

April 2021

MOVING FORWARD ON THE PROBLEM OF CONSCIOUSNESS?

Haoying Liu
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2



Part of the [Philosophy of Mind Commons](#)

Recommended Citation

Liu, Haoying, "MOVING FORWARD ON THE PROBLEM OF CONSCIOUSNESS?" (2021). *Doctoral Dissertations*. 2120.
<https://doi.org/10.7275/20694146> https://scholarworks.umass.edu/dissertations_2/2120

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

MOVING FORWARD ON THE PROBLEM OF CONSCIOUSNESS?

A Dissertation Presented

by

HAOYING LIU

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirement for the degree of

DOCTOR OF PHILOSOPHY

February 2021

Philosophy Program

© Copyright by Haoying Liu 2021

All Rights Reserved

MOVING FORWARD ON THE PROBLEM OF CONSCIOUSNESS?

A Dissertation Presented

by

HAOYING LIU

Approved as to style and content by:

Joseph Levine, Chair

Louise Antony, Member

Alejandro Pérez Carballo, Member

Kyle Cave, Member

Phillip Bricker, Department Chair
Philosophy

ACKNOWLEDGMENTS

I would like to thank my advisor, Joseph Levine, for his support and guidance over the years. Without his persistent challenge to my earlier thoughts on panpsychism, my attitude toward panpsychism would not have shifted in recent years. I would also like to extend my gratitude to the members of my committee, Louise Antony, Alejandro Pérez Carballo, and Kyle Cave, for their helpful comments and suggestions on all stages of this project.

A special thank you to all those whose support and friendship helped me to stay focused on this project and who have provided me with the encouragement to continue when the process got tough. Without your support, I probably would not have had the mental and physical health to finish this project.

ABSTRACT

MOVING FORWARD ON THE PROBLEM OF CONSCIOUSNESS?

FEBRUARY 2021

HAOYING LIU, B.A., FUDAN UNIVERSITY

M.A., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Joseph Levine

Abstract:

The problem of consciousness has been an issue in philosophy of mind for decades, and in recent years panpsychism and panprotopsychism have gained attention among philosophers who are still dedicated to finding a complete explanation of consciousness. In this dissertation, I criticize panpsychism and panprotopsychism by examining their metaphysical plausibility and their epistemic prospects. Concerning the metaphysical plausibility of panpsychism and panprotopsychism, I explain the “combination problem” of panpsychism and criticize several major accounts of panpsychism and panprotopsychism that aim at solving this problem, including Seager’s panpsychist infusionism, Goff’s phenomenal bonding proposal, and Coleman’s panqualityism. I also examine the proposal of motivating panpsychism and panprotopsychism by drawing support from the integrated information theory (IIT) of consciousness. I argue that this proposal to merge panpsychism or panprotopsychism with IIT does not work. Concerning the epistemic prospect of panpsychism and panprotopsychism, I raise concerns about whether panpsychism and panprotopsychism can guide a fruitful research program about

the nature of consciousness. First, I argue that a theory based on Russellian Monism would not have a promising epistemic prospect. I compare Russellian Monism to Thomistic Hylomorphism and argue that Russellian Monism faces similar epistemic challenges that plagued Thomistic Hylomorphism. Since most theories of contemporary panpsychism and panprotopsyism rely on the metaphysical framework of Russellian Monism, the challenges to Russellian Monism also undermine the epistemic prospects of panpsychism and panprotopsyism. Second, I challenge the epistemic prospects of panpsychism and panprotopsyism by examining if there can be a viable method to study the consciousness of micro-level entities. After revisiting the reasoning process that guided the study of atoms, I argue that studying micro-level facts requires “analogies” that compare micro-level facts with macro-level models. Then I argue that an analogy that fits the need of panpsychism or panprotopsyism is hard to achieve, because we have no macro-level models to be compared to the hypothesized systems of micro-level consciousness.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	iv
ABSTRACT	v
LIST OF FIGURES	xi
INTRODUCTION	1
CHAPTER	
PART I	5
1. AN INTRODUCTION TO PAN(PROTO)PSYCHISM	5
1.0. General Introduction.	5
1.1. First Argument for Pan(proto)psychism	9
1.2. Second Argument for Pan(proto)psychism	16
1.2.1 Intrinsic Nature	16
1.2.2 Two Birds, One Arrow	20
1.2.3 Panpsychism and panprotopsyism	24
1.3. The combination problem	28
1.3.1 The combination of subjects	28
1.3.2 The combination of qualities and structures	31
1.4. Chapter Conclusion	35
2 . OF SOME PROPOSALS TO SOLVE THE COMBINATION PROBLEM.....	36
2.1. Emergentist panpsychism	38
2.1.1 Emergentism	38
2.1.2 Of Seager's Panpsychist Infusion.....	42
2.2. Constitutive panpsychism	52
2.2.1 Phenomenal bonding	52

2.2.2 Do we know what Phenomenal bonding is?	55
2.2.3 A Challenge to Phenomenal bonding	59
2.2.4 Another Challenge: What's the use of mental dusts	63
2.3. Panprotopsyism	66
2.3.1 Panprotopsyism	67
2.3.2 Of Coleman's panqualityism	68
2.3.2.1 Introducing panqualityism	68
2.3.2.2 Against Coleman on Subjectivity of Consciousness	72
2.4. Chapter Conclusion	77
3. FROM PANPSYCHISM TO HYLOMORPHISM?	78
3.1. Desiderata of Mental Combination.....	79
3.2. IIT.....	82
3.2.1 Phenomenological motivations	83
3.2.2 A Briefing on the Machinery of IIT.....	84
3.3. Does IIT fit with Panpsychism?	90
3.3.1 IIT and Panpsychism	90
3.3.2 Mørch on Panpsychist IIT.....	94
3.3.3 Panpsychism unfit for IIT.....	98
3.3.4 IIT and Russellian Panpsychism	100
3.4. IIT and functionalism	102
3.4.1 Functionalism	102
3.4.2 IIT: not exactly the same as functionalism.....	103
3.5. Hylomorphism	109
3.5.1 An Account of Hylomorphism	110
3.5.2 IIT and Hylomorphism	114
3.6. Chapter Conclusion.....	120
PART II	121

4. RUSSELLIAN MONISM, THOMISTIC HYLOMORPHISM, AND THE MYSTERY OF CONSCIOUSNESS	121
4.1. Introduction.....	121
4.2. Epistemic Inadequacy of Russellian Monism	126
4.2.1 The Inscrutability of the Inscrutable	126
4.2.2 The Puzzle of the Combination Principle.....	134
4.2.3 The Significance of the Epistemic Inadequacy	136
4.3. Thomistic Hylomorphism	141
4.3.1 Thomistic Hylomorphism Introduced.....	141
4.3.2 Epistemic Inadequacy of Thomistic Hylomorphism.....	145
4.4. Russellian Monism and Thomistic Hylomorphism	151
4.4.1 Methodological Similarities	151
4.4.1.1 Establishing New Metaphysical Offices	152
4.4.1.2 Fundamental Stuff and Ordinary Individuality	155
4.4.2 Epistemic Inadequacy of RM and TH, Compared	158
4.4.2.1 The Inscrutable and Prime Matter	158
4.4.2.2 Combination Principle and Substantial Form.....	164
4.5. Concluding Remarks	168
5. CAN WE STUDY THE MICRO-LEVEL FACTS IN PAN(PROTO)PSYCHISM? ..	173
5.1. What is a fundamental theory	175
5.1.1 What a fundamental theory should look like: the case of atomic physics	176
5.1.2 Elements of a fundamental theory.....	179
5.2. What we know about the micro-phenomenal (according to pan(proto)psychists).....	182
5.2.1 About the properties of units in pan(proto)psychism	183
5.2.2 About the properties of systems in pan(proto)psychism	187

5.2.2.1 Of The Palette problem	188
5.2.2.2 Of the Structure Mismatch Problem	193
5.3. How is the study of the basic units possible: Transdiction	197
5.4. Transdiction in panpsychism?	214
5.4.1 Analogy (i)	218
5.4.2 Analogy (ii)	226
5.4.2.1 Possible macro phenomenal elements in (ii)	228
5.4.2.2 Reconsider (b) the relations among macro units in (ii)...	231
5.4.2.3 Further reflection on (b) of analogy (ii).....	247
5.5. Concluding Remarks	250
BIBLIOGRAPHY	258

LIST OF FIGURES

Figure	Page
1. Figure 1. A conscious system and its “zombie” counterpart	104
2. Figure 2. Micro and macro systems in transdiction	211
3. Figure 3. Analogy (i) and Analogy (ii)	217
4. Figure 4. Analogy (i)	219
5. Figure 5. Issue with Analogy (i), (c)	223
6. Figure 6: Analogy (ii)	227
7. Figure 7. Physical Space vs. Phenomenal Space	232
8. Figure 8. Issue with Analogy (ii), (b)	237
9. Figure 9. Analogy (ii), Fix 2 to the “mismatch” issue with (b)	242

INTRODUCTION

The problem of consciousness has been an issue in philosophy of mind for decades, and in recent years panpsychism and panprotopsychism (henceforth “pan(proto)psychism”) has gained attention among philosophers who are still dedicated to finding a complete explanation of consciousness. In this dissertation I criticize pan(proto)psychism by examining its metaphysics and epistemology. Concerning metaphysics, I explain the challenges facing pan(proto)psychism and argue that pan(proto)psychism cannot address them well (Ch. 1-3). Concerning epistemology, I question how much we can know about consciousness with pan(proto)psychism (Ch. 4-5).

In Chapter 1, I introduce pan(proto)psychism, the arguments for it, and its main challenge. Panpsychism is the view that there is consciousness in the entities at the fundamental level of the world, such as the basic physical particles. Panprotopsychism is the view that there are special properties (proto-phenomenal properties) at the fundamental level which allows a system to be conscious, although such special properties themselves do not make the micro-level entities phenomenally conscious. Pan(proto)psychism is introduced as an approach to a potentially complete explanation of consciousness. It is also introduced to answer the question concerning the fundamental intrinsic nature of the physical world. The main challenge of pan(proto)psychism is the “Combination Problem”, which concerns how micro-level consciousness or proto-consciousness combines into the macro consciousness that we have.

In Chapter 2, I explain several existing pan(proto)psychist theories and their difficulties. According to Seager's emergentist panpsychism, macro-level consciousness results from the fusion of the consciousness of the micro-level entities according to certain natural laws. According to Goff's phenomenal bonding proposal, macro consciousness arises when entities with micro consciousness are in a sort of special relation known as "phenomenal bonding." According to Coleman's "panqualityism," fundamental entities have no consciousness but "qualities" which are analogous to conscious qualities but are not given to any conscious subjects; consciousness arises when a system forms a representation of these qualities so that these otherwise unexperienced qualities obtain a subject that experiences them. I criticize all three approaches in terms of their metaphysical plausibility.

Given that existing theories of pan(proto)psychism face various metaphysical challenges, in Chapter 3 I discuss integrated information theory (IIT), which appears to be close to panpsychism and has some potential to solve the combination problem. According to IIT, consciousness is identified with the irreducible internal causal power within a system, known as the system's "integrated information." I ask what kind of metaphysics of consciousness is the best fit for IIT. I argue that pan(proto)psychism doesn't fit well with IIT. I also argue that IIT is not a standard version of functionalism, although both IIT and functionalism share the emphasis on causal structures. Finally, I argue that IIT can be interpreted with a version of hylomorphism.

After the metaphysical discussion on pan(proto)psychism in Ch. 1-3, I turn to some epistemological problems of pan(proto)psychism in Ch. 4-5. In Chapter 4, I argue

that Russellian monism faces some epistemic challenges. Since contemporary pan(proto)psychism typically relies on Russellian monism as a component, the challenges to Russellian monism also undermine pan(proto)psychism. Russellian monism is the view that physics only provides information about the causal structures of the world but not about the fundamental intrinsic nature of the world, and that the fundamental intrinsic nature is involved in the grounding of macro consciousness. In Ch. 4, I compare Russellian Monism with Thomistic Hylomorphism. Thomistic hylomorphism attempts to account for the world through the notions of prime matter and substantial forms. Arguably, Thomistic Hylomorphism provides little explanatory illumination. Its notions of prime matter and substantial form are unknowable and unexplanatory. I argue that Russellian monism has methodological similarities to Thomistic hylomorphism. And, given the observation about the epistemic inadequacies of Thomistic Hylomorphism, I argue that Russellian monism has similar epistemic inadequacies.

In Chapter 5, I present further epistemic challenges of pan(proto)psychism, without assuming that pan(proto)psychism presupposes Russellian monism. First, using atomic physics as an exemplar, I argue that a theory about the micro-level reality should account for (i) properties of the micro-level components, and (ii) relations among these components and the resulting properties of a system. Given this, I show that as a matter of fact, pan(proto)psychists have little to say about these. Second, I attempt an explanation of why it is difficult to offer details in accounts of pan(proto)psychism. The issue is a “transdiction” problem of inferring micro-level facts from macro-level observable facts. Reflection on physical science shows that transdiction relies on

“analogies” that compare micro-level facts with macro-level facts, so that macro-level properties and relations can be inferred to exist and operate at the micro-level. Therefore, if pan(proto)psychists are to proceed in the study of micro-level (proto)consciousness and its work, they need analogies in terms of macro-level consciousness and conscious systems. I argue that it is difficult to see if there are analogies between micro-consciousness and macro-consciousness that could help.

PART I

CHAPTER 1

AN INTRODUCTION TO PAN(PROTO)PSYCHISM

1.0 General Introduction.

The title of my dissertation is *Moving Forward on the Problem of Consciousness?*. The question mark expresses my skepticism about the possibility of “moving forward on the problem of consciousness” (Chalmers 1995). The problem of consciousness has been an issue in contemporary philosophy of mind for decades, and in recent years panpsychism/panprotopsychism (henceforth “pan(proto)psychism” to refer to both together), the view that fundamental components of the world such as quarks have consciousness or proto-consciousness, has gained attention among philosophers who are still dedicated to the possibility of a full explanation of consciousness (e.g., Alter & Nagasawa 2012, 2015; Brüntrup & Jaskolla 2017; Chalmers 2015; Goff 2017b).

The growing interest in pan(proto)psychism can be seen in the growing volume of the publications on this and similar topics. Panpsychism has an arguably long philosophical lineage in western civilization. Skrbina (2005) has surveyed the panpsychist elements in western philosophy from pre-Socratic era till the early 2000s, showing a long tradition of panpsychist-leaning ideas among prominent western philosophers such as (upon interpretation) Plato, Aristotle, Spinoza, Leibniz, Kant, James, etc.¹ Some earlier philosophers in the analytical tradition also suggested certain ideas

¹ See also Clarke (2004) for more panpsychist ideas in the history of philosophy.

which are interpreted as in line with pan(proto)psychism (Russell 1927; Feigl 1958; Maxwell 1978). However, for the most part in the (rather short) history of philosophy of mind in the analytical tradition, pan(proto)psychism has few advocates. For example, in Jaegwon Kim's textbook on philosophy of mind (3rd edition published in 2011), the topic of pan(proto)psychism didn't appear. In the 1970s, Thomas Nagel published a short article on panpsychism (Nagel 1979), in which he suggested an argument which would foreshadow the revival of philosophical interest in panpsychism three decades later. But Nagel's article didn't raise broader interest in the topic when it was published.

David Chalmers's early work on the topic of consciousness (1995, 1996, 2003) contributed significantly to the revival of pan(proto)psychism. Chalmers's formulation of the "hard problem" of consciousness (1995) prepared much of the vocabulary and concepts for the ensuing philosophical debates about consciousness. Furthermore, Chalmers's typology of different positions on the problem of consciousness and his appraisal of Russellian monism as a promising position (1996, pp. 153-155, 305-308; 2003, pp. 129-133) indicated a route which will be taken by many future theorists of pan(proto)psychism. However, although Chalmers's work in the 1990s raised a significant amount of interest in the problem of consciousness, interest in pan(proto)psychism didn't closely follow. During the 1990s and the early 2000s, there were some philosophers of mind such as William Seager (1995) and Galen Strawson (1994) who advocated panpsychist ideas or something akin to panpsychism. However, as Skrbina has pointed out, panpsychism wasn't welcomed in mainstream analytic philosophy, and was mostly a target of ridicule (2005, p. 235).

The attitude toward pan(proto)psychism shifted significantly around 2006, when *Journal of Consciousness Studies (JCS)* dedicated an entire issue to panpsychism, with Galen Strawson's (2006) "Realistic Monism" as a target article, followed by 17 comments, most of which were written by mainstream analytic philosophers of mind (including Peter Carruthers, Frank Jackson, Colin McGinn, David Papineau, Georges Rey, David Rosenthal, Peter Simons, Daniel Stoljar, etc.). The issue ended with a long response article by Galen Strawson. The *JCS* issue on pan(proto)psychism initiates the current revival of pan(proto)psychism. Since then, there have been continuous discussions on pan(proto)psychism in academic philosophy (for example: Basile 2010; Blamauer 2011, 2013; Coleman 2012, 2014; Goff 2009a, 2009b, 2017a, 2017b; Morris 2017; Roelofs 2016; Seager 2010, 2012; Shani 2010; Skrbina 2009).

Around the same time, Russellian monism became a significant topic in the discussion about pan(proto)psychism. Early traces of Russellian monism can be found in Chalmers's discussion of "Type-F monism" (2003) and Strawson's notion of "real materialism" (2006), but Russellian monism acquired a more focused attention later in Torin Alter and Yujin Nagasawa's (2012) article on Russellian monism and David Chalmers's "Hegelian synthesis" of physicalism and dualism (presented in Amherst College in October 2012; first published 2015).

With the publication of a collection on Russellian monism (Alter & Nagasawa 2015) and a collection on panpsychism (Brüntrup & Jaskolla 2017), pan(proto)psychism has earned a place for itself in today's mainstream analytic philosophy of mind.

In this dissertation, I want to accomplish two goals.

The first goal is to evaluate the metaphysical plausibility of pan(proto)psychism. Part 1 of this dissertation will be dedicated to this goal. In this part, I will focus mostly on the metaphysics of pan(proto)psychism. Some metaphysical concerns will be used to criticize the existing proposals of pan(proto)psychism.

The second goal is to reflect upon contemporary pan(proto)psychism, which will occupy Part 2 of this dissertation. As a reflection on the recent interest in pan(proto)psychism, I want to raise the concern that pan(proto)psychism might not be intellectually illuminating. First, I will put pressure on Russellian monism (RM), which is the metaphysical schema followed by most pan(proto)psychists. Then I will raise more concerns against the epistemic prospect of pan(proto)psychism, without assuming that pan(proto)psychism presupposes RM. If my arguments in Part 2 of the dissertation are on the right track, then there are reasons to think that contemporary pan(proto)psychism has little to contribute to our understanding of consciousness beyond metaphysical speculations.

The current chapter is an overview of pan(proto)psychism. To prepare for the discussion to come, in this chapter I will explain the philosophical motivations of pan(proto)psychism, especially Russellian monist pan(proto)psychism. To present the motivations, in section 1 and section 2 I will explain two arguments for pan(proto)psychism. In this process, I will explain several key concepts, including panpsychism, panprotopsyichism, and Russellian monism. Panpsychism and panprotopsyichism are slightly different, but they share the same basic motivations. Thus

in sections 1 and 2 I will first explain the arguments for RM panpsychism. At the end of section 2, I will bring panprotopsyichism into the arena. After explaining the arguments for pan(proto)psychism, in section 3 I will explain the major challenge to pan(proto)psychism, namely, the combination problem.

1.1 First Argument for Pan(proto)psychism

In this section I explain the first major argument for panpsychism. This argument proceeds from the concern of avoiding the “epistemic gap” between explanans and explananda, which seems to exist in purely physicalist explanations of consciousness.

An earlier expression of this argument for panpsychism can be found in Thomas Nagel’s (1979) argument. Nagel’s argument contains the following premises:

1. *Material composition.*

Any living organism, including a human being, is a complex material system. It consists of a huge number of particles combined in a special way. Each of us is composed of matter that had a largely inanimate history [before becoming part of an organism]... Anything whatever, if broken down far enough and rearranged, could be incorporated into a living organism. No constituents besides matter are needed. (p. 181)

2. *Nonreductionism.*

[O]rdinary mental states like thought, feeling, emotion, sensation, or desire are not physical properties of the organism... and they are not implied by physical properties alone. (p. 181.)

3. *Nonemergence.*

There are no truly emergent properties of complex systems. All properties of a complex system that are not relations between it and something else derive from the properties of its constituents and their effects on each other when so combined. Emergence is an epistemological condition: it means that an observed feature of

the system cannot be derived from the properties currently attributed to its constituents. (p. 182)²

4. *Realism.*

[Ordinary mental states like thought, feeling, emotion, sensation, or desire] are properties of the organism (p. 182).

Nagel thought that these premises together imply panpsychism, the thesis that every basic bit of matter has some mental property. For, given Realism, mental properties do exist in organisms. Also, according to Material composition, every organism is made only of micro-level particles, which are physical entities. Conscious creatures like us are composed only of particles, which ultimately are microphysical entities. Also, according to Nagel, Nonemergence means that all properties of an organism “that are not relations between it and something else” must “derive from the properties of its constituents and their effects on each other when so combined,” and “the properties of its constituents and their effects on each other” in this case are the properties of the microphysical components and their relations (p. 182). Therefore, the mental properties of an organism must also be able to be epistemically derived from the properties and relations of the micro-level constituents of the conscious creature, since Realism states that mental properties do exist in organisms.³ However, Nonreductionism implies that if all properties

² Nagel denies the existence of “truly emergent properties of complex systems”:

[Emergence] is a reason to conclude that either the system has further constituents of which we are not yet aware, or the constituents of which we are aware have further properties that we have not yet discovered. (1979, p. 182)

³ Nagel is clear that “Emergence is an epistemological condition.” Nagel’s notion of ‘Nonemergence’ is an epistemological one, that there should be epistemic derivation of the higher-level facts from lower-level facts. Thus, ‘Nonemergence’ in Nagel’s idea is close to an a priori entailment thesis.

of the microphysical components of an organism are physical properties, then one cannot “derive” mental properties from facts about the basic constituents of a creature, contrary to what Nonemergence would imply about organisms with mental properties. Therefore the microphysical components must have something other than physical properties. As Nagel puts it,

If the mental properties of an organism are not implied by any physical properties but must derive from properties of the organism’s constituents, then those constituents must have nonphysical properties from which the appearance of mental properties follows when the combination is of the right kind. (1979, p. 182)

By “the organism’s constituents” Nagel seems to mean particles rather than brain parts. Therefore, Nagel is arguing that the basic components of organisms with mental properties must have nonphysical properties. Since there is no reason to think that the basic components of organisms with mental properties (e.g., the atoms that compose a cow) differ from other such components (e.g., the atoms in grass), we should conclude that all basic components of the material world have such nonphysical properties.

This argument from Nagel was published forty years ago, based largely on the premise against “emergence”. Its basic idea is still effective in today’s philosophical context. Some later arguments for panpsychism (Seager 1995, Strawson 2006) also proceed from the idea that consciousness cannot be a product of “radical emergence”. But forty years are long enough for some shift to happen in philosophy. Therefore to present this first argument for panpsychism in contemporary philosophy, I will make some revision to Nagel’s premises.

First, Nagel's "Realism" is about mental states more generally, but contemporary panpsychism focuses on consciousness. Therefore, I would replace Nagel's "Realism" with *Phenomenal Realism*, which states that consciousness exists. The idea is that, intuitively, the existence of conscious experience is self-evident. It seems compelling that there is something it is like to see a rose, to sip coffee, or to see a patch of whiteness. Denying the reality of consciousness is accepting eliminativism, which seems a counterintuitive, excessive reaction against dualism. In this dissertation, I will simply assume that consciousness exists.

Second, Nagel's argument is in terms of "nonemergence" and "nonreductionism", but I prefer to avoid these terms. Although Nagel's intention in using these terms is discernible, since the notions of "emergence" and "reduction" have undergone complex evolution since Nagel wrote this argument for panpsychism, it is inconvenient to present the same idea in terms of "emergence" or "reduction". Thus I would use the more recent term of "epistemic gap". Nagel's Nonemergence would then be replaced by the thesis *No Epistemic Gap*, which states that if all fundamental facts are known, then other facts are in principle derivable from the fundamental facts. In the case of an organism and its consciousness, there is no epistemic gap between the fundamental facts of the components of this organism and the facts about its consciousness: the facts of consciousness must be derivable from facts about the basic components. Nagel's explanation of emergence as "an epistemological condition" (and his later talk of the emergence of mental properties as that "the mental properties of an organism are not

implied by any physical properties” (1979, p. 182)) suggests that his notion of emergence is about the epistemic gap.

I think the epistemic gap can be more conveniently described in terms of a failure of *a priori entailment* between the ontology of the fundamental facts and the (non-fundamental) facts to be understood. A standard way to test such lack of a priori entailment is to see if it is conceivable that the basic facts can hold without the facts in question. Thus, for instance, as it is conceivable that all microphysical facts could obtain without the relevant phenomenal facts (as in a zombie world), all microphysical facts don’t a priori entail all phenomenal facts. This constitutes the epistemic gap between the microphysical and the phenomenal.

Now because I replace Realism with *Phenomenal Realism* and because I express Nonemergence as *No Epistemic Gap*, Nagel’s Nonreductionism should be revised accordingly. Nagel’s Nonreductionism can be interpreted as acknowledging the existence of the epistemic gap between physical properties of the components and mental properties of organisms. Therefore Nonreductionism can be revised in terms of the epistemic gap. Also considering the fact that the contemporary issue with panpsychism is about consciousness more specifically, I would revise Nonreductionism into the thesis of *Physical-Phenomenal Gap*, which states that there is an epistemic gap between microphysical facts and phenomenal facts. The gap is the failure of a priori entailment from the totality of physical facts about micro-entities to the facts about consciousness.

Thus, preserving Nagel’s Material composition, a similar argument for panpsychism, equipped with more recent concepts, can be presented as follows:

Material Composition. “Any living organism, including a human being, is a complex material system. It consists of a huge number of particles combined in a special way... No constituents besides matter are needed.” (Nagel 1979, p. 181)⁴

Phenomenal Realism: Consciousness exists.

No Epistemic Gap: If all fundamental facts are known, then other facts are in principle derivable from the fundamental facts.

Physical-Phenomenal Gap: There is an epistemic gap between microphysical facts and phenomenal facts.

The conclusion would be a thesis of panpsychism, i.e., that there are fundamental facts beyond microphysical facts. These four theses imply that the totality of fundamental facts which includes only microphysical facts cannot account for one section of reality, i.e., phenomenal consciousness. In response, panpsychists propose to add more fundamental facts into the world—in the current case, *phenomenal* properties of the micro-level fundamental entities in one’s head—so that consciousness in creatures might be explained from a larger pool of fundamental facts. Call the smallest entities at the fundamental level “ultimates”.⁵ This move of positing phenomenal properties at the micro-level is comparable to the following move in science: To explain the existence of

⁴ Notice that this is a thesis for the metaphysical sufficiency of the physical particles as fundamental entities, even though “fundamental” wasn’t a popular term in metaphysics when Nagel wrote this passage.

⁵ A term borrowed from Strawson (2006).

electric charge in a composite, one may appeal to the existence of charge in subatomic particles.

Let me go over the justification of these theses briefly. Both (1) *Material Composition* and (2) *Phenomenal Realism* are intuitively plausible. (3) *Physical-Phenomenal Gap* is also supported by well-known intuitions. It seems that, given all facts which physical sciences can provide (including all microphysical facts about the fundamental entities), one cannot fully explain why there is conscious experience of such-and-such features, or why there is consciousness at all. Thus the determination of consciousness by facts about the fundamental entities seems arbitrary. Similar problems have been reported in a series of challenges to physicalism (Chalmers 1996; Jackson 1986; Kripke 1980; Levine 1983). Thus physicalism has been thought by many to be inadequate in explaining consciousness. If the world consists only of microphysical entities with physical properties, then consciousness seems left out due to the epistemic gap between microphysical facts and facts about consciousness. This is also the core of “the hard problem” of consciousness (Chalmers 1995). (4) *No Epistemic Gap* is controversial. But most philosophers attracted to panpsychism worry about the epistemic gap. The worry is that consciousness cannot be explained at all if the epistemic gap exists (Chalmers 2015, Coleman 2012, Nagel 1979). The epistemic gap between the microphysical (as the fundamental) and the phenomenal (as non-fundamental properties in organisms) jeopardizes the neat picture in which all facts are determined by the existence and nature of the fundamental entities. Therefore, philosophers may have a

motivation to avoid the epistemic gap and insist that all basic facts a priori entail all facts.⁶

1.2 Second Argument for Pan(proto)psychism

The second argument for panpsychism proceeds through Russellian panpsychism, which I will explain in due course. It focuses on the need to understand the intrinsic nature of physical beings. I will explain this argument in section 1.2.1. There is also a hybrid argument which integrates both the concern of the second argument and part of the concern of the first argument, which can be called the “two birds, one arrow” argument. This hybrid argument will be explained in section 1.2.2.

1.2.1 Intrinsic Nature

Many philosophers believe that physical sciences only describe extrinsic properties and the spatiotemporal and causal structures, but not the intrinsic natures of the things. For example, in physics the mass of an object is defined in terms of the object’s

⁶ Some philosophers do not worry about the epistemic gap per se, or they do not believe that the epistemic gap entails possibility. As a result, they would not believe that some undesirable ontological consequence would be entailed by an epistemic gap, and so they would have little motivation to adopt panpsychism. Thus, those who are labeled “Type-B” physicalists by Chalmers (2003), or “a posteriori physicalists” by Goff (2011), are generally not attracted to panpsychism. For examples of such physicalists, see Block & Stalnaker (1999), Levine (1983), Hill & McLaughlin (1999), Loar (1990), Papineau (2002), and Perry (2001). For a response to panpsychism from Type-B physicalists, see McLaughlin (2017). For attacks on Type-B or a posteriori physicalism from sympathizers of panpsychism, see Chalmers (2009) and Goff (2011). Physicalists assuming the “phenomenal concept strategy” against dualism belong to Type-B physicalists. Interestingly, philosophers attracted to panpsychism also tend to reject the phenomenal concept strategy for physicalism; see Chalmers (2007) and Goff (2011) for examples.

resistance to change of motion, its gravitational attraction to other objects, etc. For another example, electric charge is defined in terms of an object's tendency to attract or repel other objects, its tendency to change motion in an electromagnetic field, etc. Physical properties so defined are suitable for understanding objects as they are causally related to other things in spacetime, but such properties are dispositional, explicated in terms of relations rather than intrinsic features. As Russell puts it, speaking about properties of matter:

A piece of matter is a logical structure composed of events; the causal laws of the events concerned, and the abstract logical properties of their spatio-temporal relations, are more or less known, but their intrinsic character is not known. (Excerpts from *Analysis of Matter*, in Alter & Nagasawa 2015, pp. 30-1)

A similar idea is expressed by Frank Jackson (1998):

“When physicists tell us about the properties they take to be fundamental, they tell us what these properties *do*. This is no accident. We know about what things are like essentially through the way they impinge on us and our measuring instruments... [this] suggest the possibility that (i) there are two quite different intrinsic properties, *P* and *P**, which are exactly alike in the causal relations they enter into, (ii) sometimes one is possessed and sometimes the other, and (iii) we mistakenly think that there is just one property because the difference does not make a difference (as the point is put in information theory). An obvious extension of this possibility leads to the uncomfortable idea that we may know next to nothing about the intrinsic nature of the world. We know only its causal cum relational nature.” (pp. 23-24)

Chalmers has expressed the same point when he interprets the Russellian view of physics:

[C]lassical physics tells us a lot about what mass does — it resists acceleration, attracts other masses, and so on — but it tells us nothing about what mass intrinsically is. We might say that physics tells us what the mass role is, but it does not tell us what property plays this role. (Chalmers 2015, pp. 253-4)

Therefore, it seems widely acknowledged that we understand the theoretical terms to the extent that the properties they refer to have their place in a world described through mathematical structure. But we are ignorant of the intrinsic properties playing the roles in this structure. However, it's natural to think that such dispositional properties should have intrinsic bases in the objects themselves. For if two things have different tendencies of behavior, it must be because the two things themselves are different. There must be something in the internal constitution of an object that makes it behave one way or another. Consequently, there is a question about what the fundamental intrinsic properties of the physical world might be.

Panpsychism, especially the Russellian versions, partly aims at answering this question about the fundamental intrinsic nature of the physical. Panpsychism addresses the demand for understanding the intrinsic nature of things by proposing that consciousness is the fundamental intrinsic nature of physical entities. The idea is that we seem to have no clue about the fundamental intrinsic properties. But in consciousness, we encounter phenomenal properties as something intrinsic (Strawson 2006; Seager 2006). It seems intuitive that my conscious experience instantiates in me intrinsically, in the sense that it's fully determined by the state within me (as a conscious organism): insofar as my internal state stays the same, I can have the same experience as I have right now even if I am the only existence in the world (Seager 2006, p. 143). It is also plausible that my perfect duplicate would have the same conscious experience as I do.⁷ Therefore if the

⁷ These are not conclusive proofs that consciousness is intrinsic, for no one really has a perfect criterion of intrinsicity. But these ideas should suffice to make the intrinsicity of consciousness plausible.

fundamental entities that compose my body have some intrinsic nature, it seems promising to suppose that their intrinsic nature might be similar to phenomenal properties in consciousness. Consciousness apparently has provided access to the intrinsic nature of a part of the physical world.

