta-problem of consciousness? *Journal of Consciousness Studies*, *27*(5-6), 201–26.

Cohen, S. (2010). Bootstrapping, defeasible reasoning, and a priori justification. *Philosophical Perspectives*, *24*(1), 141–59.

Coleman, S. (2013). The real combination problem: Panpsychism, micro-subjects, and emergence. *Erkenntnis*, *79*(1), 19–44.

Coleman, S. (2017). Panpsychism and neutral monism: How to make up one's mind. In G. Brüntrup, & L. Jaskolla (Eds.) *Panpsychism: Contemporary Perspectives*, (pp. 249–82). New York: Oxford University Press.

Crane, T. (2001). The significance of emergence. In C. Gillett, & B. Loewer (Eds.) *Physicalism and its Discontents*, (pp. 207–24). New York: Cambridge University Press.

Crane, T. (2010). Cosmic hermeneutics vs. emergence: The challenge of the explanatory gap. In G. Macdonald, & C. Macdonald (Eds.) *Emergence in mind*, (pp. 22–34). New York: Oxford University Press.

Dasgupta, S. (2014). Metaphysical rationalism. *Noûs*, *50*(2), 379–418.

Dennett, D. C. (1991). *Consciousness Explained*. New York: Little, Brown and Company.

Dennett, D. C. (2016). Illusionism as the obvious default theory of consciousness. *Journal of Consciousness Studies*, *23*(11-12), 65–72.

Dennett, D. C. (2017). *From Bacteria to Bach and Back: The Evolution of Minds*. New York: W. W. Norton & Company.

Dennett, D. C. (2019). Welcome to strong illusionism. *Journal of Consciousness Studies*, *26*(9-10), 48–58.

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, *23*(11-12), 11–39.

Frankish, K. (Ed.) (2017). *Illusionism as a Theory of Consciousness*. Exeter: Imprint Academic.

Frankish, K. (2019). The meta-problem is the problem of consciousness. *Journal of Consciousness Studies*, *26*(9-10), 83–94.

Goff, P. (2006). Experiences don't sum. *Journal of Consciousness Studies*, *13*(10-11), 53–61.

Goff, P. (2015). Against constitutive russellian monism. In T. Alter, & Y. Nagasawa (Eds.) *Consciousness in the Physical World: Perspectives on Russellian Monism*, (pp. 370–400). New York: Oxford University Press.

Goff, P. (2017a). *Consciousness and Fundamental Reality*. New York: Oxford University Press.

Goff, P. (2017b). The phenomenal bonding solution to the combination problem. In G. Brüntrup, & L. Jaskolla (Eds.) *Panpsychism: Contemporary Perspectives*, (pp. 283–302). New York: Oxford University Press.

Goff, P. (2024). How exactly does panpsychism help explain consciousness? *Journal of Consciousness Studies*, *31*(3), 56–82.

Graziano, M. S. A. (2016). Consciousness engineered. *Journal of Consciousness Studies*, *23*(11-12), 98–115.

Graziano, M. S. A. (2021). Understanding consciousness. *Brain*, *144*, 1281–83.

Hartshorne, C. (1941). *The Philosophy and Psychology of Sensation*. Chicago: University of Chicago Press.

Hempel, C. G. (1980). Comments on goodman's ways of worldmaking. *Synthese*, *45*(2), 193–99.

Howell, R. (2015). The russellian monist's problems with mental causation. *Philosophical Quarterly*, *65*(258), 22–39.

Hume, D. (1739/2000). In D. F. Norton, & M. J. Norton (Eds.) *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning Into Moral Subjects*. New York: Oxford University Press.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, *32*(April), 127–36.

James, W. (1890). *The Principles of Psychology*. London: Dover Publications.

Jolley, N. (1984). *Leibniz and Locke: A Study of the New Essays on Human Understanding*. New York: Oxford University Press.

Kadić, N. (2024). Monadic panpsychism. *Synthese*, *203*(2), 1–18.

Kim, J. (1993). The nonreductivist's troubles with mental causation. In J. Kim (Ed.) *Supervenience and Mind: Selected Philosophical Essays,*, (pp. 336–57). New York: Cambridge University Press.

Kim, J. (2006). Emergence: Core ideas and issues. *Synthese*, *151*(3), 547–59.

Kripke, S. (1972/1980). *Naming and Necessity*. Cambridge: Harvard University Press.

Ladyman, J., & Ross, D. (2007). *Every Thing Must Go: Metaphysics Naturalized*. New York: Oxford University Press.

Leibniz, G. W. (2014). The monadology. In L. Strickland (Ed.) *Leibniz's Monadology: A New Translation and Guide*. Edinburgh: Edinburgh University Press.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*(October), 354–61.

Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy*, *74*(4), 549–67.

Lewis, D. K. (1966). An argument for the identity theory. *Journal of Philosophy*, *63*(1), 17–25.

Lin, M. (2012). Rationalism and necessitarianism. *Noûs*, *46*(3), 418–48.

Locke, J. (1689/1975). In P. H. Nidditch (Ed.) *An Essay Concerning Human Understanding*. New York: Oxford University Press.

Lockwood, M. (1993). The grain problem. In H. Robinson (Ed.) *Objections to Physicalism*, (pp. 271–91). New York: Oxford University Press.

Maxwell, G. (1978). Rigid designators and mind-brain identity. *Minnesota Studies in the Philosophy of Science*, *9*, 365–403.

McCloskey, M., Washburn, A., & Felch, L. (1983). Intuitive physics: The straight-down belief and its origin. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(4), 636–49.