Thus we reach *Russellian Panpsychism*, which is the idea that phenomenal properties are the intrinsic nature of the fundamental physical entities. Russellian panpsychism is a version of a more general metaphysical thesis, known as *Russellian Monism*.⁸

Russellian monism (RM): there are fundamental intrinsic properties (aka. the “inscrutable”)⁹, the categorical bases of paradigmatically physical properties, of which physical sciences cannot inform us what they are; some of the fundamental intrinsic properties that ground paradigmatic physical properties also ground consciousness.

RM is the metaphysics used by most supporters of panpsychism. The idea is that physical sciences tell nothing about the fundamental intrinsic nature, but *something* must

⁸ The idea is called “Russellian” because Russell has suggested a similar idea that relates consciousness to intrinsic nature. To account for “percepts” (experiential, conscious qualities, such as in red sensation), Russell distinguished causal knowledge provided by physics from knowledge of intrinsic characters, and assigned the place of intrinsic characters to percepts (Russell 1927, p. 384). For contemporary development in Russellian Monism, see Alter & Nagasawa 2012, 2015; Chalmers 1996, 2015; Goff 2017b; Maxwell 1978; Pereboom 2011; Seager 2006.

⁹ The term “inscrutable” is used by Alter and Nagasawa (2012), who have attributed this term to Montero (2015, p. 213). Chalmers (2015) uses the term “quiddities”, and Pereboom (2011, 2015) uses the phrase “absolutely intrinsic properties” to refer to the same kind of things.

be doing the job of the fundamental intrinsic properties that ground paradigmatically physical properties. So, maybe phenomenal properties are doing this job. As put by Seager,

... matter must have an intrinsic nature to ground the relational or structural features revealed to us by physical science. We are aware of but one intrinsic property of things, and that is consciousness. It is plausible to assert physicalism — we are physical beings and our consciousness is a feature of certain physical structures [...] Therefore, consciousness is an intrinsic property of matter. (Seager 2006, p. 136)

And similarly, Goff

Physics tells us much about the dispositions of fundamental natural entities, but leaves us completely in the dark about their categorical nature. [...] Everything natural science has to tell us about electrons concerns their behaviour [...] The only thing we know for certain about the categorical nature of natural entities is that at least some of them, for example me and you, are conscious. (Goff 2017a, p. 283)

Goff then suggests that if we “suppose that the categorical nature of fundamental particles, such as electrons and quarks, is constituted of some form of consciousness”, then we can have a simple and unified worldview, in which “the nature of macroscopic things is continuous with the nature of microscopic things.” (Ibid.)¹⁰

1.2.2 Two Birds, One Arrow

We have seen two arguments for panpsychism. The first argument (section 1) is based on the idea that because “emergence” should be rejected, explaining consciousness requires that the micro components (ultimates) of conscious creatures have more than microphysical properties. The second argument (section 2.1) is based on the idea that

¹⁰ See also Alter & Nagasawa 2012, 2015; Chalmers 2003, 2015; Goff 2017b.

something should play the role of the fundamental intrinsic nature. There is also a hybrid argument in the literature, an argument for panpsychism which integrates the ideas of both aforementioned arguments. In this section I will explain this “two birds, one arrow” argument.

This argument uses one thesis (“arrow”) to answer two concerns (“birds”). The first concern is about the puzzling relationship between consciousness and the physical world, which is similar to that of the first argument (section 1) but framed differently. The puzzling relationship between consciousness and the physical can be understood as a tension between two desiderata. On the one hand, it is better to account for consciousness as different from what we know through physical sciences. There are intuitions suggesting that conscious experiences are not simply physical properties.¹¹ On the other hand, it is also better not to distance consciousness too much from the physical world, since a dualist picture that lacks a unifying picture of mind and body is undesirable.¹² Because of this tension, neither physicalism nor dualism in their traditional form seems satisfactory. Physicalism tends to fail the first desideratum, because it can hardly do full justice to the distinctiveness of consciousness. Dualism tends to fail the second desideratum, because its ontology contains a realm which seems separated from the realm of the physical world.

¹¹ This is close to the *Physical-Phenomenal Gap* thesis.

¹² Another concern is the causal closure of physics. It seems that once we know all the neurophysiological events that lead to a conscious intentional bodily movement, we know all that is needed that causes this move. Thus, barring epiphenomenalism and overdetermination, the causal relevance of consciousness must somehow depend on the causal power of the physical, because consciousness is not supposed to add more causal power to the world, but it is presumably relevant to the causal process.

The second concern is the same as that of the second argument (section 2.1), about understanding the intrinsic nature of the physical. As explained in the last section, the concern is that the physical sciences leave out the fundamental intrinsic nature (aka. “the inscrutable”). I will not repeat the details here.

Now comes the “arrow” which addresses two concerns at once, i.e., Russellian Panpsychism. The core of Russellian Panpsychism is that all properties, including paradigmatically physical properties, must have some fundamental intrinsic nature, and such fundamental intrinsic nature is phenomenal property (section 2.1). The way Russellian Panpsychism addresses the concern of intrinsic nature is straightforward: it proposes that the intrinsic nature is consciousness. How does it address the other concern, i.e., the puzzling relationship between consciousness and the physical?

In the first place, how to do justice to the distinctiveness of consciousness? According to Russellian panpsychism, explaining consciousness needs the inscrutable, which is the fundamental intrinsic nature. However, physical properties as studied and described by science are considered as causal-dispositional or structural-dynamic properties, rather than fundamentally intrinsic. Therefore according to Russellian Panpsychism, consciousness reveals a domain that is of a different category from physical properties, and so consciousness and physical properties are distinct.

How does Russellian Panpsychism accommodate consciousness in a unifying metaphysical picture? Russellian panpsychists can hold that in the Russellian metaphysical picture, the inscrutable and physical properties are distinct but closely related owing to the metaphysical relationship between the categorical base that *grounds*

a disposition and the disposition that is *grounded* by the base, for physical properties as dispositions are realized by phenomenal properties as their categorical base. Therefore, even if Russellian Panpsychism distinguishes the phenomenal inscrutable from physical properties, Russellian Panpsychism nevertheless “encourages a tighter connection between the phenomenal and the physical than traditional versions of dualism.” (Alter & Nagasawa 2012, p. 82)¹³

Therefore the “arrow” of Russellian Panpsychism can answer both the concern about the fundamental intrinsic nature and the concern about the consciousness-physical world relationship. But why treat Russellian Panpsychism’s answers to the two “birds” as one argument rather than two? Because the key idea in this case is to use the answer to one concern to address the other concern. Russellian Panpsychism uses the inscrutable understood in terms of consciousness to answer the puzzle of consciousness. The totality of physical science seems insufficient for explaining consciousness, but this failure of explanation is, according to Russellian panpsychists, due to our ignorance of the inscrutable, which serve as the intrinsic base of all concrete properties, including paradigmatic physical properties. According to Russellian Panpsychism, the inscrutable must also have a role in producing consciousness. If the physical sciences, which may completely describe all structural-dynamic features of the world cannot explain

¹³ Concerning mental causation and the closure of physics, to avoid overdetermination, RM theorists can acknowledge that causal relations exist only at the level of the causal-dispositional properties, which explains why causality is closed under physics. However, to avoid epiphenomenalism, RM theorists can hold that causal-dispositional properties are grounded in fundamental intrinsic nature, i.e., phenomenal or protophenomenal inscrutable that (also) ground consciousness. Thus RM seems to allow consciousness to be causally relevant without disrupting the causal order of physics.

consciousness, then consciousness must be explained in part by the fundamental intrinsic nature, which is missed by a complete description of structural-dynamical facts.¹⁴

In sum, Russellian Panpsychism utilizes an expanded metaphysics of properties to open a category (i.e., fundamental intrinsic properties), where consciousness or the “seeds” of consciousness could be planted, a category whose membership excludes what is commonly understood as the causal or physical aspects of the world. Russellian Panpsychism then accounts for the relationship between consciousness and physical properties with the picture of a metaphysical dependence between the new category of properties (fundamental intrinsic nature, understood as consciousness) and the category of more familiar physical properties (causal-dispositional nature).

1.2.3 Panpsychism and panprotopsyism

I have shown in the earlier sections that panpsychism seeks to address both the concern about the physical-consciousness relationship and the question about the intrinsic nature of the physical. More precisely, panpsychism offers the phenomenal properties of some ultimates in conscious creatures as the fundamental properties that are missing in

¹⁴ Alter & Nagasawa put the underlying logic as a hiring-and-job-seeking argument:

Physics wants to hire help: it wants to employ something outside its purview to ground the structure it so elegantly describes... Consciousness wants a job: it wants to be integrated into nature by playing a role in the causal nexus known as the cosmos. Seen in this way, a unified solution suggests itself: consciousness can be employed to ground fundamental physical relations — which is what Russellian monism says, with the one qualification that on some versions of Russellian monism it is not consciousness itself but its components (protophenomenal properties) that ground the properties found in physics. (2012, p. 89)

the traditional physicalist worldview. The added fundamental properties may close the epistemic gap between the fundamental facts and facts about consciousness (section 1). Also, these properties may serve as the fundamental intrinsic nature of physical things (section 2.1). The ideas of these two arguments can also be integrated to form a hybrid argument (section 2.2).

However, it is debatable whether phenomenal properties of ultimates are necessary to satisfy the two philosophical concerns in the two arguments. Concerning the epistemic gap, the goal is to find a set of the fundamental facts so that these facts can a priori entail the existence of consciousness. Microphysical properties of ultimates as the only members of fundamental properties are apparently insufficient for this task, and so phenomenal properties of ultimates are enlisted. But it might be thought that the task can be accomplished without phenomenal properties of ultimates. If there is a fundamental property X of the ultimates such that facts about X and other fundamental (microphysical) facts can a priori entail that there is consciousness, then X fits the bill even if X is not a phenomenal property (that is, if there is no such thing as “what it is like” to instantiate X.) Since X is not phenomenal but grounds phenomenal properties, X is a “protophenomenal” property. Thus, contra panpsychism, one may opt for protophenomenal properties in conscious creatures to close the epistemic gap.

Similarly, one may choose protophenomenal rather than phenomenal properties of ultimates to account for the fundamental intrinsic nature. Insofar as protophenomenal properties are intrinsic properties of the ultimates, it doesn't seem that they are worse than phenomenal properties as the intrinsic nature of physical things. Recall that

Russellian monism (RM) requires that something be the fundamental intrinsic properties and that some of them must ground consciousness. But RM doesn't state that such fundamental intrinsic properties must be phenomenal. Maybe some non-phenomenal special properties can fulfill this job. Thus an RM theorist has two options, depending on her view about the fundamental intrinsic nature (the inscrutable). If she takes the inscrutable as phenomenal consciousness, then her RM is Russellian Panpsychism. Or, she may consider the inscrutable as "protophenomenal" rather than phenomenal. Then the resulting version of RM is Russellian *Panprotopsyism*.

Therefore both arguments for panpsychism (section 1, section 2.1) and the hybrid argument (section 2.2) can also be used to support panprotopsyism. Panpsychism and panprotopsyism belong to the same family of metaphysical views. Both address the concern about the place of consciousness in the physical world. Both look for the solution at the level of ultimates; in particular, both appeal to the intrinsic properties of ultimates. Finally, both are compatible with RM. The difference lies in the different intrinsic properties that they attribute to ultimates, which brings about the further difference in their accounts of how fundamental intrinsic properties ground consciousness. In Ch. 2, I will explain and criticize several theories of panpsychism and panprotopsyism. In this introductory chapter, I will be satisfied with introducing and clarifying these two positions.

Panpsychism: the fundamental building blocks of the universe (ultimates) have phenomenal properties. For example, if M-strings are ultimates, then each M-string has consciousness, in the sense that it has some phenomenal properties.

Panprotopsyichism: the fundamental building blocks of the universe (ultimates) do not have phenomenal properties, but they have some special properties (i.e., protophenomenal properties) that enable them to compose entities that have consciousness/phenomenal properties.

To summarize sections 1 and 2, pan(proto)psychism might offer the resource to address both the puzzle concerning the relationship between consciousness and the physical world, and the question concerning what the intrinsic nature of things is like. It addresses the former by expanding the ontological base to include (proto)phenomenal properties of the micro-entities, with the hope that macro-level consciousness is a priori derivable from the fundamental (proto)phenomenal nature of micro-entities. It addresses the latter concern by assigning the intrinsic status to the (proto)phenomenal features of the micro-entities, in the hope that the (proto)phenomenal properties of micro-entities are the intrinsic nature of physical things.

In the next section, I explain the major threat to pan(proto)psychism, i.e., the combination problem. Since the combination problem is concerned with the relationship between the consciousness we know of at the macro-level and the (proto)phenomenal state of affairs at the micro-level, talking about this problem will need constant reference to things at the micro- or macro-level. For the convenience of discussion, I will use some terms which are prefixed with “micro-” or “macro-”:

- *micro-subjects*: the beings that have the status of being micro-level subjects of experience. In the case of panpsychism which locates consciousness in ultimates, I will use “micro-subjects” to refer only to ultimates.

- *macro-subjects*: the beings that have the status of being macro-level subjects of experience. For example, human beings.
- *micro-experience/micro-consciousness*: experience/consciousness of a micro-subject.
- *macro-experience/macro-consciousness*: experience/consciousness of a macro-subject.

The following terms will also appear, but less frequently:

- *micro-qualities*: the phenomenal qualities of micro-experience / consciousness.
- *macro-qualities*: the phenomenal qualities of macro-experience / consciousness.

1.3 The combination problem

Broadly speaking, the combination problem is the challenge of explaining how micro-level (proto-)conscious entities combine to make macro-level consciousness. Even if we suppose that experience is present in the fundamental entities, it's still hard to see how these entities can combine to give rise to the experiences that we have. This problem has been the most serious challenge to pan(proto)psychism. So an essential task in the development of pan(proto)psychism is to solve the combination problem.

1.3.1 The combination of subjects

When most contemporary panpsychists look back, they trace the origin of the combination problem to William James's *Principle of Psychology*. James observed that it is unclear that a composite made of conscious ultimates will experience anything. He

argued with an analogy of a collection of a hundred people, each with his/her own experience. It doesn't seem that the collection is a conscious being:

Take a hundred of [conscious elemental units that have feelings], 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 it with them, and one could never deduce the one from the others, or (in any intelligible sense) say that they evolved it. (James 1891, p. 160)

In other words, given entities that are subjects of experience, it doesn't seem to follow that the composite made of these entities would be a subject of experience. This analogy illustrates the "subject-summing problem", i.e., the concern that micro-subjects don't combine to form macro-subjects, and thus ultimates don't explain the existence of consciousness in creatures like human beings. (Chalmers 2017, p. 182)

The subject summing problem can be construed as an epistemic gap between the basic facts of panpsychism (physical facts + intrinsic phenomenal properties of ultimates) and the facts about macro-level consciousness. This diagnosis has been summarized by Phillip Goff in the following principle:

No Summing of Subjects (NSS): It is never the case that the existence of a number (one or more) of subjects of experience with certain phenomenal characters a priori entails the existence of some other subject of experience. (Goff 2009a, p. 302)

According to this principle, micro-subjects *could* form macro-subjects. It is not claimed that the formation of macro-subjects is an inconsistent notion. The issue is that the

existence of micro-subjects cannot *a priori entail* the existence of a macro-subject. Thus, the combination problem is construed as a challenge from an epistemic gap between micro-subjects and macro-subjects.

Based on this principle, Chalmers (2017) has put the problem into the following argument:

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

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

[Therefore] For any group of subjects (with certain experiences), it is possible that the subjects in *S* exist (with their experiences) and no other subjects exist. (p. 187)¹⁵

Whereas Goff puts the worry of the epistemic gap in terms of a lack of a priori entailment, Chalmers expresses the same concern in terms of conceivability in the argument above. This epistemic gap challenges panpsychism just as another epistemic gap challenges the traditional physicalist worldview. If panpsychists are entitled to find fault with a picture of merely physical micro-entities because of the epistemic gap between the fundamental facts of this picture and the macro-level consciousness, then panpsychists should reject their own position because there is an epistemic gap between

¹⁵ Goff (2017a, pp. 291-2) has a similar argument concerning the lack of summing of subjects, but Goff has added concerns about the “transparency” of consciousness into the argument to enhance the premise from conceivability to possibility. Since for most philosophers, if they don’t believe that conceivability implies possibility then they are more likely to settle down on physicalism regarding consciousness, and are less likely to consider panpsychism anyway, I assume that most philosophers who care about panpsychism are those who find conceivability a more-or-less trustworthy guide in metaphysics, and therefore won’t raise much issue on inferring the possibility of micro-phenomenal zombies from their conceivability. Therefore, I think that Chalmers’s argument against subject summing is sufficient as a problem of panpsychism.

the fundamental facts of the panpsychist worldview (i.e., micro-entities as subjects) and facts about consciousness.

1.3.2 The combination of qualities and structures

Besides the relatively easy-to-recognize combination problem concerning the summing of subjects, Chalmers (2017) has explained two other problems of mental combination, as follows:

The quality combination problem: How do microqualities combine to yield macroqualities? [...] It is natural to suppose that micro experience involves microqualities, which might be primitive analogs of macroqualities. How do these combine? (p. 183)

The structure combination problem: How does microexperiential structure (and microphysical structure) combine to yield macroexperiential structure? Our macroexperience has a rich structure, involving the complex spatial structure of visual and auditory fields, a division into many different modalities, and so on. How can the structure in microexperience and microstructure yield this rich structure? (Ibid.)¹⁶

The quality combination problem is easier to grasp. A more specific version of the quality combination problem is vividly put as “the palette problem”, i.e., why are there *so many* phenomenal qualities in macro-experience, such as reddishness, greenishness, bitterness of taste, pain in the wound, etc.? How can the varieties of macro-qualities be explained in terms of (probably) a small varieties of micro-qualities—of (probably) a small varieties

¹⁶ Chalmers (2017) has also mentioned several other problems of mental combination, including “the unity problem”, “the boundary problem”, “the awareness problem”, and “the grain problem.” In this introduction to the combination problem, I consider “the grain problem” as an issue of structure mismatch.

of ultimates? By analogy, from only a boring color of grey, it is impossible to get a colorful spectrum of qualities.

The structure combination problem is more difficult to appreciate. As Seager has commented, the issue seems “somewhat obscure.” (2017, p. 245) The problem is about how micro-experiences of the ultimates combine to make macro-experiences of a certain structure that we seem to observe in our consciousness. To appreciate this problem, one should first find the structure of macro-experience puzzling, and then one might be motivated to ask how the structure of macro-experience is produced from the combination of micro-experiences. The puzzle about macro-experiential structure is highlighted by noting the “mismatch” between the macro-experiential structure and macro-physical structure. That is the difference between the apparent structure of our experience and the structure of the nervous system.

Michael Lockwood has described the mismatch in the following passages:

[T]he phenomenal objects of introspective awareness are far less finely structured than are any plausible physiological correlates. Consider, for example, a phenomenally flawless auditory experience, of a note, say, on a violin. Its physiological substrate, presumably, is a highly structured, not to say messy, concatenation of changes in electrical potential within billions of neurons in the auditory cortex, mediated by the migration of sodium and potassium ions across cell membranes, and of molecules of transmitter substances within the chemical soup at the synapses. How do all these microstructure discontinuities and inhomogeneities come to be *glossed over*, in such a way as to generate the elegant perfection of auditory phenomenology that we associate with the playing of a Yehudi Menuhin? How are we to make philosophical sense of such phenomenological *coarse-graining*? (1993, p. 274)

The phenomenal contents of awareness don’t appear to have the *right kind* of structure; what is ostensibly lacking, here, is even the most approximate isomorphism between states of awareness and the underlying physiological goings-on [...]. In particular, three-dimensional spatial arrangement, and changes

therein, seem central to all physical structure. Where, then, are their phenomenological counterparts? (Ibid.)

Thus there are at least two apparent differences between the macro-experiential structure and the macro-physical structure. One is concerned with the “grain” of experiences and neural events. The other is about the apparent lack of spatial structure in experience.

Assuming that we understand the structure of physical things, the above reflection on the structure of macro-experience might suffice to show that there is something novel about macro-experiential structure which is worth exploring. Thus, for panpsychists who take macro-experience to be the result of micro-experience, there is a question about how micro-experiences combine to produce the structure of macro-experiences, and how this combination is different from the combination of physical things.

Chalmers deepens the puzzle by presenting it as something close to a contradiction in panpsychism, in terms of the “structure mismatch problem”:

Macrophysical structure (in the brain, say) seems entirely different from the macrophenomenal structure we experience. Microexperiences presumably have structure closely corresponding to microphysical structure (this is especially clear on a Russellian view), and we might expect a combination of them to yield something akin to macrophysical structure. How do these combine to yield macrophenomenal structure instead? (Ibid., p. 183)

The crucial move is in the claim that the microexperiential structure “closely correspond to” the microphysical structure (Ibid.). In other places, Chalmers also talks about microexperiential structure being “isomorphic” to microphysical structure (Ibid., p. 191, p. 206). Given the isomorphism at the micro-level, because macro-experience is a result

of micro-experience while the macro-physical is a result of the micro-physical, panpsychists should believe that macro-experiential structure is also “isomorphic” to macro-physical structure. Therefore, if it turns out that macro-experiential structure in our phenomenology “mismatches” with macro-physical structure in the brain, then apparently there is a contradiction in panpsychism.

The comparison between macro-experiential structure and macro-physical structure is sufficient to single out the apparent structure of macro-experience as a phenomenon to be explained. Chalmers’s claim about the isomorphism between micro-experiential and micro-physical structures is meant to sharpen the concern of “structure combination” by problematizing it. But Chalmers’s *problem* of structure mismatch is not entirely clear. First, it’s unclear what “microexperiential structure” is. Second, it’s unclear what is for microexperiential structure to be isomorphic to microphysical structure. And third, it’s unclear what is for microexperiential structure to “yield” macroexperiential structure. In any case, to the extent that panpsychists recognize that macro-experience has certain structural features, they face the task of explaining it in terms of the combination of micro-experiences.

The subject summing problem by far seems to be the most discussed in the family of the combination problem. Solutions to other problems, including the “quality combination problem” and “structure combination problem”, all come after particular proposals to solve the subject summing problem (Brogaard 2017; Seager 2017; Coleman 2017; Goff 2017b), and are thus consequences or by-products of particular solutions to the subject summing problem. Therefore, when I discuss the combination problem as a

metaphysical challenge (Ch. 2, Ch. 3), my main focus will be the subject summing problem.

1.4. Chapter Conclusion

In this chapter, I have explained the arguments for pan(proto)psychism, which has Russellian monism as its natural ally. Pan(proto)psychism is motivated by commitments to some received wisdoms in contemporary metaphysics. It is also an attempt to break the standoff between traditional physicalism and dualism. In this introduction I have also explained the combination problem of pan(proto)psychism. In the next chapter (Ch. 2), I review and criticize several proposals to solve the combination problem.

CHAPTER 2

OF SOME PROPOSALS TO SOLVE THE COMBINATION PROBLEM

In the opening chapter of this dissertation (Ch. 1), I introduced panpsychism and panprotopsyism, explained some major arguments for them, and explained the main challenge to pan(proto)psychism, namely the combination problem. To solve the combination problem is to explain macro-consciousness as a consequence of micro-consciousness or protophenomenal properties at the micro level. The combination problem is also a source of motivation for developing different approaches to pan(proto)psychism. In this chapter, I will examine several versions of pan(proto)psychism that are developed to address the combination problem.

A basic division among the various approaches to pan(proto)psychism is between *Constitutive Pan(proto)psychism* and *Emergentist Pan(proto)psychism*. The basic ideas of these views can be characterized roughly as follows:

Constitutive Pan(proto)psychism: Macro-consciousness of a larger-scale entity is nothing over and above the distribution and relation of the (proto)phenomenal properties (together with all other micro-level facts) of the ultimates.

Emergentist Pan(proto)psychism: The sort of pan(proto)psychism that is not constitutive pan(proto)psychism. That is, there is more to macro-consciousness than

all the micro-level facts (including facts about the (proto)phenomenal properties of the ultimates.)

Recently, several philosophers have developed pan(proto)psychism in different directions, which has resulted in different accounts of pan(proto)psychism that are fairly representative of the different approaches as categorized under the classification of panpsychist/panprotopsylist, and constitutive/emergentist approaches. Goff (2017a, 2017b) has offered a “phenomenal bonding” proposal that could represent constitutive panpsychism. Seager (2010, 2017) has offered a “infusion” view that could be representative of emergentist panpsychism. While Goff and Seager follow the path of panpsychism, Coleman (2012, 2014, 2017) has developed a version of panprotopsylist, which has been termed “panqualityism”. (Chalmers 2015, p. 270) I think all these metaphysical proposals face serious problems. Therefore, pan(proto)psychists are yet to find a clear and plausible metaphysical account.

This chapter will proceed as follows. In section 1, I focus on emergentist panpsychism and evaluate it. In particular, I will present and criticize Seager’s version of emergentist panpsychism. In section 2, I consider and evaluate constitutive panpsychism by explaining and criticizing Goff’s phenomenal bonding proposal. In section 3, I consider panprotopsylist by examining and criticizing Coleman’s panqualityism. Section 4 concludes.

2.1 Emergentist panpsychism

In this section I will explain emergentism further (section 1.1.) Then I will explain and evaluate Seager's infusionist version of emergentist panpsychism (section 1.2).

2.1.1 Emergentism

Loosely speaking, emergence happens when a property (call it "emergent property") that is instantiated at a higher-level is not instantiated at the lower-level.

There could be notions of emergence of different "strength". The basic distinction between a "strong" and a "weak" version of emergence come in whether the emergent property at the higher-level can be deduced from the lower-level properties (call them "base properties"). For example, according to Brogaard, 'strong emergence' requires that "truths about the emergent property are not deducible, even in principle, from the low-level phenomenon." In contrast, 'weak emergence' happens when "truths about the emergent property are deducible, at least in principle, from the low-level phenomenon" (2017, pp. 131-132). Similarly, according to Seager, 'radical emergence' requires special natural laws which "must be appealed to in addition to the laws of the submergent [low-level] domain", which means that 'radically emergent properties' are not deducible from low-level facts alone. In contrast, in 'conservative emergence', a higher-level property of an object "follows logically from the specification of the properties of [this object's] constituents (plus environment) and the laws governing these submergent properties" (2017, p. 232).

According to standard emergentist account of consciousness as a phenomenon based on neurophysiology or psychological functions, consciousness comes into being when an organism is complex enough to manifest some high-level structural/functional features, such as the ones enabling cognitive capacity. Consciousness is said to “emerge” in this view, because below the level of complexity of a conscious creature, there is supposedly no consciousness, and so consciousness seems to emerge at a high-level. At this point, given that there is a general distinction between strong/radical and weak/conservative emergentism, one can distinguish between strong/radical and weak/conservative emergentist accounts of consciousness.

- According to weak/conservative emergentism of consciousness, facts about consciousness simply follow logically from the structural/functional features of the complex organism. This emergentist view of consciousness is the most naturally acceptable for standard physicalists. The explanation of consciousness under this view is essentially no different from an explanation of the locomotion of an entire animal body.
- According to strong/radical emergentism, consciousness must be derived from the relevant high-level structural/functional features *plus* something else. Some special natural law is needed to specify what consciousness would arise under what psychological or neurophysiological conditions.

When philosophers wonder about emergentism, they are more interested in the possibility of “strong” or “radical” emergence. Strong emergence can be stated in terms of a natural law connecting the lower-level properties and the higher-level emergent properties. This idea can be seen in Van Cleve’s (1990) definition of emergence in terms of supervenience:

If P is a property of w, then P is emergent iff P supervenes with nomological necessity, but *not* with logical necessity, on the properties of the parts of w. (1990, p. 222)

Thus, a property that is a case of “weak” or “conservative” emergence “logically” follows from facts about the low-level phenomena, whereas a “strong” emergent property follows through “nomological necessity”. But positing natural laws of strong emergence might be inadequate. If the higher-level emergent properties cannot receive an explanation in terms of the mechanisms at the lower level, this emergence sounds like magic, for it has a sense of *Ex Nihilo*, i.e., “something” (the emergent property) comes out of “nothing.” Thus, most philosophers tend to reject strong emergence. Furthermore, more importantly for the current discussion on the plausibility of panpsychism, strong/radical emergentism is what panpsychists want to avoid (Ch. 1, section 1). Thus, if a panpsychist is to pursue an emergentist approach, a desired account of emergentist panpsychism would be weak/conservative emergentism.

One may wonder what makes the difference between constitutive pan(proto)psychism and weak emergent pan(proto)psychism. I admit that emergentist panpsychists, if they want a theory that is intelligible, want a “conservative” theory, and it

does seem that the most obvious option is to choose constitutive panpsychism because it is conservative enough. But there are differences between constitutive panpsychism and weak emergentist panpsychism. The difference can be clarified by the following further explication of constitutive panpsychism and weak emergentist panpsychism:

- According to Constitutive Panpsychism, the existence of micro-consciousness with the right sort of micro-level relation/organization of the micro-entities *constitutes* macro-consciousness; micro-level facts *alone* entail macro-level facts. (E.g., a collection of H₂O molecules *is* water.)
- According to Weak Emergentist Panpsychism, there is an account of macro-consciousness in which micro-level consciousness are involved, but micro-level facts alone do not entail macro-level facts. A system of which all the relevant micro-level facts obtain is not identical to a system with macro-consciousness. Thus, naturally, some extra factor is needed to generate macro-consciousness from the micro-level facts, so that the facts about macro-consciousness follow from the micro-level facts (including those of micro-consciousness) plus that extra factor. (Seager's "conservative" emergentist panpsychism is of this sort.)

Furthermore, it's worth noting that if one prefers emergentist panpsychists, one is probably dissatisfied with constitutive panpsychism. Thus, Seager rejects constitutive panpsychism by rejecting mental combination (2017, pp. 236-238). Seager doesn't

believe that the presence of micro-consciousness constitutes macro-consciousness. Thus, weak emergentists such as Seager are looking for a middle way between constitutive panpsychism and strong/radical emergentist panpsychism.

In what follows, I will examine Seager's attempt to propose an emergentist panpsychism that is weak/conservative.

2.1.2 Of Seager's Panpsychist Infusion

In this section I will introduce and evaluate Seager's infusionist panpsychism.

According to Seager's idea of combinatorial infusion (2010, 2017), if micro-consciousness in micro-subjects accrue in the right way, then following some contingent natural law, a subject of macro-consciousness would emerge and "substitutes" the collection of micro-subjects. The infusionist law would state that an aggregate of conscious things with certain phenomenal properties would fuse into one conscious thing with a new phenomenal property. In Seager's term,

[Combinatorial infusion] follows James's suggestion... that the transmutation from the hypothetical micro psychic features assigned to the fundamental entities of the physical world to the macrostates of consciousness... requires the generation of a new state which infuses its precursors, or, to use Whitehead's term, substitutes a new state for the set of precursor states. (2017, p. 238)

Since the natural law that connects micro-consciousness and macro-consciousness is contingent, Seager's infusionist view may explain why the failure of subject-summing seems conceivable (Seager 2017, p. 243). Since subject summing in infusionism involves contingent laws, one may know all facts about the micro-consciousness without knowing

the laws of infusion, and is thus able to conceive of the failure of subject summing. Also, given that strong/radical emergence is often rejected as implausible, Seager doesn't intend this contingent law of infusion to become a case of strong/radical emergence. He insists that infusion should be "a form of conservative emergence", for which there should be "some kind of intelligible link between the micro psychic features assigned to fundamental physical entities and resultant complex states of consciousness." (Ibid., p. 238) So, in what follows, we shall examine Seager's proposal of infusion laws and ask whether it offers a plausible "conservative emergence" account.

What would an infusion law that connects macro phenomena with micro-level facts look like? Levine (2017) has observed that Seager's infusion law might work differently from other natural laws that are involved in the explanation of macro-level phenomena. He contrasts Seager's infusionist law with the physical laws about water molecules:

Take the property of being liquid at room temperature. We want to say that this property is realized by (grounded in) the various properties of the molecules and the laws governing their behavior. The way you then justify an a priori entailment thesis — required by the claim [...] that realization, or constitution, entails an a priori entailment from descriptions of the micro to descriptions of the macro — is, for example, to identify liquidity (a priori) with some functional/causal role property from which it follows that sufficiently large aggregations of H₂O molecules exhibit this macro-property. But I don't see how this would work for combinatorial infusion laws. What sort of identity claim of the macro-property would support such an inference? Of course, you can just attribute to the micro-phenomenal properties the disposition to infuse in this particular way, but that seems a cheat. The kinds of physical laws governing H₂O molecules that are needed to yield the relevant inference for water really govern their purely micro-behaviors.

Levine's criticism seems to be that, in cases of conservative/weak emergence, using physical laws to explain macro phenomena in terms of micro-level things is not simply a matter of using a natural law. Natural laws operate in a compositional explanation at the micro level. The explanatory step from the micro level to the macro level is realized through a step of *identification* of macro phenomena with a causal role property that is manifested by (aggregations of) micro things. Thus the explanation proceeds basically with two steps:

- [1] 1. Facts about (aggregations of) micro individuals (their properties and relations),
 2. **natural laws**

Facts about the behaviors/dispositions of these (aggregations of) micro individuals.

- [2] 1. Facts about the behaviors/dispositions of (aggregations of) micro individuals
 2. A macro property X obtains = that (aggregations of) micro individuals have the described behaviors/dispositions)

A macro property X obtains.

Thus a macro property isn't really an item in a natural law. Natural laws complete the description of the micro-level facts, especially the causal orders. A macro property is then

identified with the micro-level facts. Levine's concern with Seager's suggested infusion law seems to be that if a compositional explanation of macro consciousness in terms of certain natural laws should proceed in this two-step manner, then there should be a step like the following:

[2_{macro consciousness}]

1. Facts about the behaviors/dispositions of (aggregations of) micro conscious individuals
2. macro consciousness obtains = that (aggregations of) micro conscious individuals have the described behaviors/dispositions)

macro consciousness obtains.

However, although we could grasp an identity between fluidity and behavioral dispositions of aggregations of water molecules, it's unclear what behavioral dispositions of micro consciousness aggregations can be identified with macro consciousness.

Whereas, if Seager's idea of the role of infusion laws in the explanation of macro consciousness simply requires one step that connects micro states with macro consciousness, as in

1. Facts about the (aggregations of) micro individual conscious properties and relations

2. the infusion law

Certain macro consciousness obtains.

then it “seems a cheat”, because it makes an account of emergence too easy. (As Levine seems to be pointing out) An account of conservative emergence should proceed in the two-step fashion, with both *an account of micro-level facts* and *an identification between a macro-level state and a (cluster of) micro-level state(s)*.

On my interpretation, it does seem that Seager’s explanation in terms of the infusion law would work in this one-step fashion, in which no “identity claim of the macro-property” in terms of causal role properties would be made. This is because, in the model of compositional explanation suggested by Levine, the identity is between a macro property and a causal role property *manifested in a totality of micro-level behaviors*. The micro-level things and their properties still exist, and the identity with a macro property is merely a translation of vocabulary and concept. But in Seager’s fusion view, the fusion would be a process in which the micro entities and properties fuse into one. Macro consciousness is generated from the pre-fusion collection of micro consciousness. But macro consciousness cannot be *identified* with micro consciousness collectively organizing and behaving in a certain way, because the existence of that macro

consciousness *devours* the micro consciousness and their properties. Infusion requires that a new macro-level state occurs and “substitute” the “precursor” micro-level states.

Therefore, since it’s clear that Seager’s infusionist account doesn’t follow the familiar model of compositional explanation (which can be safely adopted as a general account of conservative emergence), we have a good reason to consider Seager’s account a “cheat”. Although Seager claims that his infusionist account is of “conservative emergence”, the infusionist account seems more radical than conservative.

Why does Seager think that infusion could be “conservative emergence”? One measure Seager has taken to ensure the “conservativeness” of infusion is adopting panpsychism, since Seager believes that the generation of macro consciousness from non-conscious micro entities is radical emergence. To avoid this, the micro entities should have consciousness. However, adopting panpsychism could only address the concern that the generation of macro consciousness from non-conscious entities is radical emergence. It’s not yet established that the generation of macro consciousness through infusion of micro-consciousness is “conservative.” After all, we have already seen that a story of infusion seems “too easy” to be a story of conservative emergence. Thus, to avoid strong/ radical emergence, Seager has more work to do.

How to make sense of infusion as “conservative” emergence? Seager’s strategy, apparently, is to argue that “infusion” could offer a sort of “intelligible link” (Ibid., p. 238) between micro-level and macro-level states. For this purpose, Seager uses two examples of “infusion” in physical science to argue that the notion of infusion of micro-consciousness could be “intelligible”.

The first example is based on Paul Humphreys's (1997) interpretation of quantum entanglement. In the words of Seager,

In the standard example of entanglement, the so-called singlet state, two particles interact so as to form a new state whose mathematical representation cannot be decomposed into a product of the representations of the constituents. The system acts as a unified state insofar as measurements on one part instantaneously put constraints on measurement results on other parts of the system and there is no way to determine whether the particles are entangled by any local measurement performed on the parts. (Ibid., pp. 238-39)

Quantum entanglement is thus regarded as a case of “fusing of entities into new systems” (Ibid., p. 239), because the entangled entities behave in a unified system. A new unified system comes into being, which makes this a case of infusion. There seems to be emergence in this phenomenon. The system of entangled particles has a property that cannot be derived from the properties of the individual particles, because the mathematical representation of the “new state” of the system “cannot be decomposed into a product of the representations of the constituents.” So there is not even a way to describe the states of the particles individually from which the state of the system can be derived.

Seager regards the emergence in the case of quantum entanglement to be intelligible and “conservative” because it has been recognized in physics, predicted by “the applicable physical theories” and their fundamental properties (Ibid., p. 239.) However, even though appealing to an accepted account of a scientific phenomena to illustrate fusion might render fusion a respectable notion, it does not show that fusion is conservative emergence unless it derives the state of the whole from the lower-level

states. But quantum entanglement under Seager's preferred interpretation is a case where the system has a property which cannot be derived from lower-level states. Therefore, Seager cannot treat quantum entanglement as a case of weak/conservative emergence.¹⁷ By citing quantum entanglement, he might have shown that emergence doesn't have to stay weak/conservative.¹⁸

Moreover, Seager is relying on one interpretation of quantum entanglement, while other interpretations of this queer phenomena in quantum physics are also available.¹⁹ Therefore, the fusion account of entanglement cited by Seager might not be a sufficiently well-accepted account in science. Its force to support the plausibility of fusion seems dubious.

Seager's second example of physical infusion comes from the "no hair" conjecture of black holes. According to Seager's interpretation of this conjecture, black holes "can be exhaustively characterized by three physical properties: mass, electric charge, and angular momentum..." Seager thinks that this implies that

In a certain sense, a black hole forms something like a fundamental particle. The physical entities that form a black hole can be said to fuse into a new entity which cannot be understood as a relational structure of its precursor entities. [The precursor entities] have gone out of existence. (Ibid., p. 239)

¹⁷ I am grateful to Joseph Levine for pushing me on this point.

¹⁸ And, if emergence doesn't have to be weak/conservative, then perhaps the emergence of macro-consciousness from non-conscious entities (admittedly a case of radical emergence) is acceptable.

¹⁹ See for example Bohm & Hiley (2006) and Teller (1986), among many others.

Thus black holes are understood as entities resulting from infusion of physical entities. There is emergence here according to Seager, because the state of a black hole “cannot be understood as a relational structure of its precursor entities.” This is because the entities that compose a black hole “have gone out of existence”, which means that the black hole isn’t constituted by its precursor entities being arranged in a certain way. Also, Seager holds that because black holes are predicted by respectful physical theories, the emergence in this case should be recognized as intelligible (Ibid.).

However, this analogy with black holes is implausible for two reasons. First, like in the case of the earlier analogy with quantum entanglement, even though appealing to an accepted account of a scientific phenomena to illustrate fusion might render fusion respectable, it does not show that fusion is conservative emergence unless it derives the state of the whole from the lower-level states. However, the “no hair” conjecture of black hole is also a case where the system’s properties cannot be derived from lower-level states of the system’s components (because a black hole “swallows” all that falls into it.) Therefore, Seager cannot treat black hole as a case of weak/conservative emergence, as an analogy to support the possibility of conservative emergence.²⁰

The second reason is this. If Seager is right, then black holes may offer a model to show the possibility of infusion. But this model might not be sufficient for explaining consciousness, because consciousness presents itself as having structures, and so we need an infusion account that allows structures. A black hole fuses things by *destroying* the structures. No physical thing entering a black hole retains any structure. This is how in a

²⁰ I am grateful to Joseph Levine for pushing me on this point.

black hole, it seems that all things come into one: when all structures are eliminated in an entity, there is no way to distinguish any parts in this entity, and thus all things have been fused into one. However, if macro consciousness is generated through the fusion of micro consciousness, it doesn't seem to rely on the destroying or collapsing of all structures. As far as we know, the physical structure of a brain does not collapse or get simplified when consciousness arises. Also, it seems that the generation of macro consciousness requires a sort of *enhancement* of physical structures. For presumably consciousness requires a physical substrate of amazingly complex structure, and furthermore the complex features of macro consciousness are associated with the complexity of its physical substrate. It is also problematic to compare black holes to the fusion of micro consciousnesses. Maybe the micro-experiences should merge into one, but they should not be "collapsed" in the way that no structure is discernible in the resulting macro-experience. The end result is apparently a macro consciousness with a complex structure. But fusion in black holes doesn't seem to leave structures in the black holes. Therefore, I acknowledge that the black-hole analogy may illustrate infusion, but I contend that this analogy falls short as a model of the generation of a rich macro consciousness. Because consciousness presents structures, we need an account that allows such structures.

In sum: Seager's infusion view of panpsychism is problematic. Seager rejects the idea that micro entities without micro-consciousness can combine or fuse to produce macro-consciousness. In Seager's view, physical infusion such as the quantum mechanisms in the brain doesn't help explaining consciousness, and the fusion of non-mental ultimates into consciousness is "radical emergence" (Ibid., p. 239). Seager holds

that ultimates as fundamental physical entities have micro-consciousness, and micro-consciousness will fuse into macro-consciousness in a manner of “conservative emergence”, “under certain conditions of which we remain quite ignorant” (Ibid., p. 239). Seager tries to show the “conservativeness” of infusion with two examples from physics. The message is that infusion is intelligible because it is not beyond science, and therefore the infusion of micro-consciousness into macro-consciousness should be viewed as similarly intelligible. However, I have shown that Seager’s two examples are problematic. Therefore, his picture of conservative infusion of consciousness still seems ungrounded.

2.2 Constitutive panpsychism

In this section I examine the phenomenal bonding proposal, which is a representative of constitutive panpsychism.

2.2.1 Phenomenal bonding

Panpsychism should account for the combination of subjects. So the question is: What makes a plausible panpsychism which might answer the subject summing problem? Recall that the subject summing problem can be understood as a problem of the insufficiency of fundamental facts. The idea is that even if the ultimates are conscious subjects, the fundamental facts about their phenomenal consciousness together with other (physical) fundamental facts about them don’t entail that there is consciousness beyond the ultimates. So, if adding fundamental facts about the consciousness of ultimates is

insufficient, perhaps more fundamental facts shall be added. In particular, if facts about intrinsic properties of ultimates are not enough, perhaps facts about *relations* among ultimates could help.

Thus, Goff (2017a) argues for the possibility that “there [are] some state of affairs of certain subjects of experience *being related in some specific way* which necessitates the existence of some distinct subject of experience.” (p. 292) This is the first step toward phenomenal bonding, i.e., to consider relations among ultimates.

What kind of relations among ultimates could help derive the existence of macro-subjects? If the relation among ultimates are described merely in physical terms, such as spatiotemporal relations or physical forces, it seems that facts about micro-subjects still cannot a priori entail anything about macro-subjects (Goff 2017b, pp. 182-3). For example, suppose one knows that all ultimates in my brain are conscious, and one also knows everything about the physical features of the ultimates—e.g., mass, charge, gravitational forces among each other, etc. One can still conceive that I, composed of these ultimates, am not conscious. Therefore, mere physical relations don’t suffice.

But what if the relation among ultimates also has a phenomenal aspect? What if ways of combination can be described in phenomenal terms? Perhaps the existence of micro-subjects combined in certain phenomenal relations could explain the existence of a macro-subject in the manner of a priori entailment. This is Goff’s suggestion, after he completes his argument for “Metaphysical Isolation of Subjects” (MIS), which concludes that subjects themselves don’t suffice for making a further subject:

MIS implies that there is no state of affairs of the form <subject of experience S1 exists with phenomenal character x, and subject of experience S2 exists with phenomenal character y> which necessitates <subject of experience S3 exists with phenomenal character z>. But it does not imply that there is not some state of affairs of the form <subject of experience S1 with phenomenal character x bears relationship R to subject of experience S2 with phenomenal character y> which necessitates <subject of experience S3 exists with phenomenal character z>. (Goff 2017a, p. 292)

Thus, Goff posits a “phenomenal bonding” relation, which is the relation that “bonds subjects together to produce other subjects of experience.” (Ibid.). The idea is that the conjunction of all physical facts (intrinsic and relational) and all phenomenal facts (intrinsic and relational) about the ultimates may a priori entail facts about macro consciousness.

There might be an additional reason to posit the phenomenal bonding relation between ultimates. For one thing, besides consciousness, panpsychism (more generally, any theory of consciousness) should also explain *lack of consciousness*. A theory of consciousness should explain all facts about consciousness, not only its existence and characters but also its lacking. However, the lack of consciousness has rarely been problematized. Neither standard physicalists nor dualists find the lack of consciousness remarkable, because the concept of “physical” in standard physicalism and dualism entails the opposite of “mental”, and so physical beings are construed by default as having no mental (including phenomenal) properties. Under this framework, there is little motivation to ask why a collection of atoms have no consciousness. It’s the existence of consciousness that needs explanation, not the absence of consciousness. However, since panpsychism deviates from this default absence of consciousness by posting phenomenal

properties that exist ubiquitously at the fundamental level, the absence of consciousness becomes more remarkable for panpsychism than for standard physicalism or dualism.

It is plausible that some composites such as human beings have consciousness while other composites such as tables and chairs do not. Furthermore, human beings are unconscious during dreamless sleep. How to explain the fact that one is sometimes conscious and sometimes unconscious? The difference in conscious and unconscious states cannot be explained solely by the intrinsic phenomenal properties of ultimates, if the ultimates' intrinsic natures are fundamental and stable. Indeed, there is no reason to think that the intrinsic properties of the fundamental entities that compose my brain would change when I fall asleep. Therefore, the intrinsic phenomenal properties of ultimates are insufficient for explaining consciousness and unconsciousness. We need more, i.e., how the ultimates are related differently when the composite gains or loses consciousness. Phenomenal bonding as the special relation among ultimates could be the additional factor that explains both consciousness and unconsciousness in a composite.

2.2.2 Do we know what Phenomenal bonding is?

What do we know about the nature of phenomenal bonding? Goff admits that we know little about what phenomenal bonding is like. It seems that there is nothing further we can say about this relation. We know only the role it is supposed to play in panpsychism: if micro-subjects are in this relation, then their composite is a macro-subject. Goff recognizes that our science doesn't tell us the nature of phenomenal

properties, including phenomenal bonds. Neither do we have perceptual or introspective knowledge of them (Goff 2017a, pp. 292-93).

Goff speculates that phenomenal bonding, the relation that confers phenomenal properties to the composites, may be the “real nature” of spatiotemporal relations (2017a, p. 294; 2017b, pp. 183-86). He argues that we may lack a “transparent understanding” of the “deep nature” of spatial relations, which could be the reason why it is possible to conceive of micro-subjects physically related in space in the form of a macro creature without producing macro consciousness. Maybe micro-subjects appropriately related in space will build a creature with macro consciousness, because the “deep nature” of spatial relations confers phenomenal properties to the whole. In Goff’s term,

Lacking a transparent conception of spatial relationships, we have no grounds for denying that micro-subjects, spatially related, intelligibly produce [macro-subjects.] The micropsychist can reasonably argue... that the subject summing problem results from our lack of understanding of the deep nature of the spatial relation. If we transparently understood the nature of spatial relationships, then it would be clear to us how micro-subjects bond to make [macro-subjects].(Goff 2017b, p. 185)

Goff believes that there is “plenty of hope for theoretical progress” concerning the “real nature” of spatial relations. However it is unclear even on Russellian monism that spatial relations should have categorical bases as “real nature.”²¹ It seems that Goff’s logic is that if something is relational, then it must have an intrinsic property as its base. However, perhaps there is a confusion between relation and disposition here. Dispositional properties, such as a tennis ball’s tendency to roll, have some other properties as their

²¹ Thanks to Joseph Levine for pushing me further on this point.

categorical bases. Perhaps physical properties such as mass (defined in terms of inertia) also have their categorical bases, because they are dispositions. But many relations are not dispositions.²² Spatiotemporal relations don't seem to be dispositions. And, if a property is not a disposition, it is less obvious why it requires an intrinsic nature as its "real" or "deep" nature. Therefore, Goff doesn't have sufficient reason to posit "the deep nature of the spatial relation".

But, even if we concede that there is some (phenomenal) "deep nature" of space, still we have little clue on it. As Goff admits, he tends to be "cautiously pessimistic" toward the prospect of knowing the "deep nature" of space, because "[i]t's hard to see how we could somehow guess at the deep nature of a feature of reality we are not acquainted with, just as it's hard to see how a congenitally blind woman could somehow guess at the phenomenal nature of red." (Ibid., p. 186)

For a phenomenal bonding theorist, it is unsatisfying if nothing can be known about the phenomenal bonding relation. Thus one might attempt to find some understandable candidates of phenomenal bonding. Chalmers (2017) tentatively suggests that co-consciousness, the phenomenal unifying relation among experiences, might be such a candidate. Co-consciousness is a phenomenal relation that one could observe among the experiences in one's own consciousness. Suppose, while you are having an iced coffee, you experience coffee taste, and at the same time you experience the

²² Also, it seems that many dispositions are not really relational properties. Thanks to Joseph Levine for this point.

temperature and texture of the ice. You *notice* the experiences *simultaneously*. The consciousness of the experiences *as together*, we call it “co-consciousness.”

Chalmers’s suggestion has been developed by Miller (2017), who has offered several arguments to support co-consciousness as phenomenal bonding. (i) In the first place, co-consciousness is a “phenomenal” relation that we can find in our experience: it is “the relation in virtue of which conscious experiences have a conjoint phenomenology or a conjoint what-it-is-like-ness.” (p. 548) (ii) Also, phenomenal bonding is supposed to be a relation among subjects, and Miller argues that co-consciousness is a relation that could happen between subjects. This can be understood by considering the case of split-brain. A split-brain patient might have two conscious subjects in her skull, with two disjoint conscious perspectives and different experiences. Whereas, in a normal human person, one has a unified consciousness that joins the power of both cerebral hemispheres. Thus, perhaps in our unified consciousness we could observe co-consciousness at work in relating subjects: had I had split-brain, there would have been two separate subjects in my skull; but because I’m not split-brain (and so my brain supports the co-consciousness between the two subjects in the two hemispheres), there is one unified consciousness, and *I* am this conscious subject that results from this co-consciousness (pp. 554-55).²³ (iii) Furthermore, co-consciousness seems to be a relation such that if subjects are in such relations then a conjoined “larger” subject is bound to exist, for

²³ In Ch. 5, section 4.2.2, I will question the idea that a phenomenal relation like consciousness is held among subjects.

[C]onsider a simple set of micro-subjects [...] with their certain micro-experiences, and consider that the co-consciousness relation holds between the experiences of each of the members of the set. Once each member of the set of subjects and their experiences becomes bonded by co-consciousness, there exists an experience that corresponds to the set and the co-consciousness relations between them: there is the set and its conjoint phenomenology. If there is a conjoint phenomenology corresponding to the set, then there is a what-it's-likeness corresponding to the set, and, hence, there is a macro-subject which corresponds to the set. (p. 557)

Therefore co-consciousness seems to fit the model of an easy-to-accept notion of phenomenal relation: it's "phenomenal"; it's a "relation among subjects"; and it unites subjects into further subjects. And it seems understandable, as something that people may grasp in their own experience. Thus, without a better contender, co-consciousness may be treated as the best candidate for phenomenal bonding. But there are problems with this co-consciousness version of phenomenal bonding, to which we now proceed.

2.2.3 A Challenge to Phenomenal bonding

One worry about the co-consciousness notion of phenomenal bonding is that co-consciousness doesn't *explain* mental combination. Coleman has expressed the challenge to co-consciousness as follows:

But this is to *describe* the (desired) outcome of a certain process, without telling us at all *how* it is meant to be achieved. [...] When we drag two experiential packets out of respective microsubjects, whence does the new subject come for whom they are to be co-conscious? To say that experiential packets are related now by co-consciousness is certainly to *imply* that a new subject has come into being for whom they are phenomenally unified, but it is not to tell us how this happens, nor whether it is possible — the things we wanted to know. In the absence of further positive content, what this notion of phenomenal bonding

really amounts to is a *schema*: it specifies what an explanation of subject combination must achieve, without providing any of that explanation. It is a mere black box. (Coleman 2017, p. 257)

Thus co-consciousness seems to be the explanandum, something to be explained rather than to do the explanation of mental combination. If Coleman is right, then appealing to co-consciousness is a way to *state* the fact that mental combination happens without *explaining* how it happens. Similarly, one may argue that co-consciousness presupposes a unified subject, for to say that two subjects are co-conscious one must presuppose that there is a “larger” subject that has unified the two subjects. The combination of two distinct subjects, i.e., how the two subjects get into the co-consciousness relation, is left unexplained.

The criticism that co-consciousness is a “mere black box” might be unfair. As Miller (2017) has pointed out, phenomenal bonding understood in Goff’s fashion is indeed a “mere schema”, but co-consciousness is not a mere schema. Co-consciousness as phenomenal unity among experiences is a “positive notion,” “a relation that we are all acquainted with in our day-to-day conscious lives.” (p. 561) If co-consciousness as a relation with which we are acquainted in our consciousness can be treated as a *primitive* notion (just as the simultaneous existence of two objects in physical space can be a primitive notion), then it may be a notion that explains other phenomena but doesn’t need an explanation for itself. Therefore, Coleman’s criticism may be sound if we had no idea about phenomenal bonding or co-consciousness except for its metaphysical function of combining subjects or experiences. But, given that we seem to understand what co-

consciousness is in consciousness, we do know something about co-consciousness as a candidate of phenomenal bonding. Perhaps an account of mental combination can be given in terms of a basic co-consciousness relation, as a primitive notion that we may observe in experience, just as the spatiotemporal relation can be a primitive notion (in physics) that has observable instances.

Therefore, phenomenal bonding as co-consciousness may be better than a “mere black box.” But, our available notion of co-consciousness is definitely insufficient. If “A and B are co-conscious” is a statement of a primitive fact, there should be some understanding as to whether such a statement is true. Concerning a statement like “the greenish figure on my left and the blackish figure on my right are co-conscious”, there may be a straightforward sense of when it is true: I “look inward” and check if this statement truly describes my experience. However we have little idea about how to judge, more generally, whether a random token experience A and a random token experience B are co-conscious. Furthermore, our intuitive cases of co-consciousness have always been cases where we *already know* that *there is one subject* in which co-consciousness holds among multiple experiences (as in “the greenish figure on my left and the blackish figure on my right are co-conscious”). We don’t have a case in which we recognize co-consciousness of multiple experiences *without knowing* that these multiple experiences all belong to one subject. For this reason, co-consciousness as we know it seems to presuppose the unification of subjects. We should admit that, if co-consciousness as we know it in our own consciousness can be rightly construed as co-consciousness of multiple subjects and their experiences, then co-consciousness helps in showing that

subject combination *really happens*. It's not impossible. We experience it. However, given our available notion of co-consciousness, we cannot start with several experiences or subjects of which *we don't know* if they are unified in one consciousness, and then find co-consciousness among them and conclude that they unify—we don't know if the co-consciousness relation applies or whether it can be found in such a case. I think this is the real worry that subject combination is presupposed in co-consciousness, and therefore co-consciousness (as phenomenal bonding) is insufficient for understanding subject combination. (A similar worry will be explained in a more general context in Ch. 5, section 4.2.2.)²⁴

There is a related challenge for the co-consciousness relation as phenomenal bonding. That is: If the co-consciousness relation combines micro-subjects, where is the end to this combination? As Dainton (2011) has pointed out, the micro-subjects of our bodies are closely adjacent to micro-subjects outside. At the micro level, the relationship between two micro-subjects in my body might not be any different from that between a micro-subject in my skin and an outside micro-subject besides it. If co-consciousness combines the ultimates in my body, does it also combine ultimates beyond the skin? What determines that some ultimates combine and make further subjects, while others don't? Let's name this problem as *the Stop-Combination problem*, since it is about ceasing the combination of micro-subjects.

²⁴ I am grateful to Joseph Levine for pressing me on this point.

The Stop-Combination problem also applies to Goff's suggestion that phenomenal bonding might be a hidden nature of spatial relations, since spatial relations go beyond ultimates within a body. The problem is that,

To yield human consciousness, we presumably want phenomenal bonding to bond a limited multiplicity of microsubjects associated with the human organism, without bonding these to microsubjects elsewhere. It is not at all easy to see what sort of fundamental microphysical relation has this character. Fundamental spatiotemporal and causal relations do not seem to. (Chalmers 2017, p. 201).

If there is no account of phenomenal bonding (be it in terms of co-consciousness, spatial relations, or something else) that stops it around the skin, then we have to conclude that mental combination through phenomenal bonding may extend indefinitely, even throughout the cosmos, which gives rise to a conscious cosmos. This picture of “cosmopsychism” might be coherent²⁵, but its plausibility seems highly questionable.

2.2.4 Another Challenge: What's the use of mental dusts

Philosophers are accustomed to thinking that phenomenal consciousness is an intrinsic property. Since it is natural to think that having experience is an intrinsic property of conscious creatures, it also seems natural to think that if there is experience at the fundamental level, having experience must be an intrinsic property of fundamental micro-entities (Chalmers 2015; Goff 2017b). As Ch.1 has shown, this satisfies the demand on the intrinsic nature of the physical world to some degree.

²⁵ See Nagasawa & Wager (2017), Shani (2015), and Goff (2017b) on cosmopsychism.

Unfortunately, we have seen that there is a problem of combining the micro-entities that have intrinsic phenomenal properties. We have also seen that the way to overcome this problem is to add basic ontological items other than intrinsic phenomenal properties, such as special micro-level relations (e.g., phenomenal bonding) or special natural laws (e.g., panpsychist infusion laws). In fact, most of the work of handling the combination problem should be done with the added principles. Intrinsic phenomenal properties of ultimates don't help much.²⁶ Therefore, one might think that no work is done by the supposed intrinsic phenomenal properties of ultimates.

This calls into question the need of “small consciousness” of ultimates. One may wonder: what if there is, at the fundamental level, no intrinsic phenomenal property?

Abandoning intrinsic phenomenal properties at the fundamental level may violate Russellian monism (RM), a common theme in pan(proto)psychism which takes the fundamental intrinsic nature of the physical world to ground consciousness (Ch. 1, section 2). But it is possible to keep RM without fundamental intrinsic phenomenal properties, if it is possible to make sense of fundamental intrinsic properties that are non-phenomenal.

²⁶ This is echoed by Coleman's (2017) criticism of Goff's phenomenal bonding. Coleman observes that

[T]he glaring truth about this 'explanation' [of mental combination, in terms of phenomenal bonding] is that the fact of the phenomenally bonded ingredients' being *subjects* plays no role whatever. All the work is done by the phenomenal bonding relation: it is a relation such that, by definition, it generates a macrosubject. (pp. 257-58)

Furthermore, there is a worry about RM. The worry about identifying the intrinsic nature of the physical with phenomenal properties emerges upon reflecting on the relationship between intrinsic and extrinsic properties, or between a disposition and its categorical base. It is assumed under RM that a dispositional property must have an intrinsic/categorical base. But is there any further connection between dispositional and categorical properties? One such connection could be that a dispositional property can be explained partly as the result of its categorical base. For example, consider glass's disposition to break. This disposition has a categorical base in glass, i.e., its physical composition. This physical composition helps explain why the glass is disposed to break. We may expect that this should hold true for all purported categorical base of dispositions: the categorical base should help explain the disposition. But if categorical basis should help explain the relevant disposition, then if RM is true, then phenomenal properties should be able to figure into explanations of physically described properties. The question is, does it even make sense that phenomenal properties figure in an explanation of physically described properties, such as mass and charge? Positing consciousness as an intrinsic nature of the mathematical structures described by physical sciences seems arbitrary. Phenomenal properties don't explain the physical properties of

dispositions as such. The prospect of such an explanation is unclear.²⁷ So there is a reason to doubt that the intrinsic nature of matter is phenomenal.²⁸

This worry isn't conclusive. Maybe there is such an intelligible explanation of mass and charge as dispositions in terms of phenomenal properties as their intrinsic base, but we don't know. Maybe there is no need for an explanation that we can comprehend, but there is a *metaphysical explanation* which allows mass and charge to be grounded on phenomenal properties. Or maybe no explanation is needed for microphysical properties as *basic* dispositional properties. Maybe it's a primitive fact that such-and-such *basic* intrinsic properties ground such-and-such *basic* dispositions. There are more issues about intrinsic properties and RM, which I cannot discuss here.²⁹ But this worry (and others) should suffice to let one pause and think before embracing RM.

2.3 Panprotopsychism

In this section I will briefly survey panprotopsychism, which still traces the source of consciousness to special micro-level consciousness-generating properties, but

²⁷ Levine (2017) has expressed the same worry:

[I]n standard, non-fundamental cases, we look to the features of the categorical bases to explain the dispositions they underlie. So, if being H₂O is the categorical basis of water's disposition to freeze at 0 °C, then we expect we can explain why water freezes at that temperature by appeal to this underlying molecular structure (along with other facts of course). But is being reddish going to explain why one thing attracts another as opposed to repelling it?

²⁸ David Lewis has also pointed out that even if the fundamental intrinsic properties are qualia, we still cannot have knowledge of what they are (2009, pp. 217-8).

²⁹ For doubts on RM, see Kind (2015), in Alter & Nagasawa (2015); Howell (2015); Pautz (MS).

doesn't accept consciousness as an intrinsic property of fundamental micro entities. I will briefly explain panprotopsyism in general (section 3.1) and then focus on Coleman's panqualityism (section 3.2).

2.3.1 Panprotopsyism

Recall that according to panprotopsyism, the fundamental micro-entities do not have phenomenal properties, but they have protophenomenal properties that enable them to compose entities at the macro-level that have consciousness/phenomenal properties (Ch. 1, section 2.3).

Panprotopsyism could also be associated with what is called "neutral monism". According to neutral monism, the fundamental intrinsic properties of ultimates are "neutral" in the sense that they are neither physical nor mental. Both physical and mental properties are derived from the neutral properties when conditions allow. Neutral monism can be regarded as a version of panprotopsyism, since neutral properties are not mental (more specifically, not phenomenal) but are protophenomenal in the sense that they are the intrinsic properties of the ultimates that ground consciousness and other properties that manifest in macro creatures.

In the literature, ideas that resemble panprotopsyism have been discussed by many philosophers (e.g., Holman (2008), Pereboom (2011), Russell (1921), Stoljar (2001), Stubenberg (2014, 2015), etc.) In section 3.2, I will consider one theory of panprotopsyism in detail, which is the proposal from Sam Coleman (2012, 2014, 2017).

2.3.2 Of Coleman's panqualityism.

2.3.2.1 Introducing panqualityism

Coleman's panqualityism (2012, 2014, 2017) is a kind of panprotopsychism, because according to Coleman, fundamental entities are not conscious subjects themselves, but these entities take part in the generation of conscious experience owing to their intrinsic non-phenomenal but qualitative properties. (The meaning of "qualitative" will be immediately explained.)

Coleman's central move is to deny the existence of micro-subjects in ultimates. Coleman holds that the subjective-summing problem reveals an intuition that "subjectivity, far from being a diffused sort of 'stuff' or generalized property, inevitably comes in the form of certain quanta", and these quanta are "unsuited to the constitution of any further entity." (2017, p. 259) Since Coleman believes in this "metaphysical solidity of subjects" (ibid.), he believes that panpsychism cannot succeed in deriving macro-subjects from micro-subjects. Thus Coleman rejects panpsychism's commitment to ultimates as micro-subjects. As a result, in Coleman's account there is no need to address the subject-summing problem, of how micro-subjects combine to make macro-subjects.

Coleman has no micro-experience or micro-subjects in his account. Thus Coleman's position avoids the question concerning the combination of micro-subjects. But Coleman still has the qualities and subjectivity of macro-experience to account for. Let's start with Coleman's view on qualities. Even though Coleman denies that ultimates are conscious subjects, Coleman wants the ultimates to preserve some aspects of

phenomenal properties. Since there are two basic aspects in phenomenal consciousness, i.e. the conscious subject and the qualities that the subject is aware of, Coleman hypothesizes that ultimates have “qualities” (hence the name “panqualityism” for his position). According to Coleman, “qualities” are intrinsic properties of the ultimates that are like the phenomenal properties that one could encounter in one’s phenomenal consciousness. But these “qualities” are not phenomenal properties, for they can exist in ultimates without being experienced by the ultimates, i.e. without these ultimates being the subjects that are aware of them. The “qualities” can merely instantiate in ultimates without being experienced. And according to Coleman, they exist without being experienced most of the time in the universe.