McLaughlin, B. P. (2008). The rise and fall of british emergentism. In M. A. Bedau, & P. Humphreys (Eds.) *Emergence: Contemporary Readings in Philosophy and Science*, (pp. 19–60). Cambridge: MIT Press.

Mørch, H. H. (2014). *Panpsychism and Causation: A New Argument and a Solution to the Combination Problem*. Ph.D. thesis, Oslo.

Mørch, H. H. (2019). Is the integrated information theory of consciousness compatible with russellian panpsychism? *Erkenntnis*, *84*(5), 1065–85.

Nagasawa, Y. (2012). Infinite decomposability and the mind-body problem. *American Philosophical Quarterly*, *49*(4), 357–67.

Nagasawa, Y., & Wager, K. (2017). Panpsychism and priority cosmopsychism. In G. Brüntrup, & L. Jaskolla (Eds.) *Panpsychism: Contemporary Perspectives*, (pp. 113–29). New York: Oxford University Press.

Nagel, T. (1979). Panpsychism. In T. Nagel (Ed.) *Mortal questions*, (pp. 181–95). New York: Cambridge University Press.

Nagel, T. (2012). *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False*. New York: Oxford University Press.

Ott, W. (2016). Leibniz on sensation and the limits of reason. *History of Philosophy Quarterly*, *33*(2), 135–53.

Pereboom, D. (2011). *Consciousness and the Prospects of Physicalism*. New York: Oxford University Press.

Pereboom, D. (2015). Russellian monism and absolutely intrinsic properties. In T. Alter, & Y. Nagasawa (Eds.) *Consciousness in the Physical World*, (pp. 300–23). New York: Oxford University Press.

Pereboom, D. (2016). Illusionism and anti-functionalism about phenomenal consciousness. *Journal of Consciousness Studies*, *23*(11-12), 172–85.

Pereboom, D. (2019). Russellian monism, introspective inaccuracy, and the illusion meta-problem of consciousness. *Journal of Consciousness Studies*, *26*(9-10), 182–93.

Plantinga, A. (2011). *Where the Conflict Really Lies: Science, Religion, and Naturalism*. New York: Oxford University Press.

Pryor, J. (2000). The skeptic and the dogmatist. *Noûs*, *34*(4), 517–49.

Puryear, S. M. (2005). Was leibniz confused about confusion? *The Leibniz Review*, *15*, 95–124.

Puryear, S. M. (2012). Leibniz's alleged ambivalence about sensible qualities. *Studia Leibnitiana*, *44*(2), 229–45.

Rey, G. (2016). Taking consciousness seriously–as an illusion. *Journal of Consciousness Studies*, *23*(11-12), 197–214.

Reynolds, S. L. (2012). Effective sceptical hypotheses. *Theoria*, *79*(3), 262–78.

Robinson, H. (1982). *Matter and Sense: A Critique of Contemporary Materialism*. New York: Cambridge University Press.

Robinson, H. (2016). *From the Knowledge Argument to Mental Substance: Resurrecting the Mind*. New York: Cambridge University Press.

Roelofs, L. (2019). *Combining Minds: How to Think About Composite Subjectivity*. New York: Oxford University Press.

Roelofs, L. (2021). Is panpsychism at odds with science? *Journal of Consciousness Studies*, *28*(9-10), 116–28.

Russell, B. (1927). *The Analysis of Matter*. London: Kegan Paul.

Schaffer, J. (2003). Is there a fundamental level? *Noûs*, *37*(3), 498–517.

Schaffer, J. (2009). On what grounds what. In R. Wasserman, D. Manley, & D. Chalmers (Eds.) *Metametaphysics: New Essays on the Foundations of Ontology*, (pp. 347–83). Oxford University Press.

Schaffer, J. (2010). Monism: The priority of the whole. *Philosophical Review*, *119*(1), 31–76.

Seager, W. (2010). Panpsychism, aggregation and combinatorial infusion. *Mind and Matter*, *8*(2), 167–84.

Seager, W. (2017). Panpsychist infusion. In G. Brüntrup, & L. Jaskolla (Eds.) *Panpsychism: Contemporary Perspectives*, (pp. 229–48). New York: Oxford University Press.

Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review*, *68*(April), 141–56.

Stoljar, D. (2001). Two conceptions of the physical. *Philosophy and Phenomenological Research*, *62*(2), 253–81.

Strawson, G. (2006). Realistic monism: Why physicalism entails panpsychism. In A. Freeman (Ed.) *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?*, (pp. 3–31). Exeter: Imprint Academic.

Street, S. (2006). A darwinian dilemma for realist theories of value. *Philosophical Studies*, *127*(1), 109–66.

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

Taylor, H. (2019). Is panpsychism simple? *Analysis*, *79*(2), 265–75.

Whitehead, A. N. (1929). *Process and Reality: An Essay in Cosmology*. New York: Free Press.

Wilson, J. (2005). Supervenience-based formulations of physicalism. *Noûs*, *39*(3), 426–59.

Wilson, J. M. (2021). *Metaphysical Emergence*. New York: Oxford University Press.

Wilson, M. D. (1999a). Confused ideas. In M. D. Wilson (Ed.) *Ideas and Mechanism: Essays on Early Modern Philosophy*, (pp. 322–35). Princeton: Princeton University Press.

Wilson, M. D. (1999b). Confused vs. distinct perception in leibniz: Consciousness, representation, and god's mind. In M. D. Wilson (Ed.) *Ideas and Mechanism: Essays on Early Modern Philosophy*, (pp. 336–52). Princeton: Princeton University Press.

Wilson, M. D. (1999c). Leibniz and materialism. In M. D. Wilson (Ed.) *Ideas and Mechanism: Essays on Early Modern Philosophy*, (pp. 388–406). Princeton: Princeton University Press.