Admittedly, it is not easy to conceive of a non-conscious, non-phenomenal qualitative property. Whenever we conceive of qualitative properties, we always seem to be thinking about what they are like as we experience them. Thus, qualitative properties always seem to present to us as phenomenal. To help making sense of this notion of non-phenomenal “qualities”, Coleman has given the following example of the quality of redness:

We, in the everyday mode, think of objects as straightforwardly coloured. For example, we think of a red London bus as being straightforwardly red, much as if the redness were painted over the surface of the bus in a quite objective way. [...] As part of this conception, we take it, in an easy and uncomplicated way, that the bus is still red when we are not looking at it, and even when no-one is looking at it, as when it is waiting in its garage overnight. What is the form of this conception of the bus as still red even when unobserved? If pressed on this point, we are likely to say something like that the bus has just the quality it has when we are looking at it, it is still *being red*, ‘radiating’ redness in a sense, even when no-one is there to enjoy it. [...] We have, on the commonsense way of thinking, no

difficulty with the notion of qualities which exist with no-one there to experience them... (2012, pp. 154-5.)

Thus Coleman clarifies “qualities” by highlighting the “commonsense way of thinking” about the qualities of perceptual experience, which is a sort of naive realism about perceptual qualities. We tend to attribute the qualities in perceptual experience to the perceived objects, but a little reflection might incur the idea that the qualities are subject-dependent in the sense that they only exist with the awareness of a subject. If a tree falls in a forest and no one listens, then there are waves in the air, but would there be any sound quality? Who would have the experience of that sound quality? (The answer seems to be: No one.) Similarly, if a London bus is seen by nobody, then there are lights reflected from its surface with a certain range of wavelength, but would there be any color qualities? Who would be experiencing the color qualities? (The answer seems to be: No one.) Contrary to this reflective notion of qualities of experience, Coleman is arguing that there is a less reflective notion of qualities as a property that can merely be instantiated without being experienced, and that qualities of this sort are instantiated at the fundamental level. Perhaps it’s possible to conceive of such qualities at the fundamental level in Coleman’s account. But such qualities are not phenomenal qualities, because according to Coleman there is no micro-subject of consciousness.

Concerning the conscious subject, Coleman doesn’t deny its necessity for the existence of consciousness. However, to explain macro-subjects and macro-experience/consciousness, Coleman cannot appeal to the features of micro-subjects, since he has denied that ultimates are micro-subjects. Coleman thus appeals to some principles of

arrangement of the ultimates to explain macro-subjects of consciousness, and the principle that he uses is at a level of composition that is quite high (at a creature level.) Coleman utilizes “a sort of higher-order thought theory of consciousness” (HOT) to explain conscious subjectivity (2012, p. 159, fn. 39). The proposal is that there is a conscious subject and conscious experience in a system iff a part of this system bearing the “qualities” is in some right causal relation with another part bearing such “qualities”, which constitutes what Coleman calls “phenomenal representation” of one with the other:

I hypothesize that the *phenomenal representation* by one phenomenally-qualified item of the phenomenal quality of another such item occurs through the first item taking on the phenomenal quality of the second. The capacity of the first item to take on the quality of the second item will be enabled, the idea goes, by the two items being suitably causally related within the structure of the subject. Such causal relating, we may speculate, is a good deal of what the brain is for. (Coleman 2012, p. 159. Also see Coleman 2014, pp. 41-2.)

Let’s consider the following schematic example to see more concretely what Coleman means for one item to “take on” the phenomenal quality of another. Suppose that I look at a red London bus and have a visual experience of seeing red. According to panqualityism, what happens is something like this. Region A of my brain (e.g., a part of my prefrontal cortex) instantiates a complex of “qualities” (call it Q_A). Region B of my brain (e.g., my visual cortex) instantiates another complex of “qualities” (call it Q_B). Neither Q_A nor Q_B is phenomenal (since neither is in phenomenal consciousness, assuming that A and B are not themselves conscious)—until A and B are in the right causal relation. When I have visual consciousness of the redness, what happens is that A and B are “suitably causally related” (following HOT, it should be that the relation constitutes a representation of Q_B

with a higher-order thought about Q_B) so that A “takes on” Q_B (Coleman 2012, p. 159).³⁰

When that happens, according to Coleman, there is a higher-order representation of Q_B , which constitutes the phenomenal consciousness with qualitative properties like Q_B . Thus I experience reddishness.

2.3.2.2 Against Coleman on Subjectivity of Consciousness

There are at least two issues with Coleman’s panqualityism.

The first challenge of panqualityism is to make sense of its notion of “qualities”, which are the fundamental protophenomenal properties in Coleman’s account. Coleman claims that we could, “in an easy and uncomplicated way”, understand the notion of such “unexperienced instances of phenomenal quality” (2012, pp. 154-55). I am not sure

³⁰ Coleman’s (2012) account of this example goes like the following:

[W]hen confronted with a red thing in suitable light, the signal propagated to the phenomenal screen [B] causes it to turn phenomenal red. Subjectival awareness of this perceptual representation is then accomplished by the aforementioned mechanism of phenomenal representation: a second phenomenal screen [A], this time corresponding to the central perceptual/ experiential domain of the subject, receives a signal from the visual screen [B], and represents what it finds there, i.e. it turns, in some portion, phenomenally red also. This central screen, however, also receives inputs from the other externally-facing senses and their respective phenomenal representations, and from ‘screens’ responsive to the internal (proprioceptive, emotional, etc.) states of the system too. As these are registered in the central subject-screen, we build up a phenomenal representation of the complex state of the whole system, the organism. (Coleman 2012, p. 160)

This account suggests two points in Coleman’s (2012) proposal. First, for A to “take on” Q_B to represent Q_B is for A to be caused so that Q_A becomes similar to Q_B (“... it [A] turns, in some portion, phenomenally red also.”) Second, Coleman’s account of the mental architecture of consciousness also indicates a global workspace model (GW), since the “central screen” works like a center space that receives information from other more specialized mental faculties. Therefore, it can be questioned whether Coleman’s panqualityism follows HOT or GW theory in its account of conscious subjectivity.

whether it is that “easy and uncomplicated” to understand them. These “qualities” are not merely physical properties, such as wavelengths of lights. Qualitatively, “qualities” are supposed to be the same as phenomenal qualities in consciousness. They are *as what they would be like* if they were experienced. But “qualities” can exist outside of phenomenal consciousness, and so they are not phenomenal properties either.

If we can make sense of Coleman’s “qualities” as phenomenal properties minus conscious subjectivity, as the merely instantiated but unexperienced qualities, then Coleman’s “qualities” may offer a somewhat graspable idea of protophenomenal properties as neither physical nor phenomenal. But the success of “qualities” in this sense is contingent upon the plausibility of this “easy and uncomplicated” way to think of merely instantiated but phenomenal-like qualities. For me, it doesn’t make intuitive sense to talk about (for example) any visual quality of color instantiating in the world if no one sees. Such qualities do not merely exist in the way that they *are*; intuitively the qualities are there *for someone*, some conscious subject.³¹

The second and the main challenge to panqualityism (or to panprotopsychism in general) is concerned with the adequacy of its higher-order account of subjectivity without micro-subjects. This challenge can be described as an epistemic gap in panqualityism. As Blamauer (2013) has pointed out, the HOT account of conscious

³¹ This intuitive appeal to the need of subject for phenomenal qualities can be interpreted with further reflections. For example, Levine (2010) argues that “phenomenal colors are essentially *ways of appearing*,” and therefore “one can’t remove the experiential component and still have the phenomenal color...” (pp. 218-20.) If qualities are “ways of appearing”, since “ways of appearing” are ways of appearing *to someone* (subject), this notion of qualities as “ways of appearing” means that qualities cannot exist without a subject.

subject is inadequate. Given the higher-order representational relation, it is conceivable that there is no phenomenal consciousness. In other words, assuming Coleman's panqualityism, the existence of macro-subject and macro-experience cannot be a priori entailed from the fundamental facts assumed by Coleman. Therefore, panqualityism faces an epistemic gap between its fundamental facts and facts about consciousness. Chalmers (2015, 2017) has expressed this idea with the conceivability of "qualitative zombies", which are "beings that are qualitatively (and microphysically) identical to us [but] without consciousness" (2015, p. 273). Chalmers justifies the conceivability of "qualitative zombies" in terms of the conceivability of lack of awareness ("for any set of instantiated qualities and physical properties, it is conceivable that all those qualities and properties are instantiated without any awareness of the qualities" (ibid.)). This is equivalent to arguing for the conceivability of the lack of a conscious subject, since a conscious subject exists only insofar as there is conscious awareness of something (qualities).

How does Coleman respond to this challenge? Coleman (2017) construes Chalmers's claim as that his HOT account in panqualityism fails to "reductively capture a certain phenomenology" (p. 270). He quotes the following passage from Chalmers (2017):

[a]wareness involves *phenomenology*, and there are good reasons to think that no mere functional state can constitute phenomenology... one can conceive of any such functional state in the absence of phenomenology, and in particular in the absence of awareness. (Coleman's emphasis; original in Chalmers 2017, p. 203; quoted from Coleman 2017, p. 270.)

Then Coleman interprets the objection as follows,

The anti-HOT panqualityism zombie argument is thus to operate much like the anti-physicalist zombie argument—it hangs on a failure to reductively capture a certain phenomenology. In our case the missing target is narrower than *phenomenology in general*—the target of standard zombies. Chalmers apparently holds that, in addition to being aware of sensory qualities, we’re *aware of our awareness* of sensory qualities. This further object of awareness—awareness itself—comes with its own patch of phenomenology: a qualitative feel. It’s this feel which our HOT-panqualitative duplicates are alleged conceivably to lack. (Ibid.)

However, this interpretation seems to be a serious misreading of the issue. The complaint of panqualityist zombie is precisely that a HOT account of subjectivity fails to confer *phenomenology in general*, not just a “narrower” part of phenomenology. If Coleman is right, then it seems that panqualityist zombies lack some phenomenal experiences (“awareness of awareness itself”) but still enjoy some other phenomenal experiences (“being aware of sensory qualities”). But the real challenge of zombies is not that the panqualityist zombies lack just some experiences. The challenge is that such panqualityist zombies have no awareness—no phenomenal experience—at all, not even awareness of sensory qualities.

Thus, as a result of Coleman’s (mis)understanding of the challenge as the conceivability of the lack of a special phenomenology/quality of subjectivity or awareness of awareness, Coleman’s response consists of rejecting the phenomenology of awareness. He treats the so-called quality of awareness as a special phenomenology and denies it. Because the special phenomenology of subjectivity is not the issue, Coleman’s responses head to a wrong track.

For example, he uses the analogy of the absence of a TV camera in a scene shot to illustrate the absence of subjective awareness itself in the experience:

It is via awareness that we encounter sensory qualities and the appearances of things, but why should the faculty that presents sensory qualities to us itself make some appearance among our sensory qualities? That would be akin to the camera appearing in the periphery of every shot of a television show. (Ibid., p. 270).

Coleman also describes an observation of the “transparency” of awareness in the case of seeing redness:

In being aware of *red*, I just don’t know what my alleged awareness of my awareness of red is meant to feel like; I find only the redness. When you ask me to attend to the relational property of *my being aware of the redness*, still all I find is the redness—I don’t seem to enter the picture (in respect of that redness). (Ibid., p. 271)

Even if these examples from Coleman and the argument they support against the special phenomenal qualities of awareness are plausible, I suspect that they are irrelevant to the issue. Subjective awareness is essential for consciousness, but whether there is phenomenology peculiar to subjective awareness is quite obscure. Coleman may be right about this point.³² However, regardless of whether subjective awareness has its own phenomenology, the essential role of subjective awareness is to make consciousness possible at all. The worry of HOT-panqualityist zombie is precisely that no subjective awareness is derived from the fundamental elements of panqualityism, not that

³² A detailed discussion of this issue would bring in many complications about the experience of self-consciousness. See Zahavi 2005; Kriegel 2003, 2004, 2009; Siewert 2013; Schear 2009.

panqualityism fails to account for any specific phenomenal content of subjectivity. To the extent that awareness by a subject is necessary for consciousness, whether subjective awareness exists is a different issue from whether subjective awareness has its distinctive phenomenology. The worry of panqualityist zombie is about the absence of subjective awareness, but Coleman's response is only about the absence of its distinctive phenomenology. For this reason, Coleman's response to the worry of panqualityist zombie has missed the point.

2.4 Chapter Conclusion

In this chapter, I have criticized several major approaches to pan(proto)psychism in the current literature. These approaches include Seager's infusionist panpsychism, Goff's phenomenal bonding proposal, and Coleman's panqualityism. They all face serious challenges in terms of whether their metaphysical picture is sufficiently clear or is workable.

In the next chapter, I will further examine pan(proto)psychism by considering a currently available theory of consciousness in cognitive science (i.e., IIT) which has attracted panpsychists. I will argue that the picture in IIT doesn't fit well with the metaphysics of panpsychism. Therefore, citing IIT adds no justification to panpsychism as a metaphysical proposal.

CHAPTER 3

FROM PANPSYCHISM TO HYLOMORPHISM?

In the last chapter (Ch. 2), I have reviewed and criticized several prominent accounts of pan(proto)psychism that are proposed to address the combination problem. Therefore there is not yet a metaphysics of consciousness that answers the combination problem with success.

Panpsychists may seek help from outside of philosophy. That is, if there is a scientific theory of consciousness that seems to accomplish an account of mental combination, panpsychists may seek to combine their metaphysics with that scientific theory. And, as it happens, there is a scientific theory of consciousness which seems to have an account of mental combination. This is the Integrated Information Theory of Consciousness (IIT). Thus panpsychists may want to get support from IIT (e.g., in Mørch 2018). The idea is that, if IIT is a decent scientific theory that has a promising account of mental combination, then grafting panpsychism onto IIT would outsource much of the trouble of explaining mental combination. Furthermore, having IIT as a representative of panpsychism in cognitive science also shows that panpsychism can be compatible with empirical science.

I doubt if this attempt to combine panpsychism with IIT as a respectable scientific theory of consciousness will succeed. In this chapter I will argue against this attempt. I will also argue that, actually, another metaphysics, i.e., a certain account of hylomorphism, may have a better account of the categories to accommodate IIT.

Therefore, panpsychists cannot support themselves by citing IIT as their scientific representative. Panpsychism and IIT are not really in good company.

This chapter is planned as follows. In section 1 I explain several desiderata of an account of mental combination. In section 2, I briefly introduce IIT. Section 3 argues that IIT doesn't fit with panpsychism. Section 4 argues that IIT isn't exactly standard functionalism. Section 5 argues that IIT fits with a certain version of hylomorphism. Section 6 concludes.³³

3.1 Desiderata of Mental Combination

Before proceeding to another possible solution to the combination problem, it would be helpful if we have a list of the desiderata of mental combination. I will propose a list based on three observations on what mental combination is supposed to achieve. These desiderata, especially (ii) and (iii), are included to limit the weirdness of a panpsychist theory regarding what in the world can have consciousness. Admittedly, from a commonsense standpoint, it is weird enough to posit consciousness at the micro level. But if commonsense intuitions are still to be honored or addressed in metaphysics, panpsychists should consider these intuitions when they work on their hypotheses. Other things being equal, among the different versions of panpsychism, the one that accords the most with commonsense intuitions should be preferred.

(i) Generate macro-subjects. First and foremost, any account of consciousness should account for macro consciousness. For theories such as panpsychism in which

³³ Sections 2, 3, and 4 of this chapter draw heavily from Liu (2020a).

macro consciousness is produced by micro-level consciousness, the combination of micro consciousness should generate consciousness in creatures like us.

The combination problem is structurally similar to the question of composition in mereology. A central question in mereology is concerned with the conditions under which individual “objecthood” of a composite is obtained. For instance, can a collection of particles compose a dog as an individual, and if it does, how? Likewise, the combination problem in panpsychism is also about how an ordinary “individual” is possible. It is concerned with the conditions under which individual “subjecthood” of a macro-level creature is obtained from a collection of ultimates as micro-subjects.

(ii) Stop Combination. Also, recall that the requirement on mental combination is not simply to combine micro consciousness into further consciousness in macro-subjects (Ch. 2, section 2.3). It is also desirable to stop mental combination at a certain point.

The need to stop mental combination emerges when one considers the phenomenal bonding solution. The phenomenal bonding theory of panpsychism has difficulty in satisfying this desideratum, because the candidate relations for phenomenal bonding seem insufficiently constrained. For example, if phenomenal bondings correlate with spatial relations, then there is little constraint on what ultimates can compose a conscious subject. My brain could be combined with ultimates around my body to make a conscious subject beyond myself. Indeed, the ultimates in the entire universe may combine into a cosmic consciousness. Unless a panpsychist is willing to accept such unrestricted combination of consciousness, she should set a boundary to stop combination.

(iii) **Merge subjects.** There is one more desirable feature for panpsychists to demand from mental combination, which is plausible but not yet mentioned. That is, it is desirable to merge the micro-subjects once they combine.

To appreciate this point, let me explain the advantage of having fewer subjects.

First, distinguish “Panpsychism_{ATOM}” from “Panpsychism_{FUSE}”:

Panpsychism_{ATOM}: after micro-subjects combine and make further subjects, the micro-subjects still exist.

Panpsychism_{FUSE}: after micro-subjects combine and make further subjects, the micro-subjects cease to exist.

In panpsychism, the combination of micro-subjects such as s_1 , s_2 , etc. are supposed to make a conscious whole. Thus, according to Panpsychism_{ATOM}, s_1 , s_2 are still conscious subjects after the combination. Panpsychism_{FUSE} denies this.

Panpsychism_{FUSE} is arguably more plausible than *Panpsychism_{ATOM}*.

Panpsychism_{ATOM} allows many more subjects compared to *Panpsychism_{FUSE}*, since it allows the conscious ultimates in a conscious creature to remain conscious subjects. That allows too many subjects in a conscious creature. (Try imagine that many components of your nervous system have consciousness of their own, different from your consciousness.) Therefore, if one recognizes the intuitive need to minimize the population

of conscious subjects, a panpsychist should avoid *Panpsychism_{ATOM}* and take *Panpsychism_{FUSE}* as a constraint on her account of panpsychism.

Thus, an ideal account of mental combination should satisfy the three desiderata just explained (generate macro consciousness/macro-subject, stop combination, and merge micro-subjects). In the next section (section 2), I will introduce IIT. I will explain how IIT could meet these three requirements in section 3.1.

3.2 IIT

Integrated Information Theory (IIT) is developed by Giulio Tononi and others over the last two decades, aiming at explaining consciousness as ‘integrated information’, a property which is mathematically calculable based on a system’s internal causal relations (Tononi 2004, 2008; Oizumi et al. 2014). It has received considerable attention among philosophers who work on consciousness, especially those attracted to panpsychism (Chalmers 2017; Goff 2017; Rosenberg 2015).

IIT is advertised as respecting phenomenology, “[starting] from consciousness itself” (Tononi & Koch 2015, p. 5). Following this, my introduction will start from the phenomenology that IIT is intended to capture. I will then sketch the machinery that follows from the phenomenology and the proposed identity between consciousness and IIT machinery.³⁴

³⁴ This introduction mostly follows IIT 3.0, based on Oizumi et al. 2014, Tononi 2015, and Tononi et al. 2016. Both Oizumi et al. 2014 and Tononi 2015 contain details of the mathematical modeling. For a simplified guide on IIT with mathematical formulas, see Aaronson 2014. My presentation is a simplified version without mathematical calculation.

3.2.1 Phenomenological motivations

IIT starts with a list of intuitive ‘axioms’ about the features of consciousness:

- *Intrinsic existence*: Consciousness exists; ‘it exists from its own *intrinsic perspective*, independent of external observers.’ (ibid., p. 7)
- *Composition*: A conscious state, like a phenomenal field, covers multiple phenomenal aspects. For example, a conscious visual field contains many patches of color in experience.
- *Information*: Consciousness has information; the experience of one state differs from that of another state. For example, seeing orange is experienced as different from seeing white.
- *Integration*: Consciousness is a unity, a field that seems irreducible to its parts. For example, my visual scene of seeing many patches of color at once is a unified visual experience; my visual experience is also integrated with other sensory experiences; etc. All my experiences at one moment are in one unified field.
- *Exclusion*: A system has only one consciousness, which is “definite in its content and spatio-temporal grain.” More precisely, for any human experience and its content, the conscious subject isn’t simultaneously “having an experience with less content... or with more content...” Also, “[t]he duration of the instant of consciousness is also definite, ranging from a few tens of milliseconds to a few hundred milliseconds, rather than lasting a few microseconds or a few minutes...” (Tononi et al. 2016, p.

452) The intuitive idea is that at any moment, a conscious subject has only one consciousness, rather than a superposition of multiple conscious perspectives.

IIT then seeks to model the features of a physical system that realizes these phenomenological features. The modeling follows the axioms of *Information*, *Integration*, and *Exclusion*, which suggest that the feature of a system that realizes these phenomenological axioms is an irreducible system-level property that provides information. IIT theorists identify this irreducible property as “integrated information” (Φ) and propose a mathematical account to calculate its value.

3.2.2 A Briefing on the Machinery of IIT

An account of IIT should explain (a) what could bear integrated information, and (b) how integrated information is derived.

(a). The entities that could bear integrated information are ‘mechanisms’ or ‘systems’. A *mechanism* is a collection of entities that could have an informational state in virtue of its causal relations with other parts of the system. Examples of mechanisms include a neuron, a logic gate, or a collection of such things. Mechanisms can be combined, which enables the state of one mechanism to affect the state of another mechanism. A *system* is a complete combination of such mechanisms. An example could be the nervous system of an organism or a network of logic gates.

(b). The ultimate purpose of IIT is to understand consciousness as integrated information and describe integrated information as an *irreducible* property that is a

function of the information conveyed by the elements within a system.³⁵ IIT has three steps to compute the integrated information that corresponds to a system's consciousness.

Step 1 in IIT is to identify the relevant "information" in a system. This is executed by Postulates of *Information*. Whereas integrated information is a property of a system, the "information" in the system are properties of its composing mechanisms. A mechanism's "information" is defined as its "cause-effect information," which is its causal power with regard to other mechanisms in the system. Its evaluation provides a complete profile of the mechanism's current causal relations with the possible past/future states of other mechanisms in the system.

For example,³⁶ in a system of three interrelated elements A, B, and C, the state of A at time t causally constrains what states B and C could be at an earlier time ($t-1$) and a later time ($t+1$). This constraint on the past/future states of B and C constitutes the cause-effect information of A at time t . Mutatis Mutandis for the cause-effect information of B and C at time t . The totality of the cause-effect information of the mechanisms in a system provides the "information" for the derivation of integrated information.

³⁵ Since according to the axiom of *Composition* a subject's consciousness at a time could have multiple aspects of experiences, IIT has two levels to cover both individual aspects of conscious experience and the entire conscious state of a subject. IIT first accounts for the 'mechanism' level by characterizing each experiential aspect as *the irreducible* informational property of a mechanism. Then IIT proceeds to the system level by considering collections of mechanisms as a whole and characterizing *the irreducible* informational property of the system. For both 'mechanism' and 'system' level, there are similar but not identical postulates of Information, Integration, and Exclusion. To simplify the introduction, I ignore the subtleties in the two-level account.

³⁶ The toy example in this section is inspired by Tononi et al. 2016, Supplementary Information 1.

Step 2 is to derive an irreducible property from the “information” previously defined, which goes beyond a mere juxtaposition of the causal powers (cause-effect information) of the individual mechanisms. This irreducible property is obtained through Postulates of *Integration*. The natural idea is to find an irreducible causal power of this system (considered as a composite of mechanisms), given the “information” previously defined as causal powers of the mechanisms. According to IIT, to determine whether a composite has irreducible causal power, we should check if any partition of this composite changes its causal power. If there is a partition that preserves all causal power in this composite, then there is no need to consider this composite as a whole—it doesn’t contribute anything more than the *mere* aggregation specified by the partition. The composite therefore has no irreducible causal power. If no partition of the system preserves all cause power within it, then this system has an irreducible property, i.e., an irreducible causal power. This irreducible property is integrated information (Φ).

For example, in the three-element system ABC, if the partition of ABC into AB and C doesn’t reduce any causal relations within ABC (e.g., when AB and C do not causally affect each other), then $\Phi_{ABC}=0$, and we can consider AB and C as two individuals without treating ABC as one. It turns out that a system has integrated information “only if every one of its parts has *both* causes and effects within the rest of the system,” for “otherwise from the intrinsic perspective of the system that part would not exist.” (Tononi et al. 2016, Supplementary Information 1)³⁷

³⁷ For examples of partitions of systems that indicate lack of integrated information, see Figure 7(A, C, D) in Oizumi et al. 2014.

A system's Φ is evaluated by considering the cause-effect information in the composite as a whole (i.e., all cause-effect information in it) *compared to* the cause-effect information in *all possible partitions* of the composite. In our example of system ABC, Φ_{ABC} is determined by comparing the totality of cause-effect information in ABC with those of all its partitions (AB/C, A/BC, AC/B, etc.). If a composite has irreducible information, then its Φ is non-zero. The composite would also have a “quality,” known as its “conceptual structure,” determined by the sets of cause-effect information from which its Φ is derived.³⁸

Step 3 in IIT derives the *maximal* integrated information (Φ^{MAX}). Now that there is a composite that has Φ as an irreducible property, maybe this composite is itself a part of a more complex composite that has a different Φ . For example, perhaps ABC itself is a part of system ABCD. If so, we may find multiple composites with different Φ 's in a system. Which Φ or Φ 's shall be identified with consciousness?

This is answered by Postulates of *Exclusion*.³⁹ *Exclusion* requires that consciousness be identified with “a definite set of cause-effect repertoires over a definite set of elements, neither less nor more, at a definite spatio-temporal grain, neither finer nor coarser.” (Tononi et al. 2016, p. 452) Accordingly, only one Φ shall be identified with

³⁸ The conceptual structure is visually representable as a ‘constellation’ of ‘stars’ in a multidimensional space that characterizes possible past and future states of this composite under the current state. The constellation has a geometric shape in the multidimensional space. (Oizumi et al., 2014; Tononi, 2015.)

³⁹ More precisely, this specific question regarding selection of Φ to identify consciousness is executed through the postulate of *Exclusion* at the system level. The postulate of *Exclusion* at the mechanism level gets into the complications about how ‘information’ is to be evaluated (under ‘purviews’ (Oizumi et al., 2014)) for individual mechanisms. Here I ignore this because it matters little for my purpose.

consciousness. Moreover, *Exclusion* suggests that it's the maximal Φ that makes consciousness in a system. For example, even if ABC is a part of ABCD and both have positive Φ , if $\Phi_{ABC} > \Phi_{ABCD}$, then Φ_{ABC} has priority in determining the consciousness of ABCD. But, if the system ABCD is organized in such a way that both ABC and ABCD have positive Φ and that $\Phi_{ABCD} > \Phi_{ABC}$, then in this case Φ_{ABCD} has priority in determining the consciousness of ABCD.

The “quantity” of this “largest” irreducible information in a system is its Φ^{MAX} , which is the contribution to the causal power within the system made by the entire system *over* the sum of its parts. The “quality” of maximal integrated information is known as the “maximally irreducible conceptual structure” (MICS). Just as other ‘conceptual structures,’ MICS is a function of all causal profiles of the mechanisms within a composite.⁴⁰ What’s special about MICS is that the Φ corresponding to MICS is Φ^{MAX} , which excludes all Φ ’s of other conceptual structures within the system.

With this characterization of integrated information, IIT is designed to address questions about the existence and specific qualities of consciousness. The question of whether consciousness exists in a system becomes “how much consciousness is there in this system?”. For IIT, a system is conscious if its Φ^{MAX} is non-zero. A high value of Φ^{MAX} indicates a high level of consciousness. The conscious quality of a system in a state is identified with MICS, the ‘conceptual structure’ that results in Φ^{MAX} .

⁴⁰ MICS is also mathematically representable as a constellation of stars in a multidimensional space.

Some have criticized IIT for its counterintuitive consequences. For example, Aaronson (2014) argues that IIT entails that an ordinary DVD player is conscious when it starts working because an operating DVD player has a high value of integrated information. Also, calculating integrated information for a complex organism is virtually impossible, as the amount of calculation is astronomical. Also, there are challenges from philosophers, who think that IIT fails to capture the subjective aspect of consciousness (Peressini, 2013), or that IIT fails to capture consciousness at all (Cerullo, 2015; Mindt, 2017). But IIT also has some advantages. Most significantly, IIT can explain the distribution of consciousness found in many ordinary observations. For example, it can explain the apparent presence of consciousness in many cerebral activities by appealing to the fact that these activities facilitate a high value of Φ . In contrast, consciousness doesn't seem to exist in cases of coma, general anesthesia, or deep sleep, because the (remaining) neural activities in these cases do not produce much Φ . Also, since IIT measures consciousness by measuring the internal causal connections of a system, it offers a clue about the existence (or lack) of consciousness in cases where ordinary observations tend to be silent, such as infant, animal, and machine consciousness (Tononi & Koch 2015). It also explains why consciousness exists at the level of brain regions, rather than at other scales such as within individual neurons (Ibid., p. 10).

3.3 Does IIT fit with Panpsychism?

3.3.1 IIT and Panpsychism

IIT attracts panpsychists. Several sympathizers of panpsychism have cited IIT favorably (Chalmers 2017, Goff 2017, Rosenberg 2015, etc). To see why panpsychists find IIT interesting, let us note that, in the first place, since IIT attributes consciousness to any system with Φ^{MAX} , it allows small scale systems (such as photodiodes) or even subatomic entities to have consciousness (Oizumi et al. 2014; Tononi & Koch 2015). This point makes IIT particularly close to panpsychism, since we barely find any other theory of consciousness from cognitive science that allows consciousness at such micro-level scales.

Second and more importantly, IIT provides a systematic account of the derivation of consciousness from micro-level facts. IIT offers an account in which conscious structures can be incorporated into a further conscious structure, which effectively realizes the combination of consciousness. The key is IIT's identification of consciousness with Φ^{MAX} . As section 2 has shown, in IIT consciousness ($= \Phi^{\text{MAX}}$ in IIT) is mathematically derivable from causal facts about the elements of a system. Furthermore, if the components of a conscious creature are themselves smaller structures that can individually have consciousness, IIT can explain their joint contribution to the creature's consciousness by deriving the creature's Φ^{MAX} from the Φ 's of these individual components. For this reason, IIT seems capable of solving the combination problem. In more detail,

(i) **Generate macro-subjects, with IIT.** In the first place, if something X with non-zero Φ is appropriately connected with others such that the resulting system Y has a larger Φ , then IIT predicts that conscious system Y emerges from a combination that includes X. Thus Φ represents the principle that makes macro consciousness by properly organizing the parts (including parts that could be micro-subjects).

For example, suppose both ABC and DEF have positive Φ 's. If they both exist by themselves, their Φ 's would make them two conscious subjects. Now suppose that ABC and DEF are connected into ABCDEF which also has its own integrated information Φ_{ABCDEF} , and suppose that Φ_{ABCDEF} is the maximal Φ in the system ABCDEF. IIT would imply that ABCDEF emerges from ABC and DEF as a conscious subject at a larger scale. Thus, IIT offers a way to generate macro consciousness from consciousness at a smaller scale, if the construction of larger structures could produce a higher value of integrated information.

Note that IIT could account for the generation of macro-level consciousness. But IIT doesn't imply that consciousness is always determined by the causal structure/network in the system that has the most elements. For example, if in a system ABCD, $\Phi_{ABC} > \Phi_{ABCD}$, then Φ_{ABC} has priority in determining the consciousness of ABCD. The consciousness in ABCD could then be determined fully by the structure of ABC.

(ii) **Stop Combination, with IIT.** Next, according to IIT, Φ^{MAX} sets the boundary of consciousness. A system X may have a Φ_X which is larger than any other integrated information that can be derived from its other parts, and X may have nothing outside of it that can at that moment combine with it to constitute a further irreducible system X+ with

a Φ . Or, perhaps there is an X^+ with integrated information Φ_{X^+} , but it happens that Φ_{X^+} is smaller than Φ_X , and so Φ_{X^+} is not the Φ^{MAX} in the system. In these cases, IIT would imply that within this system X there is consciousness, and that X and its consciousness is not a part of a further conscious system. Thus Φ sets a boundary for conscious systems and stops unending generation of conscious individuals.

For example, suppose a system ABCD has integrated information, but ABCD is not connected with anything else into a larger system that has integrated information. Since only what has integrated information could have consciousness, in this case the consciousness determined by the integrated information in ABCD would be limited in ABCD.

Or suppose that ABCD has integrated information and is connected with one other outside entity E so that the system ABCDE also has integrated information. But suppose that $\Phi_{\text{ABCD}} > \Phi_{\text{ABCDE}}$. In that case, the Φ^{MAX} in system ABCDE would be determined by the sub-structure ABCD in ABCDE. Thus the consciousness in ABCD is still bounded in ABCD.

(iii) Merge subjects, with IIT. Moreover, because IIT requires that Φ^{MAX} represents the consciousness of a system whereas the non-maximal Φ 's (of the system's parts) which contribute to the Φ^{MAX} do not count for individual consciousness, Φ^{MAX} effectively represents a unification of the non-maximal Φ 's which are contributed by the system's

parts. Therefore Φ^{MAX} represents the single merged consciousness of the system. It leaves a system as only one conscious individual.⁴¹

For example, suppose both ABC and DEF have positive Φ 's. If they both exist by themselves, their Φ 's (Φ_{ABC} and Φ_{DEF}) could make them two independent conscious subjects. Suppose that ABC and DEF are connected into ABCDEF which has its own integrated information Φ_{ABCDEF} , and suppose that Φ_{ABCDEF} is the Φ^{MAX} in ABCDEF. Following the Postulate of *Exclusion* (section 2.2.(b), “**step 3**”), IIT would imply that ABCDEF has only one consciousness, i.e., the one determined by its maximal integrated information, Φ_{ABCDEF} . Φ_{ABC} and Φ_{DEF} in ABCDEF lose the status of maximal Φ 's, and so the individual consciousness of ABC and DEF no longer exist in ABCDEF, even though the structures of ABC and DEF and their Φ 's still contribute to the Φ^{MAX} in ABCDEF. More generally, since a system with integrated information has only one Φ^{MAX} , a system has only one consciousness, one conscious subject.

In sum: IIT defines the quantity Φ^{MAX} in terms of the causal structure within a system and uses this quantity to characterize the presence or absence of consciousness. With Φ^{MAX} , IIT allows the generation of macro consciousness from smaller conscious subjects. It sets boundaries that separate conscious systems from things outside. Also, it produces at most one unified consciousness in a system. Therefore, IIT suggests an account of consciousness which could satisfy the three desiderata on mental combination

⁴¹ As Mørch (2019) puts it, “According to the *Exclusion* postulate, only maxima of Φ are conscious, hence, whenever a complex system becomes conscious by gaining maximal Φ its parts will lose micro-consciousness...”

(section 1). For this reason, panpsychists may find IIT amenable to panpsychism and seek to combine the two.⁴²

3.3.2 Mørch on Panpsychist IIT

How to combine IIT and panpsychism? Hedda Hassel Mørch's (2019) proposal (henceforth "M-IIT") to combine panpsychism (more precisely, Russellian Panpsychism, i.e., RP) with IIT might be an example. M-IIT is supposedly "Russellian" since its author intends it to follow Russellian monism (RM). Recall (from Ch. 1, section 2) that RM is the thesis that physical sciences left out the fundamental intrinsic nature of the world, and that such fundamental intrinsic nature plays a crucial role in producing consciousness.

⁴² I should point out that the combination of panprotopsyichism+IIT is not a good position to begin with. The reasons against panprotopsyichism+IIT are the following. (1) First, a main challenge in panpsychism is the combination of subject, which is precisely where IIT can help with its account of Φ^{MAX} . Also, panpsychism accepts conscious entities at extremely small scale, which is also allowed in IIT. But panprotopsyichism differs from panpsychism in these two aspects. Panprotopsyichism doesn't have to commit to small-scale conscious entities; it is committed only to the existence of special fundamental properties that produce consciousness in creatures. Thus it doesn't have to grant consciousness to smaller-than-brain entities (as panpsychism and IIT do), and consequently doesn't have to address the combination of multiple conscious entities. Therefore, unlike panpsychism, panprotopsyichism won't profit much in its alliance with IIT. Moreover, given that many cognitive psychological or neuroscientific theories of consciousness avoid IIT's counterintuitive cases (e.g., conscious photodiodes), a panprotopsyichist may have a good reason to prioritize other theories over IIT. (2) Second, an IIT theorist should also be skeptic to panprotopsyichism, because panprotopsyichism undermines IIT's epistemic selling point. IIT's formal tools are meant to capture consciousness with naturally understandable concepts, e.g., cause/effect and mathematics. Characterizing consciousness with mathematical constructs such as Φ offers a sense of systematic clarity that other theories lack. Adding non-phenomenal quiddities ruins this clarity. Since non-phenomenal quiddities are neither physical (as characterized in science) nor phenomenal, there is no positive characterization of such quiddities except that they produce macro-level properties under (unspecified) right conditions. If Φ cooperate with such unknown elements, then the clarity of Φ isn't genuine.

Here I will call such fundamental intrinsic properties "quiddities", following Chalmers's (2015) use of the term.

In Mørch's proposed alliance between Russellian panpsychism (RP) and IIT, the main concern is that IIT identifies consciousness only at the level of causal arrangements that determines Φ^{MAX} ,⁴³ and therefore "allows for multiple microphysical realizability of systems with identical consciousness" insofar as the micro-physically distinct systems have the same Φ^{MAX} . This conflicts with RP, according to which all physical properties supervene on phenomenal quiddities, and therefore physically different systems must also differ phenomenally. For example, IIT allows a carbon-based system and a silicon-based system to be completely identical in their phenomenal property if they have the same Φ^{MAX} (thus IIT allows a sort of multiple realizability of consciousness), whereas RP implies that these two systems cannot be completely phenomenally identical—the physical difference means that at some level (maybe at a microphysical level of carbon and silicon atoms) there must be phenomenal difference, even though the carbon and the silicon systems may have highly similar or even identical phenomenal consciousness at the system level.

⁴³ And MICS (Maximal Intrinsic Conceptual Structure). Mørch's (2019) discussion of the issue with IIT (and Russellian Monism) focuses on the derivation of Φ^{MAX} , but the issue is applicable to the derivation of MICS, as both Φ^{MAX} and MICS of a system are determined by the internal causal structure of the system. For convenience of discussion, I will only mention Φ^{MAX} when I refer to integrated information, without distinguishing between Φ^{MAX} and "(maximal) intrinsic conceptual structure" in Tononi's sense.

One may think that there is a fix to this problem: couldn't the macro-consciousness of two physically different creatures be the same, even though their micro-consciousnesses are different?⁴⁴ Mørch (2019) has a reply to this fix:

One might think this conflict can simply be resolved by supposing that although silicon brains and organic brains have identical macrophenomenal properties, they have different microphenomenal properties. Silicon neurons (or molecules/atoms) would have silicon-type micro-experience, while the organic neurons (or molecules/carbon atoms) would have different carbon-type micro-experience, while both constituting the same human macro-experience...

But this is ruled out by IIT's Exclusion postulate, according to which consciousness never overlaps—only the system with maximal Φ is conscious. Neurons (or molecules/atoms) cannot have different experiences if they are excluded from having experiences of their own in the first place... (Mørch 2019, p. 1076).

According to IIT, given its postulate *Exclusion*, only the structure with Φ^{MAX} determines the consciousness of the whole system. The Φ 's of other structures in the system do not. As a result, in two systems with the same neural network structure but made of different materials, IIT would allow both to have the same macro-consciousness, but not the micro-phenomenal properties due to the non-maximal Φ 's. If the lower-level realizers have Φ 's and thus could have consciousness (if causally isolated from the system), IIT's

⁴⁴ Thanks to Joseph Levine and Alejandro Pérez Carballo for this point.

postulate of *Exclusion* entails that such Φ 's do not make any consciousness; only Φ^{MAX} works to make consciousness.⁴⁵

Since RP is against the sort of multiple realizability of consciousness that is allowed by IIT, combining RP with IIT requires eliminating the possibility of this multiple realizability of consciousness. This requires including factors other than Φ^{MAX} as ingredients of consciousness.

M-IIT provides an additional ingredient, but the additional ingredient is still in line with IIT, as it is in terms of Φ . Specifically, M-IIT considers non-maximal Φ 's of “fine-grained microphysical structure” as the supplementary determinants of consciousness: both the physical structure that determines a system's Φ^{MAX} (e.g., connection pattern of the network) and its more fine-grained structures with non-maximal Φ 's (e.g., the material that composes the network) contribute to consciousness, albeit to different aspects. More specifically, Φ^{MAX} contributes to the ‘structure’ of the consciousness, whereas non-maximal Φ 's determine conscious ‘qualities’. For example, as Mørch has imagined, a silicon duplicate of humans also sees tomato and grass as different and prefers pleasure over pain. So its phenomenal structure is about the same as ours, which is explained by the fact that its Φ^{MAX} is the same as ours. However, the

⁴⁵ What if the lower-level realizers have micro-phenomenal properties, not because of their Φ 's, but due to something else, e.g., their fundamental intrinsic nature? Mørch (2019) doesn't seem to have considered this possibility. In fact, IIT's *Exclusion* postulate apparently only excludes non-maximal Φ 's from contributing to consciousness; it does not explicitly exclude sources of consciousness other than Φ . But positing sources of consciousness other than Φ is against the general direction of IIT, of understanding consciousness in terms of causal structure and causal power. Non- Φ factors that determine consciousness are not considered at all in IIT. (See my discussion in section 3.4, this chapter.)

silicon being might experience qualities beyond our imagination, and so its phenomenal qualities differ from ours. This is explained by its material constitution that contributes to its distinctive non-maximal Φ 's. An advantage of this proposal is that it may simultaneously answer the “structure mismatch problem” (too little phenomenal structure; Ch. 1 section 3.2) and the “palette problem” (too many phenomenal macro-qualities; Ch. 1 section 3.2), because it allows that “the missing structure is somehow *encoded* in the extra qualities” (Mørch 2019), as the physical structures with non-maximal Φ 's contribute to phenomenal qualities but not to phenomenal structures of consciousness.

I should note that M-IIT doesn't seem "Russellian" enough for me, because M-IIT doesn't make much use of a key Russellian notion, i.e., quiddities. As a *Russellian* theory of panpsychism (RP), M-IIT should make certain conscious properties (i.e., phenomenal quiddities) the fundamental ground of microphysical properties and also explain consciousness in terms of these phenomenal quiddities. Thus M-IIT should explain how such quiddities contribute to a creature's consciousness under the framework of IIT. But M-IIT doesn't seem to address this question. It seems that M-IIT is all described in terms of Φ (which is all described in terms of causal powers). There is no work for quiddities.

3.3.3 Panpsychism unfit for IIT

Although panpsychists may have a special interest in IIT, panpsychism is arguably not a good metaphysics for IIT.

The first and most obvious problem with an attempt to combine IIT with panpsychism is that IIT doesn't need consciousness to exist in micro-level entities. In IIT, the "elements" do not carry consciousness. Systems with integrated information have consciousness, but there is never any mention that the elements have consciousness. Causal relations among elements are all that IIT requires from the elements. Therefore, if one adopts IIT, why not jettison the micro consciousness in ultimates?

Concerning Mørch's M-IIT, there is a more specific problem. That is, M-IIT doesn't cohere with IIT. Recall that the postulates of *Information* in IIT requires that physical parts must be causally related so as to carry 'information' and then contribute to integrated information. If a physical part cannot 'observe' and 'control' another part through causal relation (Oizumi *et al.* 2014), then even if both parts have Φ 's of their own, their juxtaposition doesn't make another Φ . M-IIT implies that consciousness consists of both phenomenal properties (identical to Φ^{MAX}) instantiated at a certain level of causal arrangement and other phenomenal properties (identical to all non-maximal Φ 's) instantiated at some lower-level causal structures. Therefore if M-IIT is true, IIT implies that the structures of Φ^{MAX} and non-maximal Φ 's are causally related. But this seems false. For example, it's unclear whether the relation between a neural network and its material constitution is a causal relation. It isn't the same as the relations among connected neurons. Although the different levels of physical structures have some dependence relations (for example, a neural network is *realized* by its material constitutions), that X realizes Y isn't the sort of relation that matters for the state of Y in the way that X's causing/being caused by Y gives Y 'cause-effect information' about X.

There is also a sense in which the causal relations in a neural network and its materials are causally segregated, for many subatomic/molecular/cellular events can happen within a neuron without affecting its behavior, as if the causal network at one physical level isn't causally connected to the causal network at another level. In fact, that's why we have Φ^{MAX} at one level, and other Φ 's at other levels, without the Φ^{MAX} and the non-maximal Φ 's fusing into a yet larger Φ^{MAX} . Therefore, given IIT's postulate of *Information*, the causal arrangements of Φ^{MAX} and non-maximal Φ 's must be causally related if they both are to figure into the consciousness of the system. However, it's unclear if the causal arrangements of Φ^{MAX} and non-maximal Φ 's are causally related in the way that they carry information about each other. Therefore, contrary to M-IIT, it's unclear how Φ^{MAX} and non-maximal Φ 's can contribute to a single consciousness of the system.

3.3.4 IIT and Russellian Panpsychism

Another problem with trying to fit IIT with panpsychism is that IIT doesn't fit well with its Russellian picture, which as I have explained (Ch. 1, section 2) is a significant metaphysical commitment of panpsychism. Recall that the Russellian metaphysics of panpsychism needs some phenomenal properties as quiddities to ground causal powers and consciousness in creatures. The problem is that, if quiddities themselves are phenomenal, the resulting Russellian IIT is incoherent.

If quiddities are phenomenal, as in Russellian panpsychism (RP), then these phenomenal properties should ground causal powers. For RM in general, physical properties as causal powers need categorical grounds. Phenomenal properties are

recruited in RP as the intrinsic grounds of physical properties and hence causal powers. For IIT it's the reverse: causal powers ground phenomenal properties. According to IIT, causal dispositions of the elements of a system ground its integrated information, which is identified with its phenomenal properties. Therefore for IIT, consciousness is grounded on causal powers. But, according to RM, consciousness grounds causal power. Since the property dependence directions in RP and IIT are opposite, assuming that grounding is irreflexive (i.e., one thing cannot ground itself), Russellian IIT seems incoherent if its Russellian elements (quiddities) are phenomenal.

An RP theorist might reply by hypothesizing that phenomenal quiddities are simply ungrounded—thus also ungrounded in causal powers, but all phenomenal properties of composite entities are, as IIT suggests, grounded in causal powers of Φ , which we can suppose are ultimately grounded in phenomenal quiddities.⁴⁶ In reply, I think this “mixed” RP version of IIT is coherent, but it brings worries for both IIT and RP. First, ungrounded phenomenal properties go beyond the theoretical machinery of IIT. Accepting this requires some concession from IIT. Second, even if we can ground phenomenal properties in composites with IIT, it's unclear what the role of phenomenal quiddities is in the framework offered by IIT for grounding the macro-level phenomenal properties of a system, except indirectly through grounding the system's and its components' causal powers. The result is a split metaphysical picture of phenomenal properties: phenomenal quiddities are ungrounded, while all other phenomenal properties are grounded but on causal relations. The explanation of consciousness also becomes

⁴⁶ I thank Joseph Levine for this suggestion.

fragmented: while IIT may explain macro-level consciousness, consciousness at the level of quiddities is not explained by IIT. This picture is inelegant. Thus if we follow RP, we should look further before choosing IIT.

3.4 IIT and functionalism

I have argued in section 3 that IIT doesn't fit well with pan(proto)psychism. At this point, a natural question to ask is: if IIT can be represented by a metaphysics in philosophy of mind, what metaphysics suits IIT the best?

IIT might look like a version of functionalism, but a close examination reveals that IIT's relationship with functionalism is more complicated than it may appear. IIT doesn't smoothly fit into the category of functionalism in the way that a computational theory of mind does. In this section, I will examine this complicated relationship between IIT and functionalism.

3.4.1 Functionalism

Functionalism became popular as philosophers rejected mind-body identity theory and accepted the multiple realizability of mental states. An identity theorist may treat pain as identical to C-fiber firing. However, if one recognizes that organisms such as octopuses or Martians can also have pain despite their physical differences from humans, one should recognize that it's more plausible to identify a mental state (e.g., pain) with the causal role it plays (e.g., normally caused by tissue damage, causing certain beliefs or desires, causing avoidance behavior, etc.). Thus, a functionalist holds that a mental state

shall be defined as a role within a causal network, rather than in terms of the physical matter that plays the role (Block 1980/2007). Following this, a functionalist account of consciousness will define a conscious state in terms of causal relations in the cognitive system, without specification of the material that realizes the system.

Section 2 has shown that IIT doesn't appeal to the material from which the system is made. There is no mention of brain states in IIT's machinery. Since integrated information is defined abstractly in causal terms, a system's integrated information does supervene on its physical state, but its material constitution doesn't have to affect integrated information. Therefore IIT seems closer to functionalism than to identity theories.⁴⁷

3.4.2 IIT: not exactly the same as functionalism

The multiple realizability of integrated information may suggest that IIT is a version of functionalism. However, there are reasons against this categorization of IIT. For example, Cerullo argues that IIT is not 'computational functionalism' (i.e., 'a combination of functionalism and computationalism'), because "Tononi claims that a feed-forward program running the same computations as the human brain would not be conscious and acknowledges that IIT allows for functionally identical philosophical zombies." (2015, p. 6). This idea can be illustrated by the example of a conscious system and its zombie counterpart, which is possible under IIT (Figure 1):

⁴⁷ Block (2009) has considered IIT as functionalism for a reason of this sort (that, IIT doesn't refer to neuroscience).

The two systems A and B run the same computation, because they have the same input-output profile. However, they differ in integrated information as measured by Φ^{MAX} , and thus differ in consciousness according to IIT. We can assume that in a human brain, many computations are implemented by A-type structures, which results in a high value of Φ^{MAX} . Suppose that there is a system with the same computations of a human brain but consists solely of B-type structures. Then we have a computationally identical system without integrated information, an *IIT-zombie*.

Peressini rejects treating IIT as a standard functionalist theory for a similar concern:

... a careful read of Tononi makes clear that two *functionally* equivalent neural arrangements can differ in their internal integrated information and hence won't have the same quality of experience. This could happen if the functional neural units had different levels of redundancies or if they implemented the function using a distinct but equivalent logical form... (2013, p. 203)

I agree that "a careful read of Tononi makes clear that two *functionally* equivalent neural arrangements can differ in their internal integrated information." (Ibid.) The mismatch between integrated information and function (considered as computation) suggests that IIT is not functionalism in the sense often linked to psychology by philosophers.⁴⁸ But I think this symptom of mismatch is due to two fundamental

⁴⁸ According to Block, mental states in functionalism are characterized "in terms of their causal roles, particularly, in terms of their causal relations to sensory stimulations, behavioral outputs, and other mental states." (1980/2007, p. 28) For other functionalist characterizations of mental states in terms of causal relations among psychological states, see Kim 2011, Ch. 6; Lewis 1972; Shoemaker 1981.

differences between IIT and standard functionalism: **(A)** IIT is a causal theory of consciousness, but the relevant causal relations are not psychological; **(B)** IIT is not *entirely* a causal theory, because although a system's integrated information is defined by the cause-effect information of its composing mechanisms, the state of integrated information itself is not defined by its *causal relations* with the mechanisms' states of cause-effect information.

A. To see that the relevant causal relations in IIT are not psychological, note that in standard functionalism, a state is a causal role defined according to its causal relations with other psychological states, whereas integrated information in IIT is not a state of causal role in this sense. A functionalist account of consciousness defines consciousness as a privileged type of mental state, whose privilege depends on its special psychological role in the mind. For example, it could be that there is a mental state representing another state in a certain way, and 'higher-order representationalism' would suggest that a mental state is conscious iff it is meta-represented in the way specified (Rosenthal 1997). Or, maybe a state can represent itself in a special way, as 'self-representationalism' maintains (Kriegel 2009), in which case being conscious just is self-representing in the right way. In contrast, IIT doesn't have to privilege any particular type of psychological functions to define consciousness, and indeed it doesn't. Instead, IIT introduces integrated information as a high-level property which is mathematically defined over all causal connections within a system. Consciousness and its features are then identified with this high-level property. What determines integrated information is the most complex interrelated causal arrangement in a system. To identify the most complex causal

arrangement, one should not assume that the causal relations must be individuated in psychological terms.

This difference also manifests in how IIT and functionalist accounts are developed. In functionalism, mental states are initially framed in psychological terms. For example, functionalism about pain starts by asking what pain causes or what causes it. Once we obtain a psychological theory of pain *in terms of* its causal relations with other sensory stimuli, behaviors, and mental states, we then consider the mechanism that implements this causally defined state. In contrast, IIT doesn't define integrated information with any psychological theory. Although it officially starts with phenomenological axioms, these axioms only provide clues on *what* phenomenal states *are like* in themselves, rather than how these phenomenal states relate to other mental states. Consequently, IIT's mathematical account is *formulated* in causal terms *without* referring to any phenomenal state, even though IIT's account of integrated information is *constrained* by phenomenological features stated by the axioms.⁴⁹

B. To see that IIT is not *entirely* causal, consider the fact that integrated information is not defined in terms of its relations with *other* states that are *external* to it. Although the “information” (cause-effect information) of a mechanism is defined by its causal relations with past/future states of other mechanisms, *integrated* information of a system is defined by the causal powers of the states that ‘compose’ rather than cause the system in its current state. A symptom of this distinctive feature of IIT is the following. For

⁴⁹ Therefore I agree with Peressini, when he says that “[Integrated information] aren't precisely functional properties in the standard sense, but rather... are what might be called ‘implementational properties’...” (2013, p. 203)

standard functionalism, the psychological states used to characterize a specific state are not simultaneous with the target state. For example, a functionalist definition of a belief B includes the fact that it tends to cause another belief B' as its logical consequence, and these two states B and B' are non-simultaneous. In contrast, in IIT, a system's integrated information exists simultaneously with all states that give rise to it, because a system's integrated information *at a certain time* is derived from the cause-effect information of the system's parts *at that time*. For example, for system ABC at time t , although the information of a mechanism is defined by its causation with past/future states (e.g., the characterization of the cause-effect information of A at t includes its causal constraints on the states of B and C at $t-1$ and $t+1$), Φ_{ABC} at t is determined by the information of its mechanisms *at t* . Therefore, despite being determined by causal relations, integrated information itself is not defined as a state that causes or is caused by future or past states. Rather, it resembles an irreducible composition of the states of cause-effect information that are its spatiotemporal parts. IIT's characterization of integrated information is therefore unlike standard functionalist characterizations of mental states.⁵⁰

The considerations above suffice to differentiate between IIT and standard functionalism. However, given the fact that a system's integrated information is causally determined and is multiply realizable (as noted at the beginning of this section), one may consider IIT and standard functionalism as belonging to a more general theory, according to which causal relations within a system fully determine its mental states (including conscious states). But standard functionalism of psychological states is stronger, because

⁵⁰ I am grateful to Joseph Levine for discussions on issues about functionalism.

it characterizes psychological states as causal roles in psychological theories, whereas integrated information in IIT is not a causal role in a psychological theory. Rather, integrated information is the strength of integration of a (not necessarily psychological) causal system.

3.5 Hylomorphism

I have argued that, contrary to what panpsychists may wish, IIT and panpsychism are not quite compatible (section 3). I have also argued that IIT is not standard functionalism. There are at least two points of difference between IIT and standard functionalism. **(A)** The relevant causal relations in IIT are not psychological: they are causal relations of the mechanisms that implement the cognitive system. **(B)** The state of integrated information is not defined by *this state's* causal relations with other states. In sum, IIT defines a state in terms of what happens *within* rather than *around* this state, in terms of causal relations that are *not* psychologically characterized. At most, IIT may belong to a more general theory, according to which the property in question (such as consciousness) is determined by causal relations (but not psychological states) in the system (section 4).

One can continue to ask what metaphysics of mind seems to best capture IIT. Given the earlier discussion, one may simply consider IIT as a brand of physicalism that is close but not identical to functionalism. That wouldn't be far off. But I think it is possible to consider a brand of metaphysics which, like IIT, also aims at characterizing some key category in terms of causal relations, but, unlike standard functionalism, is not

limited to using the causal relations in psychological terms, among states that happen diachronically in a process. I think that hylomorphism (of some sort) is such a brand of metaphysics. Also, note that the concepts in hylomorphism and IIT are developed (partly) with special concerns in the metaphysics of individuality. An essential concern of IIT is to define an *individual*, a unified field of consciousness within a system, in terms of the relation of the parts of the system. IIT considers consciousness as manifesting a kind of unity in a single perspective. The development of the concept of Φ^{MAX} in IIT (as shown in section 2, this chapter) aims at defining *one thing*, i.e., a unified conscious perspective. Similarly, the development of the concept of substantial form (to be explained in section 5.1, this chapter) also (partly) aims at defining *one thing*, i.e., a material individual.⁵¹ For these reasons, I propose that some sort of hylomorphism may be the best metaphysics for IIT.

3.5.1 An Account of Hylomorphism

Originated with Aristotle's metaphysics, hylomorphism dominated metaphysics until the early modern period. Recently hylomorphism has regained some momentum among philosophers. Several versions of contemporary hylomorphism have been proposed (e.g., Johnston 2006, Koslicki 2008, Rea 2011, Marmodoro 2013, Koons 2014, Jaworski 2016). There are probably more forthcoming.

⁵¹ Admittedly, standard functionalism can also help define an individual. But it requires an account of a psychological individual, as a system of psychological states. As explained earlier, the appeal to psychologically-defined causal states marks the key difference between IIT and standard functionalism.

Hylomorphism uses “form” and “matter” to explain the existence and changes of material objects.⁵² For example, a human being’s flesh and bones and other organs/tissues are the matter, while the biological structure that organizes the matter into a human being is the form. The flesh, bones, and organs/tissues need the form to make a human, while the biological structure of a human being must be embodied in the matter.

To understand hylomorphism, one must first understand that for hylomorphists, “forms” or “structures” are not merely ways of talking. Forms are ontologically “heavy” in hylomorphism. Almost every metaphysical position concerning mind and body acknowledges the existence of structure or organization in some sense. For example, a physicalist may argue that mental phenomena result from the microphysical particles organized in complex neurophysiological structures. But such talking of structures doesn’t manifest a commitment to the forms or structures in hylomorphism, because the physicalist may insist that these structures are merely derived from the spatiotemporal relations and physical forces among the particles. She might treat the notion of structures as theoretically helpful (maybe necessary), but refuse to treat them as metaphysically fundamental. In contrast, a hylomorphist would insist that forms/structures are fundamental. For example, the form of a dog is not the sum of the microphysical relations among the particles.

⁵² For those who follow a four-cause scheme of explanation, “efficient cause” and “final cause” are also often included. See Aristotle, *Physics*, Bk. II, §3, 194b24-195a2, and *Metaphysics*, Bk. V, §2, 1013a24-1013b3. Here and in what follows, all citations of Aristotle refer to Barnes (1984).

So, what is a form? What does a form contribute? In hylomorphism, a *substantial form* determines the existence of an individual object. The presence or lacking of a substantial form is thought to explain why a certain collection compose a real, unified individual or only a collection. For example, human beings (and organisms in general) are individuals due to their substantial forms, whereas a pile of bricks is merely a heap.⁵³

⁵⁴ Substantial forms are contrasted with *accidental forms*. Whereas accidental forms determine the properties which individuals can have contingently (e.g., whiteness is an accidental form that makes a man white), a substantial form determines whether an individual exists at all. Substantial forms therefore determine the existing individuals in the world: matters caught up in these forms make individuals.

Thus a substantial form unifies parts into an individual. Now what's the status of the parts in this unity? Do they count as individuals? Hylomorphists could make different choices on this question about the metaphysical status of parts. There is one approach to hylomorphism in which parts of an individual are not individuals. According to the hylomorphism from Thomas Aquinas, a substantial form unifies parts into a substance

⁵³ Pasnau 2011, sect. 24.2.

⁵⁴ Substantial forms are also supposed to determine the essences of individuals (e.g., whether one is a dog or a horse). Also, the acquiring, bearing, and loss of substantial forms are supposed to explain how new individuals generate, how individuals endure through certain changes, and how they are destroyed: "[W]hen the substantial form is introduced, something is said to come to be, without further qualification. But when an accidental form is introduced, we do not say that something comes to be, without qualification, but that something comes to be this; just as when a man becomes white, we do not say that he comes to be, absolutely speaking, but that he comes to be white. And to these two kinds of generation there correspond two kinds of corruption, namely corruption in an absolute sense, and corruption with qualification." (Aquinas, *On the Principles of Nature*, Ch. 1.) See also Pasnau 2011, sect. 24.2; Shields & Pasnau 2016, pp. 34-35.

and none of the parts of a substance has substantial forms. Since a substantial form makes something an individual and parts lack substantial forms, parts only exist “virtually”, not as individuals in themselves. Thus, for example, Kant has a substantial form and therefore exists as an individual, but Kant’s heart lacks substantial form and therefore is not an individual.⁵⁵ Hylomorphists may hold different views on the part-whole relationship, and so they can disagree on whether a heart in a body is itself an individual. The Thomistic view is unique in this aspect because it denies that parts have the status of individual objects.

The point that parts are unified into one in a substance is emphasized by Marmodoro,

... Since what is needed [to unify elements] is the shedding of only the distinctness of the elements, the role of this unifying principle [substantial form] must be just that: to strip the elements of their distinctness. (2013, p. 17)

Madden has explained the consequence of this unification in the following example,

[T]he parts of a material substance are not *substantially present* within the composition. That is, [Fluffy the cat’s] paw, tail, heart, carbon atoms, quarks, etc. are not present in her composition as *individual substances*... These parts of Fluffy are retrievably, potentially, virtually, or nominally present in Fluffy’s composition. Strictly speaking, there are no carbon atoms in Fluffy *actually as material substances*, but she subsumes the powers of such substances, and we could generate actual individuals of this kind by dividing Fluffy, for example, cutting small enough parts from Fluffy might produce individually existing carbon atoms. (2013b, p. 668)

To summarize, in the picture of hylomorphism that I have introduced, (i) there are ontologically “heavy” forms, (ii) some forms are substantial forms, which define the

⁵⁵ For the view that only one substantial form exists in a human body, see Aquinas, *Summa Theologiae*, Part I, Question 76, Article 4. See also Pasnau (2011), sect. 26.5; and Oderberg (2005), on the “unicity” of substantial forms, especially p. 84.

existence of individuals, and (iii) every object has only one substantial form, and therefore no part of an individual counts as an individual.

The hylomorphic account I have sketched is just one possible version of hylomorphism among many. But, as I will argue in the rest of this section, this hylomorphic picture fits well with IIT.

3.5.2 IIT and Hylomorphism

Recall that an ideal solution to the combination problem requires (i) generation of macro-subjects from micro-subjects, (ii) stopping combination within the boundary of an organism/its nervous system, and (iii) merging micro-subjects so that there is only one macro-level conscious subject in a conscious creature. We have seen that IIT could use Φ^{MAX} defined by a system's internal causal structure to achieve (i)-(iii). I think hylomorphism may be the metaphysics that could most closely capture IIT, because substantial forms in hylomorphism could be a general category which works in the same way as Φ^{MAX} .

To begin with, both Φ^{MAX} and substantial forms are supposed to define individuals: Φ^{MAX} defines a system as a conscious individual, with only *one* consciousness; a substantial form defines a material individual, as one object. With a definition of an individual, it is possible to figure out both the boundary of an individual and whether something qualifies as an individual, and therefore possible to address concerns like (i)-(iii) which are about the attribution of the status of individual consciousness/object.

There could be three ways to think about the connection between Φ^{MAX} and substantial forms. (a) One may consider substantial forms and the definition of material individuals to be more basic, and consider Φ^{MAX} and the notion of conscious individuals as derivative. (b) One may consider both Φ^{MAX} and substantial forms and their respective definitions of conscious individuals and material individuals as instances of a more general and basic notion of individuality. (c) One may consider Φ^{MAX} and the definition of conscious individuals as more basic and consider substantial forms and material individuals as derivative. In this section, just for the purpose of illustrating how IIT can be understood within the metaphysics ofhylomorphism, I will elaborate on option (a). The goal is to explain how Φ^{MAX} can be a derivative notion based on substantial forms and how this allows Φ^{MAX} to satisfy (i)-(iii).

How can Φ^{MAX} be a derivative notion of substantial forms? Recall that Φ^{MAX} is defined as a property of the internal causal structures within conscious creatures, while substantial forms are the individual-making internal causal structures within material individuals, including conscious creatures. Therefore Φ^{MAX} could be a certain aspect of substantial forms, esp. in conscious creatures. Also, substantial forms of creatures determine the power of creatures as individuals, e.g., growth, movement, and reaction to the environment. Φ^{MAX} can be the aspect of substantial forms that determine another power in individual creatures, i.e., consciousness.⁵⁶

⁵⁶ Hylomorphism can be a general account of the individuality of a system, in which IIT and functionalism offer different aspects of the substantial form: IIT accounts for the conscious perspective, while standard functionalism defines the psychological unity, the cognitive agent.

Now let me explain how this allows Φ^{MAX} to satisfy (i)-(iii). The basic idea is that substantial forms do something similar to (i)-(iii), concerning the individuality of material individuals.

(i) Generate macro-subjects, with HM. Recall that IIT could generate macro-subjects from micro-subjects, because for IIT the property Φ that defines subject/consciousness could “grow”: parts with Φ may connect in such a way that together they form a system of a larger Φ , in which case the larger system would have consciousness.

Thus intuitively, Φ is defined to allow other Φ ’s at higher levels of organization. Substantial forms as the individual-defining factors in hylomorphism are not explicitly defined in this way, but it is possible, in certain versions of hylomorphism, that substantial forms in smaller objects allow that certain organizations of these objects bring about a substantial form in the larger organized system. For example, oxygen and hydrogen atoms may come into a larger system, a water molecule. One cell, together with other adjacent cells, can constitute a functioning tissue or organ. One set of neurons, specialized for one function, may be connected with another set of neurons and make a more integrated system of functions. Thus hylomorphism may allow that one material individual be integrated into a larger material individual. IIT could be using the same logic in its account of conscious systems: the kind of internal causal structure that is responsible for consciousness can also “grow”.

(ii) Stop Combination, with HM. Recall that IIT stops combination by means of Φ^{MAX} , since Φ^{MAX} has boundaries determined by the internal causal structures of the systems. We can expect that Φ^{MAX} of a conscious creature is normally determined by a

system within this creature's body/nervous system. Thus, Φ^{MAX} can prevent extended consciousness.

Substantial forms in hylomorphism supposedly work likewise, since substantial forms offer a scope of composition based on the causal structure of the individual creatures. At most, only the matter caught up in substantial forms make individuals; not any random combinations of things count as individuals. Thus, although a dog exists because there is a dog form determined by the organization of this dog organism, supposedly there is no dog-and-the-leash individual because there is no dog-and-the-leash form. The leash isn't normally integrated into the dog organism (e.g., metabolism) and its functions (the leash hinders the dog's movement.) The world doesn't become one individual, because substantial forms set a boundary of composition within a limited causal structure. IIT is operating with a more explicitly characterized boundary of a conscious system, with a more explicitly defined causal structure.

(iii) Merge subjects, with HM. Recall that Φ^{MAX} in IIT is also defined in such a way that it merges conscious subjects: if smaller parts are connected into a system that has Φ^{MAX} , a value of integrated information that is higher than the values of Φ of the parts, then only the Φ^{MAX} counts. This helps panpsychism, because if panpsychists reject the idea of billions of micro-consciousness co-existing within a conscious creature, they need the subject combination to be such that ultimates (i) are micro-subjects with individual consciousness if they are not so combined, but (ii) would lose their status as individual micro-subjects once they are combined into a macro conscious creature.

In this respect, substantial forms from hylomorphism may also help. In particular, the substantial forms of the Thomistic view facilitate such a combination. As explained earlier (section 5.1), the individual-making function of substantial forms could include *both* the making of individuals *and* the cancellation of the individual status of the parts of the individuals. Multiple causal structures, each being an integrated system in itself, may combine into a strongly integrated causal structure, so that the functions of the substructures must be understood within the operation of the entire structure. Thus, for example, the substantial form of a dog makes an individual dog, but under the Thomistic view of substantial forms, the parts of the dog as they are organized in this individual dog do not exist as individual things themselves. These parts participate in the function of this individual dog, but they don't exist as individuals as the dog does. Each individual has only one substantial form, just as in IIT, each conscious system has only one integrated information that defines the system's consciousness (i.e., Φ^{MAX}).

To summarize, hylomorphism may be the metaphysics that has the right resource to accommodate IIT. IIT is in fact articulated as a theory of individuality (of conscious systems) in terms of causal structures. Meanwhile, hylomorphism (at least of the kind that I explained in section 5.1) is the metaphysics that explains individuality of material things, in terms of substantial forms determined by the internal causal structures of those things. Also, unlike pan(proto)psychism or standard functionalism, hylomorphism can do without the notions of micro (proto)phenomenal properties or functions defined in

psychological terms.⁵⁷ Therefore, it seems that among the available metaphysical theories, hylomorphism could be the one equipped with precisely the right categories for IIT.

Arguably, an IIT theorist can focus on articulating and testing her theory of integrated information without thinking in terms of the metaphysics of hylomorphism at all. But this doesn't mean that hylomorphism has no value for IIT. IIT and hylomorphism are at different levels of abstraction: if there is an empirical theory about some type of properties, then there is a metaphysics which provides the kinds of categories that this theory of specific properties will use, at a more abstract level. I take IIT as an empirical theory of consciousness, like other empirical theories of consciousness such as Global Workspace theory. Thus, it makes sense to ask what metaphysics, which ontology, suits the need of this empirical theory. My contention in this section is that hylomorphism of some sort is the metaphysics that offers the right sort of categories for IIT.

⁵⁷ Please note that the sort of hylomorphism that I introduced in section 5.1 is different from a more thoroughly *Thomistic* version of hylomorphism, since I have not included the Thomistic notion of prime matter. To my knowledge, contemporary hylomorphism can be broadly classified into Thomistic and non-Thomistic hylomorphism, according to the extent to which their philosophical notions are borrowed from Thomas Aquinas or the Aristotelian tradition represented by Aquinas. Contemporary Thomistic hylomorphism (e.g., Brower 2014) contains Aquinas's terms of substantial forms and prime matter, whereas contemporary non-Thomistic hylomorphism (e.g., Fine 1999; Jaworski 2011, 2016; Johnston 2006; Koslicki 2008) also contain some notions of forms and matter, but deviate from Thomism in various aspects. The hylomorphic account of section 5.1 doesn't refer to prime matter. Therefore, despite its notion of substantial forms which is close to that of Thomism, it is not a fully Thomistic brand of hylomorphism. In Ch. 4, I will discuss Thomistic hylomorphism, in which I will talk about prime matter.

3.6 Chapter Conclusion

In this chapter, I have argued that, even though some panpsychists may find IIT friendly to their metaphysical position, panpsychism doesn't really match with IIT. In fact, further reflection shows that a certain account of hylomorphism has a better account of the categories described in IIT. Therefore, panpsychists cannot support themselves by citing IIT as their scientific representative. Panpsychism and IIT are not really in good company.

This chapter also finishes Part I of this dissertation, which questions the metaphysical proposals of pan(proto)psychism. After reviewing the arguments for and the challenges to pan(proto)psychism (Ch. 1), I have reviewed and criticized several proposed accounts of pan(proto)psychism (Ch. 2), and have rejected the idea of supporting pan(proto)psychism by grafting it onto a respectable theory of consciousness in cognitive science, i.e., IIT (Ch. 3).

In Part II of this dissertation, instead of criticizing pan(proto)psychism on metaphysical grounds, I will criticize pan(proto)psychism as a proposed framework for further research into the nature of consciousness. I will argue that there is little prospect that more knowledge of consciousness will come out of the framework of pan(proto)psychism.

PART II

CHAPTER 4

RUSSELLIAN MONISM, THOMISTIC HYLOMORPHISM, AND THE MYSTERY OF CONSCIOUSNESS

4.1 Introduction

Recall that contemporary pan(proto)psychism mostly assumes the metaphysics of Russellian monism (henceforth RM), as Ch. 1 has explained. RM seems to have a metaphysics with something going for it, because its metaphysical framework has absorbed some advantages of both traditional physicalism and dualism while avoiding some disadvantages of both. Supporters of RM believe that physics only teaches us about the causal structures of the world but not the categorical bases that fulfill the causal roles in these structures. As a result, RM theorists believe that there must be some fundamental properties (aka. the inscrutable) which serve as the categorical bases of paradigmatically physical properties. Furthermore, they make the bold proposal that these fundamental properties that ground paradigmatically physical properties also ground consciousness. That is, the inscrutable is hypothesized as the fundamental factor that helps explaining consciousness, and our inability to explain consciousness is attributed to our ignorance of the inscrutable. The elegance of RM is that it seems to enjoy both the power of traditional dualism in accommodating consciousness and the theoretical unity of traditional physicalism. The inscrutable, which isn't included in traditional physicalism, offers RM

with further ammunition for explaining consciousness, much like the role played by metaphysically independent mental substances/properties in traditional dualism. On the other hand, since the inscrutable relate to paradigmatically physical properties through a metaphysical dependence relationship between dispositions and their categorical bases, the relationship between consciousness (or its fundamental ground) and physical properties in RM could be some form of metaphysical dependence, instead of the looser nomological association between the mental and the physical in traditional dualism. Therefore, RM has opened a niche in its metaphysics so that the inscrutable, as the candidate occupant of that niche, may provide an elegant account of both the fundamental ground of physical properties and the existence and features of phenomenal consciousness. (See Ch. 1, esp. section 2.2 for details of this argument for RM.)

RM could be criticized on many grounds.⁵⁸ For my part, I am concerned with the epistemic limitation of RM. Now that RM is bundled with pan(proto)psychism and is considered a significant footstep on the move *forward* on the problem of consciousness, it's worth asking what kind of knowledge RM could offer, if RM can be a fruitful research program. In this chapter, I will argue that, despite the appeal of RM in metaphysics, RM cannot become a fruitful research program, in the sense that RM cannot guide us toward a specific, detailed explanation of consciousness analogous to the sort that modern atomism can suggest for physical sciences. In fact, RM's "elegant" metaphysical framework⁵⁹ and lack of details make me suspect that RM is akin to a once-

⁵⁸ Ney 2015; Kind 2015; Pautz MS; Stoljar 2014; etc.

⁵⁹ To quote the title of Chapter 6 in Goff 2017b, "The Elegant Solution".

upon-a-time prominent metaphysics, i.e., Thomistic Hylomorphism (henceforth “TH”), which was rejected by most philosophers after the early modern period.⁶⁰ Generally speaking, hylomorphism (matter-form-ism) offers “matter” and “form” as two necessary components of material individuals. TH in particular uses “substantial form” and “prime matter” as two basic notions to explain material objects. As I will argue, RM and TH are similar methodologically. Each has a metaphysical system in which some beings are assigned significant metaphysical functions, to serve as “fundamental stuff” of some sort or to work as principles of unity or individuation. Because of the similarity between RM and TH, I suspect that revisiting the path and peril of TH might aid philosophers with a glimpse of RM’s future.

Being a contemporary philosopher but not a scholar in medieval philosophy, I acknowledge the richness of TH but am concerned with its limitations. In this paper I will not discuss the metaphysical merits or shortcomings of TH⁶¹, but will focus on its epistemic limitations. It is well-known in philosophical history that TH’s prime matter has been criticized as unintelligible while substantial form is criticized as unknown and limited in explanatory power. Given the epistemic inadequacies of TH, the similarity between RM and TH suggest that RM may face similar epistemic challenges. In

⁶⁰ Hylomorphists can be broadly classified into Thomistic and non-Thomistic hylomorphists (Ch. 3, section 5.2, fn. #57). In this chapter, I will focus on TH, because Thomistic theories use the notion of “prime matter.” As I will argue (section 4.1), this makes TH closer to RM. Also, since my purpose in this paper is to show the inadequacy of RM, it is sufficient if I show that RM has an inadequate companion (i.e., TH) that is similar to it in the relevant aspects. Perhaps there are better hylomorphic accounts than TH, but whether such hylomorphic accounts exist is irrelevant to my purpose.

⁶¹ For a defense of TH against some objections from contemporary metaphysical views about material objects, see Brower 2014, Ch. 7.

particular, perhaps the inscrutable in RM is unknowable. Moreover, the unknowability of the inscrutable may bring a further difficulty to RM, i.e., that it is unknowable how the inscrutable, which is supposed to exist in fundamental micro-level entities that compose all things in the universe, explains the physical properties of these entities or the consciousness of the animals which they compose. This further difficulty of RM is reminiscent of the question of how substantial forms in TH arrange prime matter. If these concerns are right, then to the extent that TH has contributed little to human knowledge of material beings (if we agree that it was the scientific revolution and its underlying mechanical philosophy that brought breakthrough to human knowledge in the material world), then there is a reason to worry that RM's contribution to knowledge of consciousness will be limited likewise.

As I will argue, the similarities between TH and RM can be summarized as follows. (1) Both RM and TH are centered around the idea of a fundamental ground of things in some sense, and the way such fundamental ground is defined precludes our knowledge about such ground. In the case of TH, there is a concern about the ultimate substratum that endures through all changes. Thus TH requires a notion of prime matter as formless. Since knowledge of something requires grasp of its form, the formlessness of prime matter makes it unknowable. As for RM, there is a concern about the ultimate intrinsic nature that realizes all causal roles, and thus RM needs a notion of the inscrutable as devoid of dispositional nature in itself. But since we acquire knowledge of something through the manifestation of its disposition, the disposition-less inscrutable is unknowable. Even though we may learn about the inscrutable indirectly by the

dispositions to which it contingently associates, we cannot tell what the inscrutable is in itself. (2) Both RM and TH have an interest in establishing the individuality of certain familiar things, which pushes both to consider principles of unity and individuality that may well turn out to be unknowable. TH needs to account for the individuality of ordinary objects, hence the TH notion of substantial forms, which has been widely considered as unknown and explanatorily limited. RM needs to account for individual consciousness in ordinary human beings, and so RM theorists belabor on “the combination problem” (i.e., how the combination of micro-level entities with fundamental intrinsic nature produces consciousness in complex entities like human beings), with little progress. In sum, both RM and TH look for metaphysical factors that serve as fundamental grounds of some sort or as the maker of some individuals in ordinary sense, and I contend that the way of defining fundamental grounds and ordinary individuality in its metaphysics limits the epistemic value of RM just as the epistemic value of TH is limited.

Thus, I will argue that RM is like a contemporary reincarnation of TH in the topic of consciousness, both in its methodology of metaphysics and its limited epistemic prospect. To reach this conclusion, I will proceed as follows. In section 2, I explain the epistemic challenges to RM. In section 3, I explain TH (section 3.1) and review some familiar epistemic inadequacies of TH (sect. 3.2). In section 4, I explain the similarities between TH and RM (section 4.1) and argue that the epistemic challenges of TH and RM are similar (section 4.2). Section 5 concludes.

4.2 Epistemic Inadequacy of Russellian Monism

As Ch. 1 of this dissertation (and section 1 of this chapter) has explained, RM has some advantages among metaphysical theories of consciousness. But RM faces an interesting challenge: it seems that we can hardly gain knowledge about its core elements. The fundamental ground of all properties that RM emphasizes is the inscrutable, the fundamental intrinsic nature of all other properties, including consciousness in macro-level creatures and the structural-dynamic properties that physics studies. But there is little prospect that we can know about the inscrutable. This also constrains further investigation into how the inscrutable may contribute to more familiar forms of individual consciousness at the macro level.

4.2.1 The Inscrutability of the Inscrutable

According to RM, the inscrutable is devoid of dispositional nature in itself. Therefore its relationship to all causal roles is contingent. The disposition-less inscrutable as a fundamental ground in RM metaphysics is unknowable.

David Lewis (2009) has argued for our inescapable ignorance of fundamentally intrinsic properties by arguing for “Ramseyan humility”.⁶² Lewis invites us to consider the “true and complete” final theory T. According to Lewis’s (1970, 1972) account of theoretical terms, a theoretical term only characterizes a causal role that some property can take with regard to other properties. Also, in Lewisian metaphysics, the causal roles

⁶² But see Ney (2015) for a disagreement from a physicalist perspective.

are ultimately realized by certain fundamental intrinsic nature (a.k.a. “quiddities” for Lewis), which are only contingently related to any role characterized by theoretical terms. Therefore Lewisian metaphysics of fundamental properties depicts the same picture that RM defines in terms of the inscrutable and causal roles (section 1.2). Crucially, Lewis argues that our knowledge only reaches the causal roles but not the fundamental role-players. Theory and empirical evidence only tell us what causal roles are played, i.e., what roles quiddities play, but not the quiddities themselves. For example, if there is a fundamental property of mass, our theory only describes the mass-role but not what quiddity plays the mass-role. For another example, suppose that the unit-negative-charge role is realized by the quiddity P1 while the unit-positive-charge role realized by the quiddity P2; P1 and P2 are only contingently related to their actual posts in the theoretical structure. There are at least two ways for quiddities to detach from the roles they play. First, it is possible that P1 and P2 switch the roles they play without changing the causal roles. We cannot tell which quiddity is playing which role. Second, there might be an alien quiddity Px, which may take the role that P1 takes, and we cannot tell whether it’s P1 or Px that is performing the role. Because for any hypothesis of the distribution of quiddities (which Lewis calls a “realization” of T) in the world, there could be competing

hypotheses but no way to tell which hypothesis is actual, we are locked in ignorance of quiddities. Humility about quiddities is therefore the right attitude.^{63 64}

Jonathan Schaffer (2005) has criticized Lewis's Ramseyan humility argument as an argument of skepticism which is on a par with e.g. Descartes's radical skeptical argument regarding the eternal world. Schaffer therefore thinks that Lewis's humility argument isn't more plausible than such radical skeptical arguments. However, it's unfair to treat Lewis's humility argument as only one of skepticism. Lewis's presentation may have suggested itself as an argument of skepticism. But this presentation/interpretation has unnecessarily granted *too much* to RM theorists or anyone who believes in "quiddities". In a skeptical argument that goes against a certain type of knowledge K (e.g., of the eternal world), the target K under the skeptical attack is something that we understand. Of an instance of K (e.g., whether I have hands or not), we tend to know what it is for it to be true. And we often have some well-motivated grounds, sometimes

⁶³ "Suppose [T] does indeed have multiple possible realizations, but only one of them is the actual realization. Then no possible observation can tell us which one is actual, because whichever one is actual, the Ramsey sentence will be true. There is indeed a true contingent proposition about which of the possible realization is actual, but we can never gain evidence for this proposition, and so can never know it." (Lewis 2009, p. 207)

⁶⁴ What if the inscrutable is homogeneous? That is, what if the inscrutable is of one kind, like P1? There are at least two problems with this proposal. First, the homogeneity of inscrutable doesn't fit well with the idea of multiple different properties in the fundamental laws of the complete theory (which is implied in Lewis's (2009) humility argument by assuming that fundamental intrinsic properties that play the theoretical roles have different "adicity"). Different theoretical properties should be based on different categorical natures, intuitively. Otherwise something else is necessary to explain why there is only one fundamental intrinsic nature but more than one fundamental dispositions. Second, if the inscrutable is homogeneous all over, we still cannot tell what it is. There can be multiple competing hypotheses of what this homogeneous and ubiquitous inscrutable is (e.g., whether it is P1 or P2), and it follows from Lewis's humility argument that no evidence can tell which hypothesis is true.

Moorean, to hold K true. But the target of Lewis's doubt, i.e., that a quiddity that plays a certain theoretical role is of a certain sort, is such that one can only *suppose* that one knows or justifiably believes *which* inscrutable (P1 or P2, etc.) is doing a certain job, with "P1" and "P2" as mere labels like "this intrinsic nature that actually serves unit-positive-charge role" or "this intrinsic nature that actually serves unit-negative-charge role". Still, nothing is known about the things themselves that P1, P2 stand for. Therefore, the humility argument is not doubting that "I have a hand" by appealing to the possibility that I am a brain-in-a-vat. The stories of quiddities of which Lewis is pleading ignorance are those of which we have no clear understanding of any of them and no *prima facie* evidence to tell which one is true.⁶⁵

The unknowability of the inscrutable is unlike the not-directly-observable entities postulated in microphysics. Microphysics is filled with theories about the features of the micro-level entities, which are supposedly testable with empirical means. Even if subatomic particles are invisible, microphysics about them still provide clues as to the empirical conditions for these theories to be tested. For example, sub-atomic particles at least have measurable properties such as charge, mass, angular momentum, even if we cannot directly observe a single particle. In contrast, of the inscrutable in RM, Ramseyan humility implies that no attribution of positive features to the inscrutable can be empirically justified in RM, and so RM supports no positive account of the inscrutable. Thus we seem to be doomed to remain ignorant of the inscrutable.

⁶⁵ See Dustin Locke (2009, pp. 228-34) and Dasgupta (2015, p. 474) for similar criticisms of Schaffer.

Ramseyan humility should be sufficient to warn against speculations on the inscrutable/quiddities toward any further specificity, e.g., of whether the inscrutable is phenomenally reddish or phenomenally greenish, or whether the inscrutable is phenomenal at all, etc. Within RM, we cannot tell.⁶⁶ To speculate on such specific features of the inscrutable is to go beyond knowledge, or at least, beyond any theoretical structure.

If Lewis's Humility Argument is correct about the nature of our ignorance on quiddities (=the inscrutable in more explicitly RM terms), then there is no *empirical evidence* for making the choice between different hypotheses about the nature of the inscrutable, only *philosophical concerns*. For example, if one thinks that phenomenally conscious micro-entities are implausible,⁶⁷ then one may prefer protophenomenal inscrutable. On the other hand, if one thinks that protophenomenal inscrutable is unknowable but phenomenal inscrutable is knowable, and that there should not be unknowable things, then one has a reason to prefer phenomenal inscrutable. Even so, the most "specific" speculations RM may justifiably reach are still extremely unspecific. In the rest of this section I will discuss the only "specific" accounts of the inscrutable that I think RM could motivate, i.e., panprotopsychism and panpsychism, and I will argue that

⁶⁶ Because of this ignorance about whether the inscrutable is phenomenal or not phenomenal, Amy Kind (2015) argues for a pessimism toward RM's prospect of solving the mind/body problem. According to Kind, RM brings forth the inscrutable, but then the debate between physicalism and dualism becomes the debate between panprotopsychism and panpsychism, of whether the inscrutable is in itself phenomenal. This is not solving the mind/body problem, but relocating the problem to another ontological level.

⁶⁷ Or, if one thinks that the "combination problem" of conjoining micro-consciousness cannot be solved (Ch. 1, section 3).

the “specificity” of the inscrutable in these accounts still leaves the inscrutable unknowable. Since my concern is about the epistemic prospect of RM, in the remaining of this section I will discuss only those philosophical considerations about the epistemic prospects of RM panprotopsyism and RM panpsychism (besides the Lewisian consideration expressed in Ramseyan humility, which to begin with shows that none of them can be preferred on empirical grounds).

Consider RM panprotopsyism first, which hypothesizes that the inscrutable is protophenomenal. So understood, the inscrutable in RM panprotopsyism is unknowable. Recall that according to panprotopsyism, fundamental entities with the inscrutable are not themselves phenomenal, but owing to the inscrutable can give rise to entities with phenomenal consciousness when organized in the right way. Note that the protophenomenal inscrutable is not described in terms of what it is in itself. It is characterized only in terms of what it is supposed to achieve under certain arrangements (i.e., the inscrutable makes conscious creatures under combination principles). By positing protophenomenal inscrutable, what RM panprotopsyists have really achieved is (i) positing some property which performs the function of enabling fundamental entities to be arranged into macro-level creatures that have consciousness, and (ii) refusing to offer any currently available positive characterizations of such property, insisting that it is neither paradigmatically physical nor phenomenal—which de facto excludes any positive characterizations that we currently could conceive of. Thus, protophenomenal inscrutable is denied characterization in terms of any known positive

feature (save its role as the fundamental intrinsic nature of physical and phenomenal properties).

Then consider RM panpsychism, according to which the inscrutable is phenomenal. If the inscrutable is phenomenal, then it seems that we may know *a little* about it, since it seems intuitive that we know what is for something to be phenomenal. We know many instances of phenomenal experience, from which we may have some grasp of what *being phenomenal* is versus what is not to be phenomenal (e.g., zombies), even though it's hard to define "being phenomenal". Furthermore, phenomenal inscrutable seems to explain more than non-phenomenal inscrutable, since the hypothesis that micro-entities have *phenomenal* inscrutable seems to explain why they compose animals that have *phenomenal* consciousness, for in general it seems easier to understand why something has a property if its components also have the same type of property (although in smaller quantities).⁶⁸ So, one may argue that, if the inscrutable is phenomenal, then we are not entirely ignorant about it.

However, besides the concern that there is no empirical test for the phenomenality of the inscrutable, the proposed intelligibility of phenomenal inscrutable is subject to two

⁶⁸ See Bruntrup (2017), who argues that (emergent) panpsychism is superior to emergentist dualism because of the "genetic argument": that, phenomenal properties must come from (simpler) phenomenal properties. According to emergent panpsychism, the micro-level entities do have phenomenal properties, but the phenomenal properties of macro-level creatures are something over and above the micro-level phenomenal properties. In contrast, according to emergentist dualism, micro-level entities have no phenomenal properties. Bruntrup takes emergent panpsychism to be tolerable, because despite its "strong emergence," it follows the principle that phenomenal properties only come from (simpler) phenomenal properties. Bruntrup regards emergent dualism as intolerable, because it requires "superstrong emergence", i.e., that phenomenal properties can come from non-phenomenal ones. (sect. 2.4.3.)

limitations. **First**, phenomenal inscrutable are also supposed to be the intrinsic nature of microphysical properties according to panpsychist RM. Therefore we should expect phenomenal inscrutable to explain microphysical properties. Yet it is unclear how such an explanation could go. **Second**, more importantly, there is no way to be any more specific about what phenomenal inscrutable is like than the most general characterization that it is *phenomenal* (period). This much *minimal phenomenality* of the inscrutable seems to be all that RM panpsychists can hold with justification. No more. The minimal phenomenality of the inscrutable may be motivated or justified by the merely-philosophical considerations mentioned above, i.e., to explain consciousness and to avoid being condemned to likely eternal ignorance if the inscrutable is protophenomenal. But what can further support any more specific hypothesis of the feature of phenomenal inscrutable, e.g., whether it is phenomenally reddish or greenish? Nothing. If our consciousness manifests some universal reddishness, i.e. if we always experience something reddish whenever we are conscious, then there may be a reason to hypothesize that the inscrutable that underlies all macro-consciousness is phenomenally reddish in some sense. However it doesn't seem that any phenomenal quality is ubiquitous in all consciousness. Therefore, if RM might have some reasons to motivate mere phenomenality of the inscrutable, which may justify a *minimal panpsychism* in the sense that the fundamental entities are phenomenal (period), any further speculations on *what phenomenal property* the inscrutable is is unmotivated. Therefore RM panpsychism would still leave the inscrutable largely unspecified.

4.2.2 The Puzzle of the Combination Principle

I have argued in section 2.1 that the inscrutable in RM is unknowable. In this section, I will argue that there is another unknowable element in RM.

RM starts with a pair of metaphysical notions (i.e., dispositional nature/categorical basis) instead of some intuitions about ordinary individuals, but RM needs to account for individual consciousness in ordinary human beings. According to RM, the inscrutable as the fundamental intrinsic nature is the ground of consciousness in ordinary human beings. But RM still needs *something* to make a macro-level consciousness, given the micro-level components of a human body that has the inscrutable as their fundamental intrinsic nature. Let's call whatever that performs this macro-level-conscious-individual-making function "Combination principles", since the making of such conscious individuals presumably depends on the combination of micro-level entities.⁶⁹

Now consider how much we can know about combination principles. For anything that functions as a combination principle, we know what they are supposed to achieve, i.e., consciousness and other macro-level properties. We can understand a combination principle as *that arrangement of the micro entities with the inscrutable which results in the macro-level properties in question (including macro consciousness)*. But this account of combination principle is too general. Can it be more specific?

⁶⁹ For RM panpsychists, combination principles are needed to explain the combination of micro consciousness into macro-level consciousness. For RM panprotopsyichists, combination principles shall explain the combination of micro-level (proto)conscious entities into macro-level conscious animal.

For the time being, no combination principle has been specified, which is shown in the combination problem.⁷⁰ If the combination problem remains unsolved, then no specific account of combination principles will be available. The question about the epistemic prospect of combination principles is equivalent to the question of the combination problem.

Can we solve the combination problem? Or, are combination principles ultimately unknown? There is a good reason to be pessimistic, that is, knowledge of the micro-entities which have the inscrutable and knowledge of their combinations are interdependent. The idea is simple: to know how things with *P*-type properties combine is to know what results from a set of things of *certain specific P-type properties* arranged *in a certain way*. Knowledge of how combinations happen always includes not only knowledge of the end results of combinations, but also knowledge of the specific features of the combined items in virtue of which the combinations happen. For example, explaining how several spices together make a certain flavor requires knowing something about these spices, including their flavors and the ratio of mixture. Similarly, explaining how bricks together form a house requires knowledge about the bricks themselves, including their sizes and shapes and impenetrability (by other bricks). Thus, if we believe that fundamental micro-entities have the intrinsic nature known as “the inscrutable” but know nothing about the inscrutable itself, then it is like believing that an object is made of “matter” but has no idea of what matter is. No knowledge of the arrangement of such

⁷⁰ For more on the combination problem, see Part 1 of this dissertation, esp. Ch.1 and Ch. 2.

unknown matter can get off the ground, except that there is some arrangement of the unknown matter. Thus, if RM theorists stick to the inscrutable as the fundamental intrinsic nature, then since the inscrutable is unknowable, combination principles are likewise unknowable.

Indeed, ignorance of the “units” and of their combinations runs both ways. We have seen that the unknowability of the inscrutable renders knowledge of combination principles unattainable. On the other hand, since combination principles are responsible for relating fundamental micro-entities to make macro-level phenomena that we know, ignorance of combination principles also blocks the road of knowing the relevant features of the entities combined (in this case, the inscrutable) by inferring the nature of these micro-entities themselves from more familiar macro-level phenomena. The challenge of knowing the inscrutable is thus an inverse problem of the combination problem. While the combination problem challenges an RM theorist to explain the macro-level phenomena (including consciousness) from the inscrutable at the micro level, the challenge of knowing the inscrutable is about inferring the inscrutable at the micro level from the macro-level phenomena. In RM, therefore, ignorance of the inscrutable and ignorance of combination principles reinforce each other.

4.2.3 The Significance of the Epistemic Inadequacy

I have argued that RM faces epistemic challenges. The inscrutable in RM is unknowable (section 2.1). Also, RM needs an account of the combination principles of the micro-level entities that have the inscrutable. However, given the unknowability of

the inscrutable and the interdependence between knowledge of the to-be-combined items and knowledge of how they combine, the combination principles are also unknowable. The combination problem is a symptom of such ignorance of both *the elements* and *their arrangements* in the domain of phenomenal consciousness (section 2.2).

What does this mean for contemporary supporters of RM? Indeed, one may ask what is wrong with the unknowable inscrutable and the combination principles? If RM is just a metaphysics, then maybe these challenges are not so serious. If an RM theorist is satisfied with developing a new metaphysics that promises to “naturalize” consciousness, then fair enough. I am not arguing against RM metaphysics in this chapter. However, it seems to me that RM is not just one more metaphysics in which the primary target is only to establish or defend the truth or credence of its theses. RM seems to be considered a “research program” or something similar, such that RM is not supposed to permanently remain an outline with only some metaphysical categories; about the metaphysical categories, there seem to be facts of details awaiting discovery, so that an explanation of consciousness could be *given* instead of merely promised. If so, then the epistemic challenges to RM must be taken seriously.⁷¹

It seems to me that RM is supposed to do more than a metaphysics that just offers new categories, eliminates/revises old categories, and/or re-organizes systems of categories. Contemporary RM theorists seem to have the ambition of new knowledge, of a new domain of consciousness “science”, where they expect new voyage and

⁷¹ I thank Joseph Levine and Alejandro Perez Carballo for their advice on this point.

discoveries. For instance, in Chalmers's concluding remark to his assessment of RM panpsychism/panprotopsychism and the combination problem, he wrote,

If we can find a reasonable solution to the combination problem for either [RM panpsychism and RM panprotopsychism], this view would immediately become the most promising solution to the mind-body problem. So the combination problem deserves serious and sustained attention. (2015, p. 274)

There is no guarantee in this passage that an RM panpsychist/panprotopsychist theory will solve the problem, but apparently there is the hope that more can be said about such theories. In general, current literature on panpsychism/panprotopsychism tends to assume that the study is about a domain of reality and some discovery is underway, whether it is some “mental chemistry” (Coleman 2012) or emergent laws (Seager 2010) or “phenomenal bonding” (Goff 2017b).

In a recent paper about philosophers's intuitions about consciousness and the relevant psychological processes (“meta-problem intuition” and “meta-problem processes”), Chalmers (2018) expresses a similar hope toward RM. In this paper, RM is hinted at in a view known as “realizationism,” as the inscrutable “realizes” the nomological roles characterized by paradigmatic physical properties.⁷² Chalmers

⁷² “[According to meta-problem realizationism] consciousness plays a role in realizing meta-problem processes. ... [T]heorists may hold that... consciousness realizes some of [the meta-problem processes], thereby playing a causal role with respect to their outcome. Perhaps panpsychist consciousness plays a role in physical dynamics. Perhaps interactionist consciousness plays a role in high-level dynamics... [Some] biological materialists may hold that consciousness is essentially biological and realizes computational processes that generate phenomenal intuitions. Likewise, some quantum-mechanical materialists may hold that consciousness is a quantum process that realizes the meta-problem processes.” (Chalmers 2018, p. 42)

acknowledges that “realizationism” faces challenges, but he is optimistic about its prospect:

I think the most promising view [about the explanation of meta-problem processes] is realizationism. The research project for the realizationist is to spell out a satisfactory version of the view showing how consciousness realizes meta-problem processes... (Chalmers 2018, p. 49)

This expectation of discovery seems most clearly and ambitiously expressed in Goff’s comment on Galileo’s role in promoting the mathematical method in modern science:

Before Galileo, philosophers took the world to be full of *sensory qualities*: colors, smells, tastes, sounds. And intuitively one cannot capture the redness of a tomato, the spicy taste of paprika, or the sweet smell of flowers in the austere, abstract language of mathematics. Galileo got around this problem by stripping the world of such qualities and locating them in the soul... By stripping external objects of any qualities other than shape, Galileo created a metaphysical picture of the material world in which it could be exhaustively described in mathematical geometry... (Goff 2017a, pp. 11-12)

Then Goff claims that this move of limiting the scope of inquiry isn’t sufficient for the knowledge of the “fundamental reality”:

... The success of mathematical physics resulted from *limiting the scope of enquiry*; by putting the sensory qualities we encounter in conscious experience—colors, smells, tastes, sounds—outside of the domain of the physical sciences, we are able to give a purely mathematical description of what’s left over. But those qualities that Galileo took out of the material world still exist *somewhere* and must still be accounted for *somehow*... At some point those qualities need to be put back into our metaphysical picture of the world. (Ibid., pp. 13-14)

If this passage expresses a common sentiment among RM theorists, then it should leave no doubt that RM theorists take phenomenal properties as a domain of inquiry and expect that some inquiry in this “new science” (perhaps an entire field beyond mathematical physics) is possible. However, if my arguments in this section is correct, then RM theorists’ hope is empty. RM cannot develop as a research program.

To make the seriousness of this epistemic challenge clearer, I will cite a philosophical approach which may serve as a historical precedent: Thomistic hylomorphism (TH). As a brand of metaphysics, TH may face some challenges. But what’s historically more interesting might be the fact that TH never served as a fruitful base for a research program, as scholastic philosophy based on Aristotelian hylomorphism failed in general as a research program about the natural world. Let’s turn to section 3 for this tale of TH.

4.3 Thomistic Hylomorphism

4.3.1 Thomistic Hylomorphism Introduced ⁷³

Hylomorphists in general use the notions of form and matter.⁷⁴ Thomistic hylomorphists, in particular, account for the existence and nature of things in terms of *substantial forms* and *prime matter*.

Form and Matter. The distinction between form and matter was introduced by Aristotle for the purpose of explaining changes.⁷⁵ According to commonsense, a change is a change of *something*, and therefore something should endure through changes. That which endures through a change (whatever it is) is called “matter”. Meanwhile, a change also involves something that comes or goes with regard to the enduring matter. That which comes or goes during changes is called “form”.

In general, matter can be construed as that which composes an individual when “informed” of the formal aspect. In Thomistic terms, the matter of an individual of a

⁷³ My knowledge of TH is indebted to the following sources: Aquinas, *On the Principles of Nature*, translated by Gyula Klima; Brower 2014; *Summa Theologiae*, Part I, Questions 75-76, 78-79, 84-86, translated by Robert Pasnau and Christopher Shields (n.d.); Madden 2013a; Pasnau 2011; Shields & Pasnau 2016.

⁷⁴ For those who follow a four-cause scheme of explanation, “efficient cause” and “final cause” are also often necessary for a complete explanation of the existence of something as having a certain feature. See Aristotle, *Physics*, Bk. II, §3, 194b24-195a2, and *Metaphysics*, Bk. V, §2, 1013a24-1013b3. (Here and in what follows, all citations of Aristotle’s work refer to Barnes (1984).) Also see Aquinas, *On the Principles of Nature*, Ch. 3. Here we are only concerned with “matter” and “form,” because they are the components of objects themselves: “[M]atter and form are said to be intrinsic to the thing, for they are constituent parts of the thing; but the efficient and the final cause are said to be extrinsic, for they are outside of the thing.” (Aquinas, *On the Principles of Nature*, Ch.3.)

⁷⁵ *Physics*, Bk. I, §7.

certain type is what *potentially* exists with regard to being an individual of this type, while the presence of form makes the individual exist in *actuality*.⁷⁶ For example, a lump of clay may be potentially a statue, in the sense that by itself the lump doesn't constitute a statue, but if a form of statue is given (by the work of a sculptor) then there is actually a statue.

Prime matter. Although this understanding of “matter” and “form” can have some intuitive appeal, it leads to a complication. Even if matter is what potentially exists in some sense, apparently what exists as matter also seem to actually exist in another sense. For example, even if a lump of clay is *potentially* a statue, it is *actually* of the clay-type. Therefore it seems that anything that one can consider as the matter of something also has some form. Hence the further question of what *its* matter is. What is the matter of this matter? If for any identified matter we can further ask what its matter is, we face an infinite sequence of things as matters. To end this sequence, TH theorists posit “prime matter.”⁷⁷ Prime matter is considered “pure potentiality.”⁷⁸ There is no question of what's the matter of prime matter.

⁷⁶ “[J]ust as everything that is in potentiality can be called matter, so everything from which something has being, whether accidental or substantial being, can be called a form... And since form makes something actual, form is also called actuality.” (Aquinas, *On the Principles of Nature*, Ch.1) See also Shields & Pasnau 2016, pp. 29-31.

⁷⁷ “[S]ome matter has some form, for example, the bronze, which is matter in respect of the statue, but bronze itself is composed of matter and form; wherefore bronze is not called prime matter, for it has matter. But that matter which is thought of without any kind of form or privation as subject to all forms and privations is called *prime matter*, because there is no other matter before it...” (Aquinas, *On the Principles of Nature*, Ch. 2.)

⁷⁸ See Pasnau 2011, sect. 3.1 for issues regarding how to understand pure potentiality.

Prime matter as matter without form is supposedly the ultimate matter that could potentially assume any forms, and endure through any changes, generations, or destructions of individuals. For example, medieval Aristotelians believed that the material world consists of four elements (Air, Earth, Fire, and Water), but they believed that the elements are not fundamental. The four elements are thought to be made out of prime matter, which underlies all changes, including the changes from one element into another.⁷⁹ Thus Thomists refrain from attributing features to prime matter, take it as nothing beyond that which can actualize a thing.

Substantial form. Also according to commonsense, some changes merely involve the same object gaining or losing properties, while other changes involve the coming-into-being (generation) or -out-of-being (destruction) of an object. Hence it is thought that there should be different kinds of forms. Some forms are such that their coming or going determines the coming-into- or -out-of-being of objects. These forms are called “substantial forms”. Those other forms whose coming or going brings changes in the properties but not coming-into- or -out-of-being of objects are called “accidental forms”.⁸⁰

Substantial form has at least three functions:

- First, the substantial form of an object determines what type of thing it essentially is.

For example, humanity is the substantial form of Kant, as Kant is essentially human.

⁷⁹ Pasnau 2011, sect. 3.2, p. 43.

⁸⁰ For example, being a bachelor is an accidental form of Kant. See Aquinas, *On the Principles of Nature*, Ch.1 and Shields & Pasnau 2016, p. 33 for more on the contrast between substantial and accidental form.

- Second, having a substantial form is necessary to qualify something as an individual.

Thus Kant has a substantial form and therefore exists as an individual, but Kant's right hand attached to his body lacks substantial form and therefore is not an individual.⁸¹

Similarly, the presence or lacking of a substantial form can be used to explain why a certain collection of objects compose a real individual or only a heap. Human beings (and organisms in general), for instance, are truly individuals, while a pile of bricks is most likely not.⁸²

- Third, continuous bearing of a substantial form is what ensures the enduring existence of an individual. For example, insofar as Kant lives, his substantial form—the organizational principle that shapes some prime matter into this human individual known as Immanuel Kant—endures and maintains the existence and identity of Kant,

⁸¹ For the view that in a human body there is only one substantial form, see Aquinas, *Summa Theologiae*, Part I, Question 76, Article 4. See also Pasnau 2011, sect. 26.5, p. 624. According to Pasnau, scholastic philosophers had a wide range of views on part-whole relation, and so they would have different opinions on whether a hand of a human body is itself truly an individual. TH is particularly strong in this aspect, for Thomistic view denies rather than weakens the individual status of any part of a living organism. This denial of “substantial” existence is inherited by contemporary Thomists. For example, as Madden (2013b) puts it,

[T]he parts of a material substance are not *substantially present* within the composition. That is, [Fluffy the cat's] paw, tail, heart, carbon atoms, quarks, etc. are not present in her composition as *individual substances*... These parts of Fluffy are retrievably, potentially, virtually, or nominally present in Fluffy's composition. Strictly speaking, there are no carbon atoms in Fluffy *actually as material substances*, but she subsumes the powers of such substances, and we could generate actual individuals of this kind by dividing Fluffy, for example, cutting small enough parts from Fluffy might produce individually existing carbon atoms. (p. 668)

See also Oderberg (2005), in terms of “the unicity of substantial form” (p. 81).

⁸² Pasnau 2011, sect. 24.2.

even though Kant experiences various changes. Therefore, substantial form is used to explain how an individual comes into being, endures through changes, and eventually goes out of being.⁸³

To summarize, according to TH, the existence and features of a material individual shall be explained in terms of its prime matter and substantial form. Its substantial form provides its essential features and supports the enduring existence of the individual, separating the individual from other portions of the material world. Prime matter provides the fundamental matter to be informed and thus composes *material* individuals, which can be distinguished from same-type individuals by its distinctive portion of matter. A material individual is at least a compound of prime matter and substantial form.

4.3.2 Epistemic Inadequacy of Thomistic Hylomorphism

Metaphysicians can dispute the pros and cons of TH. I can fully believe that a metaphysician who supports TH can stick to his post. My concern with TH is something else, i.e., that TH is epistemically inadequate if treated as a research program to guide the

⁸³ “[W]hen the substantial form is introduced, something is said to come to be, without further qualification. But when an accidental form is introduced, we do not say that something comes to be, without qualification, but that something comes to be this; just as when a man becomes white, we do not say that he comes to be, absolutely speaking, but that he comes to be white. And to these two kinds of generation there correspond two kinds of corruption, namely corruption in an absolute sense, and corruption with qualification.” (Aquinas, *On the Principles of Nature*, Ch. 1.) See also Pasnau 2011, sect. 24.2; Shields & Pasnau 2016, pp. 34-35.

study of the natural world.⁸⁴ My worry about the epistemic inadequacy of TH is that its elements are unintelligible and unexplanatory. More specifically, prime matter is unintelligible, while substantial forms are unexplanatory.

As explained in section 3.1, in TH prime matter can be understood as the “ultimate matter”, in the sense that the being which counts as prime matter endures through all changes and has nothing as its matter. However, although we may understand that the logic of form and matter could lead to something like prime matter in TH, that which performs the role of TH prime matter is really unknowable. What distinguishes TH from other hylomorphism is that TH requires that prime matter be formless and thus able to assume any form. But since understanding something requires grasping its form, the formlessness of prime matter in TH means that whatever that might be prime matter is unintelligible, because it has no form that characterizes it.⁸⁵ Prime matter supposedly has no form at all by itself and cannot exist independently of any form. On the one hand, if any positive feature is attributed to prime matter, the feature seems to belong to the entity consisted of the prime matter, not to the prime matter itself. On the other hand, if no positive feature is attributed to prime matter, then there is no story about prime matter

⁸⁴ It should be noted that when Aquinas presented his metaphysics of the natural world, he might not intend TH to contribute to *further* exploration of the nature that a modern scientist would do, whereas RM seems more explicit in its ambition as a research program. See section 5, Footnote #103.

⁸⁵ According to Aquinas:

Now, since [any] definition and cognition is [obtained] by form, prime matter cannot be cognized or defined in itself, only by comparison, as when we say that prime matter is that which is to all forms and privations as bronze is to the form of the statue and to the lack of this form. (*On the Principles of Nature*, Ch. 2.)

that can be specified or tested. Therefore TH prime matter is unintelligible.⁸⁶ The worry is not new. Early modern critics of Aristotelianism such as Descartes have already complained about the unintelligibility of prime matter.⁸⁷

Whereas prime matter is unknowable, substantial forms in TH are unknown. Like in the case with prime matter, TH offers a metaphysical concept of substantial forms: the coming or going of substantial forms determines the existence of ordinary individual objects of specific kinds. However, as in the case of prime matter, in TH it is unknown what substantial forms are. Medieval believers in substantial forms were mostly skeptical about our ability to know substantial forms.⁸⁸ Early-modern critics of scholastic philosophy were also ready to remind people about philosophers' ignorance of substantial forms.

The fact that substantial forms are unknown contributes to another major epistemic failure of substantial forms, i.e., their limitation in explanatory power. To be

⁸⁶ According to Brower (2014), medieval scholastics seemed more ready to consider TH prime matter as “incoherent”, e.g., because TH prime matter is supposed to be “pure potentiality” and have no “actuality”, and to have no “actuality” is thought tantamount to not existing (Brower 2014, pp. 120-21.) My concern here is not about whether TH prime matter is real, but whether TH prime matter can be meaningfully characterized besides in terms of its metaphysical function.

⁸⁷ Descartes said the following against prime matter before introducing his own theory of matter,

[L]et us not also think that this matter is the ‘prime matter’ of the philosophers, which they have stripped so thoroughly of all its forms and qualities that nothing remains in it which can be clearly understood. (1664/1985, “The world or Treatise on Light”, Ch. 6, p. 91.)

Aquinas also acknowledged that we have little idea about what prime matter is. See Pasnau 2011, sect. 3.3, p. 48; sect. 7.2, pp. 119-124.

⁸⁸ Pasnau 2011, section 7.2, pp. 119-21; section 27.2, pp. 635-6.

precise, I will not argue that substantial forms are utterly unexplanatory, because according to TH, substantial forms explain the individuality of ordinary objects. It seems intuitive that there is some fact of the matter that a dog is an individual that persists over time and has always been the same dog despite its change from a puppy to an old dog. Therefore according to TH theorists, something (namely, a substantial form) must be doing the metaphysical job of making the ordinary individual, i.e., the same good old dog.⁸⁹ Nevertheless, substantial forms still seem unexplanatory of why something has certain features. This criticism is captured in the well-known ridicule that scholastic thinkers would appeal to “*virtus dormitiva*” (“dormitive virtue”, which literally means no more than *the property of inducing sleep*) to “explain” how opium can induce sleep. In general, an individual of type X is said to exist because of a substantial form of X, as for instance a human being is said to be constituted by the form *humanity*. To explain in TH why someone is human is to cite its *humanity*, which properly understood in terms of metaphysical function means no more than *that which makes an individual of the type human*. Therefore, if TH theorists cite the substantial form of X to explain why there is an individual of type X, then they are saying only that *something* makes an X-type individual, without explaining what that something is. Adding the fact that the things that

⁸⁹ Or so would hylomorphists argue for substantial forms; see Pasnau 2011, Section 24.2. It might be argued that there is no need for substantial forms as some robust metaphysical principles other than (maybe) some relations or configurations of the matter that composes the objects. However, contemporary debates in mereology have shown that such reductive accounts of ordinary objects cannot be easily secured. In fact, back to philosophical history, if Pasnau’s (2011) account is correct, then the early-modern attempts to abandon substantial forms gave rise to “metaphysical chaos” in the understanding of individual objects (Ch. 27-30; the phrase “metaphysical chaos” comes from p. 633.)

perform the metaphysical function of substantial forms are widely recognized as unknown, we should conclude that TH's substantial forms don't explain why something has a certain feature.⁹⁰

Is it possible to offer an informative and explanatory account with substantial forms? To provide more explanation beyond mentioning substantial forms as metaphysical functions, one should explain what substantial forms are. For example, the

⁹⁰ My concern here corresponds to what Pasnau has summarized as the criticism that substantial form is "superfluous", offering limited intellectual benefit (2011, sect. 24.5). This is very clear in Hobbes's criticism of Aristotelian philosophy:

Then for *physics*, that is, the knowledge of the subordinate and secondary causes of natural events; [the Schools] render none at all, but empty words. If you desire to know why some kind of bodies sink naturally downwards toward the earth, and others go naturally from it; the Schools will tell you out of Aristotle, that the bodies that sink downwards, are *heavy*; and that this heaviness is it that causes them to descend. But if you ask what they mean by *heaviness*, they will define it to be an endeavour to go to the centre of the earth. So that the cause why things sink downward, is an endeavour to be below: which is as much as to say, that bodies descend, or ascend, because they do. (1651/1962, p. 678; more examples in pp. 678-679)

More generally, in Descartes's term,

[Substantial forms] were introduced by philosophers solely to account for the proper actions of natural things, of which they were supposed to be the principles and bases... But no natural action at all can be explained by these substantial forms, since their defenders admit that they are occult and that they do not understand them themselves. If they say that some action proceeds from a substantial form, it is as if they said that it proceeds from something they do not understand; which explains nothing. (1642/1991, "Letter to Regius," pp. 208-209)

Also see Gassendi's (1624/1972) *Exercises against the Aristotelians*, Bk. II, Exercise VI.

For a scientific example of the limitation of finding substantial-form-like principles, consider the fact that biology operated mainly as a discipline of taxonomy in Linnaeus's style is limited (Gardner 1965). Taxonomy with Aristotelian method misses many important aspects of knowledge in biology, such as the operating of organisms, the life history of individual organisms, the natural history of collections of organisms, the relationship among organisms and environment, etc.

substantial form that individuates a dog may be accounted for in terms of the natures and activities of the dog parts. But, if the discussion in the previous paragraph is right, a more informative account of forms would require a more informative account of matter, in which matter should be attributed certain forms. For example, if the substantial form of a dog is the causal arrangement among the portions of the matter of the dog, these portions should have *specific* causal powers that make possible the *specific* arrangement of the dog. But the portions of matter should have forms in order to have causal powers, and so the matter with specific causal powers cannot be formless prime matter. Thus, to say anything informative about form, the matter itself should also be formed.^{91 92}

⁹¹ Advocates of substantial forms as causal arrangements may believe that the matter with causal powers should be understood, ultimately, as compounds of prime matter and some more basic forms. Then there is a question of what those more basic forms are. And the lesson still holds that if an informative account can be given of those forms in terms of the matter they arrange and the ways of arrangement, then formless prime matter still has no place in such an account of forms.

⁹² Indeed, according to Pasnau, the scholastic notion of substantial form has a “physical aspect” besides being a metaphysical notion of that which accounts for the individuality and essence of material objects. Understood in its physical aspect, substantial forms tend to be understood as efficient causes which allows explanations of the manifested features, and in fact the scholastic accounts of substantial forms evolved toward the emphasis of substantial forms as efficient causes, whose functions are prone to further analysis in causal terms (2011, sections 24.3-4, pp. 558-64). With a more causal notion of substantial forms, the idea of matter also tended to become corpuscularian. Thus scholastic ideas of “substantial forms” indeed had the tendency to become informative, but its trend of development led toward the reduction of substantial forms and toward “formed” matter. This trend deviates from TH, not only because it gradually deconstructed Aristotelian substantial forms into mechanical arrangements, but also because it abandoned prime matter in TH. Therefore, attempts to offer informative accounts of substantial forms in causal terms cannot use matter as in TH’s sense of prime matter.

4.4 Russellian Monism and Thomistic Hylomorphism

In this section I compare RM with TH. I will argue that (section 4.1) they have similar methodology, and that (section 4.2) consequently they have similar epistemic inadequacies/shortcomings.

4.4.1 Methodological Similarities

If my explanation of RM and TH is on the right track, then there is reason to think that their strategies are similar. This similarity I want to draw is more methodological and epistemological than metaphysical. Despite the fact that the categories of matter and form offers hylomorphists some room of maneuver analogous to RM's move in philosophy of

mind,⁹³ I don't think that RM pan(proto)psychist theories can be smoothly interpreted as variations of hylomorphism. Most RM theorists will probably protest if I argue that their metaphysics is Thomistic hylomorphism. So I will not argue that RM is a version of TH (nor that TH is a version of RM). However, I do want to argue that RM and TH have similarity in their methodology.

4.4.1.1 Establishing New Metaphysical Offices

Both RM and TH proceed through an expansion of metaphysical framework, in the sense that both argue that certain *metaphysical functions* should be realized by some categories of beings. Speaking figuratively, some new metaphysical “offices” are opened

⁹³ Like RM, TH can consider mind and body as belonging to two distinct metaphysical categories. Thomistic hylomorphists find a metaphysical “office” (i.e., substantial form), such that consciousness (and material individuality, intentionality, life, etc.) can be categorized as occupants of this metaphysical office or dependent on occupants of this office. Moreover, the more familiar physical beings studied through physical sciences are identified as belonging to another metaphysical office, i.e., the “matter” part of objects short of substantial form, such as the bits and parts of objects and their mechanics. So a metaphysical distinction between mind and body is achieved through a metaphysical division of labor.

On the other hand, despite the metaphysical distinction between mind and body, Thomistic hylomorphists could still offer some account of mental causation and a unified ontology. Since TH is friendly to the Aristotelian framework of causation, which retains “formal cause” and “efficient cause”, a Thomist can treat mental states as formal causes, in contrast with the correlating neurophysiological states which provide “efficient causation.” (Haldane 1998, pp. 272-6.) Therefore, TH seems to explain the puzzling existence of mental causation as somehow dependent upon, but distinct from and irreducible to physical causation. Furthermore, the mind/body difference in TH doesn't split the world in the way that Cartesian dualism does, because the metaphysical offices to which mind and body belong, i.e., form and matter, are interdependent: matter and form cannot exist independently. A form must actually exist by being embodied in some matter, whereas matter cannot exist without assuming some form. Therefore, a mind-body unity is maintained in TH.

in the metaphysics bureau to perform certain functions. These metaphysical offices provide some room for certain explanations in metaphysics.

As I have presented (section 3.1, this chapter), in the case of TH, certain notions about the nature of change is accepted, including the intuition that something should endure through changes, and that some changes merely involve some object gaining or losing properties whereas others involve the coming-into- or -out-of-being of an object. With the notion of change as requires something that goes through a change, it is thought that something should perform the role of *that which endures through a change*, hence the metaphysical function of “matter”, and consequently the metaphysical function of “form” as something that comes or goes during changes. With the idea that some changes don’t involve objects coming-into- or -out-of-being while others do, it is thought that the metaphysical function of “form” should be divided further: some forms are involved in the coming-into- or -out-of-being of objects while others don’t. Hence the metaphysical function of “substantial forms” as what’s responsible for the changes concerning coming-into- or -out-of-being of objects and the metaphysical function of “accidental form” as what’s responsible for changes but not coming-into- or -out-of-being of objects. With substantial forms, there is also the need to realize the metaphysical function of “prime matter” which, together with substantial forms, takes part in the generation/corruption of individual objects. By establishing the notions of prime matter and substantial form as metaphysical functions, with the assumption that something in the reality must be

performing these functions, TH theorists have offered an account of how change in our commonsense notion is possible.⁹⁴

RM theorists also proceed by appealing to metaphysical function, although because RM theorists aim at accommodating consciousness, they need not only to establish the need of metaphysical function(s) but also make the case that consciousness or the ultimate ground of consciousness might be performing that metaphysical function. The strategy in RM is to reconstruct the metaphysical framework with the distinction between intrinsic/extrinsic or categorical/dispositional properties. Crucially, it is thought that the world cannot be dispositions all the way down: there must be something which is the fundamental intrinsic nature of the dispositional properties, even if physical sciences cannot account for it. Therefore, something must be playing the metaphysical function of fundamental intrinsic nature (and it is named “the inscrutable”). With this metaphysical

⁹⁴ Indeed, in the eyes of contemporary hylomorphists, the territory of contemporary metaphysics is too barren. They tend to seek more metaphysical offices in matter and form before launching into a specific metaphysical issue. For example, Madden (2013a) introduces the Thomistic hylomorphic “philosophy of nature” before presenting his Thomistic theory of mind. Same attitude is found in non-Thomistic hylomorphists. For example, Jaworski has made the following comment about John Searle, who calls for a radical change to the concept of mind/body or mental/physical but hasn’t achieved such a change:

There might be many reasons why Searle doesn’t abandon the traditional mind-body categories when articulating his theory, but perhaps the most obvious is that he doesn’t develop a metaphysical framework that would enable him to formulate his theory in a different way. He doesn’t dig deeply enough into the metaphysical soil to uproot the mental-physical dichotomy that he and other philosophers take to be problematic. Hylomorphism, by contrast, does. The hylomorphic framework does not imply a commitment to categorizing powers or manifestations, or the systems or subsystems that have them, as mental or nonmental, physical or nonphysical. According to hylomorphists, we can get on with our empirical investigation of powers, parts, and manifestations without ever employing those categories. (2016, p. 175)

framework, RM theorists thus suggests that there is room in their metaphysics to accommodate consciousness, which seems unaccountable by physical sciences. Thus RM theorists rely on the notion of categorical base v. dispositions to expand their metaphysics, and once they reach the metaphysical office of *fundamental* categorical base, they may open a position for (proto)phenomenal properties, which they intend to accommodate.

4.4.1.2 Fundamental Stuff and Ordinary Individuality

Aside from the common strategy of TH and RM (i.e., the maneuver in metaphysical functions), there are two more points of similarity in the metaphysical concerns of TH and RM. **(A)** First, both TH and RM are centered around the idea of a fundamental “stuff” in some sense. **(B)** Second, both have an interest in establishing the individuality of certain familiar things, which pushes both to consider principles that make such familiar individuals. These two similarities reflect in the types of metaphysical functions they posit respectively.

A. Both TH and RM are centered around the idea of a fundamental “stuff” in some sense. In the case of RM, the strategy of opening a metaphysical function for (proto)phenomenal properties suggests the posit of the *fundamental* intrinsic base that realizes all causal roles. This is because ordinary intrinsic properties (such as the roundness of a tennis ball) may be fully described by physical sciences, but *fundamental* intrinsic nature is a domain about which physical sciences seem to be silent.

Consequently the metaphysical office of fundamental intrinsic nature seems open for

(proto)phenomenal properties, as there is little constraint from physical sciences against speculating that what's in that domain is (proto)phenomenal.

As for TH, Thomistic Hylomorphism has a distinctive concern about the ultimate substratum that endures through all changes. Crucially, TH denies that the ultimate substratum has forms, like the elements of Earth, Air, Water, and Fire in medieval worldview. These elements in empirical accounts are thought to be changeable, such that one element may change into another. Given the hylomorphic view that some matter should endure through a change, some fundamental matter devoid of the forms of Earth, Air, Water, and Fire must underlie the change among the elements (section 3.1, this chapter). Thus TH needs the notion of *prime matter*, which is the *fundamental* matter that is itself formless but serve as the material base of all formed things.

B. Besides their interest in fundamental “stuff”, both RM and TH have an interest in establishing the individuality of certain familiar things. This concern is clearer in TH's notion of substantial forms. The notion of change accepted by TH theorists presupposes the commonsense view that ordinary objects such as humans and dogs exist, and that these objects undergo processes of generation, persistence, and destruction. Hence the TH metaphysical function of substantial forms that account for the existence of ordinary objects, to be performed by that which constitutes an ordinary individual (such as a dog) out of a portion of matter.

As for RM, RM theorists also have an interest in staying consistent with an ordinary notion of individuals. Recall (section 2.2, this chapter) that combination principles are what explain the generation of macro-level consciousness from the micro-

level entities which have the inscrutable as their fundamental intrinsic nature. As a metaphysical function, combination principles in RM are similar to substantial forms in TH, as far as an account of an ordinary individual is concerned. Whereas substantial forms in TH are supposed to make (from a portion of prime matter) a material individual that is recognizable in commonsense (and also separate this individual from other portion of prime matter), Combination principles in RM are supposed to make (from a set of micro-level entities with the inscrutable) a conscious individual, a conscious perspective that a human organism normally has.⁹⁵

Therefore, both RM and TH proceed by positing certain metaphysical functions (defined and organized in some desired way.) The general idea is that a metaphysical framework of sufficient complexity can accommodate the complexity of the phenomena, as in change or phenomenal consciousness. In developing their metaphysical framework, both RM and TH develop certain metaphysical functions of the fundamental “stuff” in some sense (the inscrutable in RM and prime matter in TH) and some other metaphysical functions for the making of some ordinary individuals (combination principles in RM and substantial forms in TH). These points of similarity in the methodology and concerns of RM and TH suggests that their epistemic challenges may be comparable, and therefore the epistemic track record of one (TH) may suggest what the epistemic prospect of the

⁹⁵ This paper is primarily concerned with the comparison between RM and TH (and what we may learn from their similarities). Concerning non-Thomistic, structure hylomorphism, the case of similarity with RM is weaker but still significant. Since non-Thomistic hylomorphists do not accept prime matter, it is unclear if they have an analog to the inscrutable in RM. But since they often preserve forms to make individual objects, non-Thomistic hylomorphists still have an analog to combination principles in RM.

other (RM) would be like. I turn to the comparison of the epistemic inadequacy of RM and TH in section 4.2.

4.4.2 Epistemic Inadequacy of RM and TH, Compared

If both RM and TH face epistemic challenges, given the similarity between them, it's worth comparing their epistemic challenges and find similarities. In particular, we can compare (4.4.2.1) the epistemic challenge of the inscrutable in RM and prime matter in TH, and (4.4.2.2) the epistemic challenge of the combination principles in RM and substantial forms in TH.

4.4.2.1 The Inscrutable and Prime Matter

I have argued (section 3.2, this chapter) that prime matter in TH cannot be characterized. I have also presented an argument of Ramseyan humility (section 2.1, this chapter) to show that the inscrutable in RM is unknowable. If the argument of Ramseyan humility is correct, then there could be no empirical support for/against any specific claims about the inscrutable/quiddities, e.g., of whether the inscrutable is phenomenally reddish or phenomenally greenish, or whether the inscrutable is phenomenal at all. Therefore, of both prime matter in TH and the inscrutable in RM, nothing specific can be said.

Or, there is something seemingly specific to be said about each of them. Both are understood in terms of metaphysical functions (section 4.1). Prime matter as a metaphysical function in TH is that which has no form and thus undergoes all changes.

The inscrutable in RM as a metaphysical function is the intrinsic nature which underlies all causal roles and thus itself has no causal profile. However, these metaphysical functions seem to be all that we could understand about prime matter and the inscrutable. We cannot specify what performs the metaphysical functions of prime matter or the inscrutable. Prime matter is denied characterization by definition. The inscrutable is also unable to be specified, due to Ramseyan humility.⁹⁶ Moreover, it seems that the epistemic challenge with both the inscrutable and prime matter is due to the fact that the metaphysical functions involved are *fundamental* stuff, which are *defined* in a way that

⁹⁶ Indeed, protophenomenal inscrutable in RM panprotopsyism is similar to prime matter in TH both in their unintelligibility and in the metaphysical functions that philosophers may want them to play. Consider what Madden has said (in a context against panprotopsyism), and note how close this may sound to panprotopsyism:

[T]he hylomorphist does not claim that fundamental physical particles are conscious, but only that prime matter has the potency for the power of sensation that can be actualized under certain conditions. Of course prime matter is part of the composition of fundamental particles and conscious organisms alike. That is not to say that the former has a proto-version of the power of sensation, but only that it is composed of something that has the potential for such a power when it is coupled with the right sort of substantial form. (2013a, p. 260)

Everyone in the debate about consciousness acknowledges that conscious creatures and their capacities are somehow based on the microphysical components. To this extent, everyone in this debate acknowledge some kind of “potential” for consciousness in the fundamental nature of matter. RM theorists acknowledge this by building the consciousness-grounding feature into inscrutable. A Thomist hylomorphist can acknowledge the same by saying that “prime matter has the potency for the power of sensation that can be actualized under certain conditions”, as Madden has claimed. Thus it seems that both prime matter in TH and the inscrutable in RM panprotopsyism are not consciousness in actuality but provide the potentiality.

The similarity between the inscrutable and prime matter may be stronger, for not only are they supposed to provide the potential of consciousness, but they are also supposed to participate in the metaphysical ground of all material things. In RM, the inscrutable not only ground consciousness, but also serve as the intrinsic base of physical beings. In TH, prime matter is conspicuously “pure potentiality” and has a place in all material things.

deprives the possibility of knowledge of them. The fundamental status of prime matter requires that it's formless, making it unintelligible. Likewise, in RM the fundamental status of the inscrutable requires that it's disposition-less, making the specification of the inscrutable impossible.

Thus, in their search of metaphysically fundamental stuff, both TH and RM go beyond what we can understand, with the consequence that their posited fundamental stuff become unintelligible or unknowable.

In defense of RM, one may argue that TH's epistemic situation is worse: prime matter is unintelligible, whereas the inscrutable is unknowable but intelligible. In RM it makes sense to characterize the inscrutable, despite the lack of empirical test, whereas in TH it doesn't make sense to characterize prime matter. Thus, in the case of RM panprotopsyism, at least it's not wrong to think that the inscrutable is *something*, of which we may never know, but is *something* nevertheless. Whereas, in the case of RM panpsychism, it also makes sense to think of the inscrutable as *something*; furthermore, in RM panpsychism it seems that the inscrutable is not only intelligible but also slightly more knowable than prime matter, because the inscrutable in this case is *something phenomenal*, and we seem to know intuitively what is for something to be phenomenal. We know many instances of phenomenal experience, from which we may have some intuitive grasp of what *being phenomenal* is versus what is *not* to be phenomenal (e.g., zombies).

In response, I want to make two points.

First, about the idea that the inscrutable in RM panpsychism is better known than prime matter, I must first point out that RM panpsychism is still a hypothesis. Due to Ramseyan humility, we don't really know that the inscrutable is phenomenal. The reasons we have for RM panpsychism are philosophical considerations, e.g., that macro consciousness should be explained by micro components with consciousness, and we don't yet have decisive philosophical reasons to prefer panpsychism to panprotopsyism. Also, I have argued (section 2.1, this chapter) that RM panpsychism cannot support any further speculations on *what phenomenal property* the inscrutable is specifically. In response, RM panpsychists may argue that even if we don't know *what phenomenal property* the inscrutable is, we do know what *being phenomenal* (period) is, and they may take this to show that they are in a better epistemic position than TH. However, if this minimal phenomenality of the inscrutable is sufficient for us to think that we have some understanding of the positive feature of the inscrutable, then it seems that we are also justified to think that we have some understanding of the positive feature of prime matter, i.e., its "minimal materiality." Prime matter itself is formless, but it is at

least material.⁹⁷ We have some intuitive grasp of material objects, from which we have a grasp of what is to be material at all. And indeed, in medieval concepts, people even “knew” what is to be material and what is not: human beings and many other animals are material (and thus are partly composed of prime matter) but angels are not,⁹⁸ just as it seems plausible to many contemporary thinkers that actual human beings have phenomenal consciousness but zombies don’t. If the minimal materiality of TH prime matter couldn’t be seriously taken to make prime matter sufficiently knowable to earlier scholars, it’s questionable if the minimal phenomenality of the inscrutable in RM

⁹⁷ Saying that prime matter is the *prime* matter that cannot be predicated of any form means that: prime matter cannot be characterized, in the sense that you should never say “prime matter is X” where X stands for a form that characterizes prime matter itself, in the way that “biped” characterizes human beings in “human beings are biped.” That prime matter cannot be predicated of forms that offer an informative account of something as a matter-form compound doesn’t mean that prime matter cannot be spoken of in ways that doesn’t predicate it with forms, e.g., “prime matter is matter.” This statement is true without attributing any form (e.g., hot, wet, cold) to prime matter.

Aquinas has other ways of talking about prime matter without attributing forms. For example: “[P]rime matter, as well as form, is not generated (or corrupted)”, because “every generation proceeds to something from something. That *from which* generation proceeds is matter, and that *to which* generation proceeds is form. Therefore, if either matter or form were generated, then matter would have matter and form would have form, and so on, *in infinitum*.” Also: “[M]atter is said to be numerically one in all things”, in the sense that “it lacks those dispositions which make things numerically different”. (“On the Principles of Nature”, Ch. 2.)

⁹⁸ See Aquinas, *Summa Theologiae*, Part I, Question 50: “The Substance of Angels Considered Absolutely”, Article 2: “Is an angel composed of matter and form?”, translated by Alfred J. Freddoso: <https://www3.nd.edu/~afreddos/summa-translation/Part%201/st1-ques50.pdf>.

panpsychism can make the inscrutable sufficiently knowable to contemporary philosophers.⁹⁹

Second, more importantly, for the purpose of developing a research program that has a hope of moving forward on the explanation of phenomena, a theory being unintelligible and a theory being unknowable-but-intelligible are not practically different. If a notion is unintelligible in the sense that characterizing it is logically prohibited (as prime matter in TH), then this notion cannot be figured into explanations. If a notion is allegedly intelligible because characterizing it is logically acceptable but no specific characterization of it can be described, motivated, or tested (as the inscrutable in RM), then this notion also cannot figure into explanations. Therefore, insofar as neither prime matter nor the inscrutable can be specified, regardless of whether this is due to the

⁹⁹ *OBJECTION*. One may argue that perhaps *we* humans don't know the inscrutable, but a much more powerful being might know. Isn't it nice if we can share the metaphysics of that more powerful being, even if we ourselves cannot fill in the details? *REPLY*. But any advocate of any speculative metaphysics can appeal to this potential benefit, conditioned upon the truth of their metaphysics. TH theorists also tend to attribute whatever possible knowledge about prime matter to God. None of such assurance of a more powerful knowing being would help *us*. Furthermore, given Ramseyan humility, it is unclear what powerful being can know the inscrutable, unless it has more ways of knowing than forming theories and gathering and interpreting evidence (e.g., by "revelation"). Therefore, if RM theorists appeal to an omniscient being that knows the inscrutable, they are merely expressing a metaphysical faith.

meaninglessness or the in-principle unknowability of the specification, the epistemic challenges in both cases suffice to paralyze the theories that use these notions.¹⁰⁰

4.4.2.2 Combination Principle and Substantial Form

Both RM and TH posit some metaphysical functions of fundamental stuff, in the form of the inscrutable and prime matter respectively. But we are in principle unable to specify the thing that performs such metaphysical functions of fundamental stuff. Hence the epistemic challenges with the inscrutable and prime matter. In this section, I will compare the epistemic challenges with the combination principles in RM and substantial forms in TH, arguing that their epistemic challenges come from the difficulty in specifying that which perform the metaphysical functions of making ordinary individuals in RM and TH.

Recall that combination principles are what explain the generation of macro-level consciousness from the micro-level entities which have the inscrutable as their intrinsic nature. The status of combination principles with regard to the inscrutable in RM is

¹⁰⁰ Notice that in both TH and RM, the move toward the fundamental stuff makes explanation impossible. In hylomorphism, if one explains things with “elements” of Earth, Air, Fire, and Water, it seems that one can still operate some explanation, although the explanatory success is highly questionable. But once one steps further toward the fundamental level, of prime matter, then one reaches a level which cannot be characterized. Similarly, in pan(proto)psychism, people may also operate some crude explanations if they stick to understandable but “small and simple” phenomenal qualities (e.g., explaining color consciousness by claiming that ultimates have extremely simple color qualities), even though such explanations would seem simple-minded, even stupid. But once we step further to the fundamental level, of the inscrutable, then it seems that nothing specific could be said. I will also mention this observation of the epistemic difficulty with the fundamental level in the next section (section 4.2.2, this chapter).

analogous to that of substantial forms with regard to prime matter in TH, in the sense that the combination principles perform the function of making conscious individuals that we recognize in commonsense, just as substantial forms perform the function of making commonsense material individuals.

I have argued (section 3.2, this chapter) that substantial forms in TH are unspecified and lack explanatory power of why something is of a certain type. I have also argued (section 2.2, this chapter) that combination principles in RM are unspecified and unexplanatory. Therefore, the epistemic challenges to combination principles in RM and substantial forms in TH are similar. Moreover, we can compare the sources of the epistemic challenges to combination principles and substantial forms, which are also comparable to the sources of the epistemic challenges in the case of the inscrutable and prime matter. The common source of the challenges, generally speaking, is that we could only understand these categories in terms of unspecified metaphysical functions. Previously I have argued that there could be no specific account of combination principles in RM (section 2.2, this chapter). Consequently, for anything that functions as a combination principle, we could only understand it abstractly in terms of its metaphysical function, as

that (unspecified) arrangement of the micro entities with the inscrutable which results in the macro-level properties in question (including macro consciousness)

Now consider the fact that we also know what a substantial form is supposed to be as a metaphysical function, as

that (unspecified) principle which arranges prime matter so as to result in the macro-level individual in question

Therefore, in both cases of RM and TH, we have no knowledge of the *things* that perform the metaphysical functions of combination principles or substantial forms, but only about the metaphysical functions themselves. Therefore we don't know more about the combination principles in RM than about the substantial forms in TH. Claiming that something is performing the function of combination principles that make macro-level conscious individuals given the inscrutable as fundamental intrinsic nature is as uninformative as claiming that something function as the substantial forms that arrange prime matter into ordinary material individuals.

Notice that the epistemic challenge with both combination principles in RM and substantial forms in TH is the unknowability of the principles that arrange the fundamental stuff in each of the theories. It seems that the unknowability of the organizing principles of ordinary individuals (macro consciousness or commonsense material individuals) is related to the lack of understanding of the fundamental stuff to be arranged. Previously I have argued that lack of understanding of the "units" hampers the understanding of their arrangement (section 2.2, this chapter). Similar idea might have already been recognized by Aristotle. In many of his explanations of natural phenomena, instead of appealing to prime matter, Aristotle used Empedocles's four-element theory and proceeded in terms of the combinations of Hot, Cold, Dry, and Wet manifested in

mixtures of Earth, Air, Fire, and Water.¹⁰¹ This move seems to admit that the idea of prime matter, though metaphysically useful for some purpose, hardly figures into explanations of various phenomena. If Aristotle only had featureless prime matter and never considered the four-element theory, then due to the lack of characterization of prime matter, he would not have formulated any remotely illuminating explanations of natural phenomena that attracted the attention of his medieval followers. Corpuscularian theorists who took elemental matter to be particles also assumed certain basic properties of the matter, which allowed them to formulate hypotheses in terms of the combinations of matter, although different from Aristotle's.

The lesson against the inscrutable and combination principles in RM, by analogy, is that the unknowable inscrutable is unlikely to leave a promising epistemic prospect for RM. With prime matter came only unaccountable forms but no specific combinations. An account of the combination of elemental items is possible only after retreating from prime matter to characterized, "formed" items. Similarly, staying with the unknowable inscrutable as fundamental intrinsic nature will not result in any specified combination principles.

Thus there is a reason why the combination problem of panpsychism is difficult. Its difficulty is of the same sort that Thomistic Hylomorphists would face if they attempt to specify the arrangement of prime matter. Unless RM provides some positive characterization of the inscrutable beyond its metaphysical function as the fundamental

¹⁰¹ For Aristotle's praise of Empedocles's theory of elements, see *Physics*, Bk. I, §4, 188a12-188a18; *ibid.*, §6, 189a12-189a19.

intrinsic nature, we can hardly count on RM to specify the combination principles. If the prospect of specifying the combination principles is dim, then there is little future for the combination principles of RM to be better known and more explanatory than substantial forms.

4.5 Concluding Remarks

I have argued that RM faces similar epistemic challenges to TH's. Both TH and RM have some fundamental metaphysical principles which cannot be specified. In TH, prime matter is unintelligible, whereas in RM the inscrutable is unknowable (even though it might be intelligible). Furthermore, neither TH nor RM may offer adequate explanation of the macro-level, commonsense individuals that they are supposed to address. TH appeals to substantial forms, which are unspecified and therefore explanatorily limited. RM might be held to contribute to an account of the combination principles of the micro-level entities that have the inscrutable. However, given the unknowability of the inscrutable and the interdependence between knowledge of the to-be-combined items and knowledge of how they combine, the combination principles are also unknowable. The combination problem is a symptom of such ignorance of both *the elements* and *their arrangements* in the domain of phenomenal consciousness. In both TH and RM, since the fundamental principles of prime matter or the inscrutable cannot be specified, there is little hope that elaborating the metaphysics around these notions can guide theorists toward any specific accounts.

What is wrong with the unknowable inscrutable and the combination principles?

A supporter of RM may insist that there are metaphysical functions of the inscrutable and the combination principles, and therefore some things must be performing these functions, even if they are unknowable to us. The presence of that metaphysical function implies that we know at least that something exists at that metaphysical office.

In response, I acknowledge that if RM is just a metaphysical proposal with no ambition of guiding further discovery, then fine. However, RM seems to arouse a higher ambition (section 2.3, this chapter). RM seems to promise a research program into a deeper nature of the world. In that case, RM's epistemic challenges and its similarities to TH should be relevant. By comparing the metaphysical methodology and epistemic limitations of RM and TH, to the extent that a contemporary RM theorist would agree with my account of TH, I hope to make RM theorists more self-aware of the nature, the limitation, and the (lack of) prospect of their approach.

Consider what happened to TH. In TH, it is thought that there are some metaphysical functions such as prime matter and substantial forms, which are unknown; but something must be performing those functions, whatever it is. The unknowability itself was thought by many to be a problem. Hence a tendency among scholars to say more of prime matter (and substantial forms) to make it possibly better knowable. Furthermore, attempts to reconsider matter as formed and characterizable proceeded without the terms of TH. That was what Descartes and many early-modern philosophers

did. Therefore, TH was not, strictly speaking, shown to be false,¹⁰² but was abandoned for its lack of epistemic guidance. Now, if my comparison between RM and TH is apt, we have a reason to think that RM's condition is similar. According to RM, there are some metaphysical functions, especially the inscrutable, which is unknown, but supposedly something must be performing that metaphysical function, whatever it is. But, like TH, RM offers little epistemic guidance. RM's metaphysics implies that if any intellectually fruitful theories about consciousness may develop, it can proceed without RM's term of the inscrutable. Perhaps a RM theorist, especially if she is a panpsychist, would tend to say or hypothesize more about what the inscrutable is, e.g., that the inscrutable is phenomenally reddish or so. Perhaps if more can be hypothesized about what phenomenal property the inscrutable is, this *qualified* rather than *minimal* phenomenality (like the "formed" rather than "formless" prime matter) may offer a bit more epistemic guidance, but such an account of "phenomenal atoms" could proceed without using RM terms (such as the notion of the inscrutable). Thus, a theory of consciousness or even a theory of panpsychism could develop without real use of RM's ideas, except for the purpose of satisfying the metaphysical commitment to the fundamental intrinsic nature

¹⁰² Although the idea of substantial form was attacked forcefully by John Locke's argument from changelings in *An Essay Concerning Human Understanding*.

(that supposedly grounds consciousness.) Therefore, perhaps like TH, RM won't be shown as false but would be abandoned for its epistemic limitations.¹⁰³

There is no doubt that we should take historical ideas of philosophy, including TH, seriously. An optimistic view is that RM is on the right track, and that the similarity between RM and TH suggests that there is some truth in TH that has been forgotten or under-appreciated. However, the value of historical ideas doesn't have to be truth. A pessimistic view is that contemporary RM theorists may be renewing or reliving a track of intellectual bewilderment. It is a glorious pursuit, but we are bound to fall into unknowable items on this route. I lean toward pessimism.

According to Aristotle's report, Democritus hypothesized that animals are composed partly by round atoms, and that the mobility of animals is due to the mobility of the round atoms.¹⁰⁴ This account sounds naive today, but at least there is an attempt to specify the feature of atoms to make an explanation in terms of atoms possible. Later in history, corpuscularian ideas gradually gain prominence, because specific accounts of the nature of particles *and* their combinations have been developed and attested in tandem. If RM could not specify the inscrutable, then its performance will remain worse than that of

¹⁰³ In fact, RM might be at a worse position than TH, since RM may be more ambitious than TH. TH theorists acknowledged their ignorance, but had little ambition (e.g., in Aquinas) to go for new discoveries. In contrast, contemporary RM theorists seem to have the ambition of a new domain of the science of consciousness. Therefore, if my comparison between TH and RM is right, then RM and TH have similar epistemic challenges but RM has more intellectual ambition than TH. RM theorists should realize that RM's limited epistemic prospect really doesn't square with their intellectual ambition (whereas the ambition of TH philosophers didn't go far beyond TH's epistemic prospect).

¹⁰⁴ *On the Soul*, Bk. I, §2, 403b29-404a16.

Democritus's atomic theory. I will refrain from further speculation about how RM compares with more recent theories.

CHAPTER 5

CAN WE STUDY THE MICRO-LEVEL FACTS IN PAN(PROTO)PSYCHISM?

In this chapter, I will raise more epistemic concerns about pan(proto)psychism, without assuming that pan(proto)psychism must follow the metaphysics of Russellian monism. In Ch.4, I have argued that RM's epistemic prospects are limited. Thus, given that contemporary pan(proto)psychism often assumes the metaphysics of RM, the epistemic prospects of most contemporary proposals of pan(proto)psychism will be limited. However, even without the assumption that pan(proto)psychism is committed to RM, I think there is reason to be concerned that pan(proto)psychism will have little to contribute to our understanding of consciousness beyond metaphysical sketches. It is my objective in this chapter to explain this further reason to think that the epistemic prospects of pan(proto)psychism are limited.

My argument in this chapter can be summarized as follows:

Premise 1: A fundamental theory should have certain principles to be informative.

Premise 2: Pan(proto)psychists can hardly offer these principles in their account.

Therefore, pan(proto)psychism can hardly become an informative fundamental theory.

Pan(proto)psychism has been marketed as a theory of fundamental reality (Goff 2017b). Thus, to evaluate pan(proto)psychism as a fundamental theory, I need to explain

what an informative fundamental theory should look like. Hence the need to argue for Premise 1. Since physics has so far offered a decent exemplar of a fundamental theory, I will use physics as a model to illustrate what a fundamental theory should have in order to be informative.

After figuring out what an informative fundamental theory should look like, I will ask if pan(proto)psychists can offer such principles that make it informative. In particular, given the varieties of phenomenal properties that we apparently experience, we need to ask if pan(proto)psychism can offer the principles that work at the fundamental level (or close to the fundamental level) and explain the varieties of phenomenal properties. We know that physics can offer particles and their properties and relations, based on which we can to some extent explain the varieties of chemical elements and compounds that we can see, smell, or get poisoned by, and the operations of many large molecules that make life possible. Similarly, we may expect a sort of “mental chemistry”¹⁰⁵ to come out of pan(proto)psychism, in which the fundamental (or close to fundamental) facts in pan(proto)psychism can explain the varieties of phenomenal properties in everyday experience.

Current literature in pan(proto)psychism has offered little guidance on this question. Furthermore, there is reason to think that pan(proto)psychism cannot offer informative clues. The challenge here is similar to the “palette problem” of pan(proto)psychism,¹⁰⁶ which is a special case of this more general challenge of deriving

¹⁰⁵ To borrow the title of Coleman (2012).

¹⁰⁶ For an introduction to the palette problem, see Ch. 1 section 3.2, this dissertation.

the phenomena from the fundamentals. However, many pan(proto)psychists seem satisfied *just* with showing the *metaphysical* possibility of having enough *number* of the fundamental facts/categories in their accounts, with little to no attempt to characterize such micro-level facts/categories. I believe this is far from sufficient.

Like in Ch. 4, the reason I put forward in this chapter against pan(proto)psychism is not exactly the same as the combination problem that many philosophers have been focusing on. Most philosophers seem to focus on the *metaphysical possibility* of combination. Thus their accounts may address questions such as whether micro-subjects are the sort of entities that can combine, whether there are enough basic phenomenal qualities to make the varieties of qualities in our experience, etc. I recognize that pan(proto)psychists must worry about the metaphysical possibility. But I think we should also recognize the ambition of the pan(proto)psychists that pan(proto)psychism is to ground a *science* of consciousness. For this purpose, I think we should examine the epistemological prospect of developing a study along the lines of pan(proto)psychism, not only to accommodate the possibility of the combination of consciousness, but also to informatively characterize the combination (if combination of consciousness is possible.) My worry is that such informative characterization is unattainable in pan(proto)psychism.

5.1 What is a fundamental theory

By a “fundamental theory”, I mean a theory that explicitly take its subject matter as the fundamental reality, including fundamental entities, their properties, and relations. Every theory is about something, and every (concrete) thing is ultimately made up of

fundamental things, if fundamental things exist. However, only fundamental theories are explicitly about the fundamental nature of things, or at least about the nature of something sufficiently close to the level of the fundamental entity/entities of the universe. Thus, a fundamental theory must incorporate a fundamental ontology as a component.

I want to evaluate pan(proto)psychism as a *fundamental* theory, because pan(proto)psychism is supposedly a metaphysical approach to the fundamental reality (Ch. 1, section 2). In fact, supporters of pan(proto)psychism have little choice but to work at the fundamental level, because they reject the possibility that consciousness is derived from something wholly non-conscious, and consequently they have to include consciousness of some sort among the fundamental reality. Therefore, since pan(proto)psychism is supposed to offer a fundamental theory, to determine its epistemic value, we should ask what a fundamental theory should look like in order to be epistemically valuable.

5.1.1 What a fundamental theory should look like: the case of atomic physics

In explaining what a fundamental theory should look like, I will explain the basic ideas of atomic physics as an example. The picture I present will basically follow a Newtonian view of physics, without relativistic spacetime and quantum mechanics. Although this picture is far from being true in the strict sense, it is sufficient to demonstrate what a fundamental theory should look like. Its extent of precision and applicability, though imperfect, is good enough to qualify it as a fruitful theory of nature.

I will first briefly explain its basic principles, and then use several examples to demonstrate how its principles allow it to be fruitful.¹⁰⁷

Nowadays, for good reasons, most educated people believe that the world is made of atoms. (Although atoms themselves are made of sub-atomic particles.) The kinds of atoms we know of (such as hydrogen, carbon, oxygen) and their properties have been summarized well by the Periodic table of elements. Atoms have mass, which differs according to their kinds. Atoms could also manifest electric properties, because they can gain or lose a sort of much smaller particles that carries negative charges (namely, electrons.) Atoms also move in space. Atoms are also in myriad kinds of relations. In the first place, atoms in space are in spatial relations. At any moment, between two atoms, there is some magnitude of spatial distance. Also, given their spatial relations and due to their properties such as mass, charge, and motion, atoms can exert force upon each other, in the sense that they can influence the motion of others.

Atoms, with their properties and relations, have proven fruitful in explaining the world. Given these relations among atoms, they can bind into various relative stable structures, which give rise to larger entities such as molecules, dusts, cells, stones, organisms, and so forth. Thus we obtain the familiar things in the world we live in. Also, under certain conditions, stable structures of atoms can dissolve. And thus familiar things in the world can be destroyed. There is also no shortage of examples to illustrate the success of the atomic worldview. I will mention only four.

¹⁰⁷ My account of atoms and their work is partly inspired by Feynman et al. 2013, Ch.1, Ch.3.

- (E.g. 1) Why could fire destroy things? Because fire drastically increases heat, which is essentially to accelerate the motion of small particles that compose a thing. Once the particles become too active, their relations cannot bind them into the stable structure which they are in under normal conditions. Thus the structure—the thing—collapses in fire.
- (E.g. 2) Why does salt vanish in water, but come back if we reduce the water by boiling? Because salt (Sodium Chloride) consists of two sorts of atoms which can depart from each other in water. Thus water provides a space for the atoms in salt to “swim away.” Thus salt grains vanish in water, as sodium and chlorine particles of the salt move away from the salt grains into the body of water, until most of them are moving freely in the water rather than hugging together into a solid grain. However, if we reduce the water by boiling, then we accelerates the small particles of water (which consist of hydrogen and oxygen atoms), which then leave the body of water, leaving a smaller body of water. As the water shrinks, there is less room of movement for the sodium and chlorine particles of the salt that are trapped in the water. Thus the sodium and chlorine particles are more likely to come back together, and so the salt is back.
- (E.g. 3) Why do we have to breathe, otherwise we die? Because most of our bodily function requires constant energy supply to most of our components (cells), and the energy supply in these components is in the form of an energy pack (namely, ATP, made of 10 carbon atoms, 16 hydrogen atoms, 5 nitrogen atoms, 13 oxygen atoms, and 3 phosphorus atoms), and the production of most of these energy packs requires oxygen

(through a series of events with multiple enzymes, which are relatively big compounds of atoms.) To take in oxygen, human bodies must take in the air. So, we have to breathe.

- (E.g. 4) Why do organisms reproduce themselves? Why do dogs give birth to little dogs but not sheep? Because what organisms become is largely determined by a sort of atomic compounds, known as DNA. They are constructed in such a way that they record information and can be copied. Given the information from DNA, atoms and compounds of atoms can be assembled to produce a whole organism that is about the same as the organism(s) from which the information came (aka. parent(s).) Also, because DNA can be copied, this newly produced organism will retain roughly the same information, and therefore its offsprings will be assembled from atoms and compounds of atoms based on roughly the same blueprint. This is why organisms reproduce themselves.

5.1.2 Elements of a fundamental theory

The case of atomic physics suffices to show that a theory of micro-level composition should articulate at least two sets of elements: (1) the properties of certain *units*, i.e., the components that are considered basic, and (2) the relations that relate these units into *systems*, and some properties of these systems.

(1) The properties of basic units of composition. A theory about the micro-level reality should describe the properties of the units, which involves specifying what types of such properties there are and what values these properties might have.

For instance, in the atomic physics that I have sketched, an atom is considered a unit of composition, which is supposed to have properties such as mass (and charge, in which case this charged atom would be an ion.) Modern atomic physics has developed the tools and language to tell the value of atomic mass and charge. For example, the mass of an oxygen atom can be characterized by the number of its atomic mass, approximately 16, which means that the mass of an oxygen atom is about 16/12 of the mass of a carbon atom. (As we know today, an aggregation of about 6.02×10^{23} carbon atoms have a mass of 12 grams.)¹⁰⁸

(2) System composed by atoms. A fundamental theory should also account for some relations among the units and some properties of the systems of the units. Here a system doesn't have to be a whole in a mereological sense; but a system must have multiple units in relation. The fundamental theory doesn't have to account for all properties of a system, but it should be able to describe some facts about the system. If a theory about certain units of the world has no account of what happens when many such units come together, then it cannot initiate any account of things in the world as composed of these units.

We have seen that physical sciences can account for the relations among atoms. For example, according to atomic physics there are facts about the spatial relations among the atoms. (In physics, any two non-overlapping things have spatial relations.) For example, the distance between two carbon atoms in a diamond is 1.54×10^{-10} meter.¹⁰⁹

¹⁰⁸ On the definition of atomic mass and its measure, see https://en.wikipedia.org/wiki/Atomic_mass.

¹⁰⁹ Rossi, M. (2002). "How can graphite and diamond be so different if they are both composed of pure carbon?". *Scientific American*. [Retrieved on March 28, 2019] <https://www.scientificamerican.com/article/how-can-graphite-and-diam/>.

Based on their spatial relations and their individual properties such as mass and charge, we can also derive facts about the force upon an atom in this system.

And, atomic physics can also account for some properties of the systems of atoms. For example, the total mass of a system is the sum of the masses of the individual atoms. The size of a system of atoms can be geometrically determined based on the number of atoms and their spatial arrangement.¹¹⁰ Atomic theory can explain such system-level properties, although normally we can acquire information about the mass and size of a macro-level system just by measurement.

But some other system-level properties can be discovered or understood well only if one considers what happens at the micro level. For example (following E.g. 2 in the last section), we can observe that salt (sodium chloride) dissolves in water. But to explain (instead of simply observing) the exact rate of dissolution of sodium chloride in water and the effect of heat that accelerates the dissolution, one needs to consider the forces among the atoms or atomic compounds, their motions, the dynamics of forces and motions, etc. Also (following E.g. 3 in the last section), we can observe that we breathe air, consuming a component of it known as oxygen. But there are creatures that “breathe” sulfur rather than oxygen, such as certain bacteria that infect internal human tissues and release stinky smells. Why do human beings breathe oxygen but not sulfur? Part of the explanation of this requires understanding how atoms or compounds of sulfur would work with the compounds of atoms in human body, which normally use oxygen. Also

¹¹⁰ Sizes of atoms don’t contribute much to the sizes of macro-level systems in most cases.

(following E.g. 4 in the last section), we know that DNA underlies the inheritance of biological features over generations. But the discovery of DNA presupposes some established notions about systems of atoms (especially nucleic acids, consisting basically of atoms of hydrogen, carbon, nitrogen, oxygen, phosphorus, etc., and their bondings) and techniques for observing and measuring such systems (such as X-ray crystallography.)¹¹¹ In fact, many contemporary material sciences are impossible without the theory of atoms.

I hope that section 1 has presented a decent example of a theory that deals with micro-level composition. In section 2, I will present some representative accounts of pan(proto)psychism on micro consciousness, conscious qualities, and apparent structures of consciousness, which will show that, as a matter of fact, pan(proto)psychists have said little to articulate the kinds of facts which furnish an informative account of their view of micro-level composition.

5.2 What we know about the micro-phenomenal (according to pan(proto)psychists)

Section 1 has shown that a theory of micro-level composition should articulate (1) the properties of certain units, and (2) the relations that bind these units together into systems, and some properties of these systems. In this section I will describe some representative accounts of pan(proto)psychism to show that, according to

¹¹¹ Pray, L. (2008) Discovery of DNA structure and function: Watson and Crick. *Nature Education* 1(1):100. [Retrieved on March 28, 2019] <https://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397>.

pan(proto)psychists themselves, we know little about (1) the properties of micro (proto)phenomenal units and (2) their ways of combination that produces the system-level (proto)phenomenal units. In the sections that follow (sections 3 and 4), I will argue that there is an in-principle reason for such ignorance of micro (proto)consciousness.

5.2.1 About the properties of units in pan(proto)psychism

To my knowledge, theorists of pan(proto)psychism have not yet developed serious proposals about the characterization of the properties of the basic units in their theories. Little has been said about the types and values of micro-level (proto)phenomenal properties. A consensus is that these unit properties should be similar to phenomenal consciousness. Unlike paradigmatic physical properties, there must be “something it is like” or something close to that in having micro (proto)phenomenal properties. Beyond that, the water is mostly untested.

The fact is that, in the literature of pan(proto)psychism, little is said about what the basic (proto)phenomenal properties are. For example, in Philip Goff’s (2017b) monograph that advocates panpsychism, little is said about what micro-phenomenal properties might be like. Similarly, in Seager’s articles (2010, 2012, 2017) that argue for an emergentist version of panpsychism, little is said about the character of micro-phenomenal properties. Seager admits that “No very informative answer can be given at this point.” (2017, p. 240) He expects that some future theory may tell what the micro consciousness is like. Facing this difficulty of knowing micro phenomenal properties, Seager says that

The main point is that while it seems hopelessly mysterious to point that phenomenal experience is generated by a system of entities possessed only of physical properties such as mass, spin, charge, energy and a few more abstruse quantum mechanical properties, it does not seem similarly paradoxical that a system of entities possessed of elementary phenomenal features could generate complex states of consciousness... (2017, p. 240)

Thus it seems that Seager doesn't regard it as urgent for panpsychists to characterize micro phenomenal properties, because he believes that panpsychism is metaphysically superior to standard physicalism (for reasons considered in Ch. 1, section 1), even without characterizing micro phenomenal properties.

If there happens to be some more specific characterization of certain (proto)phenomenal properties, such specifications are just toy models of (proto)phenomenal properties fashioned after macro-phenomenal properties. These toy models are designed to illustrate certain ideas. Theorists who design such toy models are not committed to the micro-level (proto)phenomenal properties being the way characterized by the toy models.

For example, in an early article on panpsychism, Goff (2006) used a toy model of pain experiences in micro units to argue that "it is contradictory to suppose that the experiential being of lots of little experiencing things can come together to wholly constitute the *novel* experiential being of some big experiencing thing." (2006, p. 58) The story goes:

Even the experience of a severely pained subject of experience is sufficiently different from the experience of slightly pained subjects of experience as to make it incoherent to suppose that the former could be formed from the latter. For the

experiential being of some little experiencing thing ‘LITTLE’ to be part of the experiential being of some big experiencing thing ‘BIG’ s for what it is like to be LITTLE to be a part of what it is like to be BIG. But it follows from this that BIG feels how LITTLE does (even if it also feels other things). Correspondingly, for the experiential being of some BIG to be wholly constituted by the experiential being of LITTLE 1, LITTLE 2, LITTLE 3 ... [...] can be nothing other than for BIG to feel how it feels to be all those LITTLES and to feel nothing else. [...] it is contradictory to suppose that they could sum together to form some novel conscious state. (2006, p. 58)

Goff wrote this passage to criticize the possibility of the combination of micro phenomenal qualities into macro phenomenal qualities. The exact reasoning doesn’t concern us here. For example, one may question Goff’s claim that “it is contradictory to suppose that [LITTLES] could sum together to form [BIG].” But let’s leave this aside. I quote this passage merely to show a toy model of a panpsychist world, furnished by some familiar experience that is chosen somewhat randomly as the micro phenomenal property. In general, if a pan(proto)psychist wants a model to vividly illustrate a point, she would probably use a toy model which supposes that the micro-(proto)phenomenal properties are just familiar properties in macro-consciousness. The experience of pain is familiar enough. The experiences of color, especially that of redness, also seem popular as a toy example of micro consciousness.

This brief survey should suffice to indicate that the state of knowledge about the properties of the units (micro consciousness) in pan(proto)psychism is unsatisfying, even though pan(proto)psychists don’t seem extremely wary about it. Obviously the problem is very hard. Everyone seems to be waiting for a solution to come. But it’s also unclear from

their accounts if pan(proto)psychists are on the right track to prepare for a solution to come.¹¹²

¹¹² Lewtas (2013) is an exception among contemporary pan(proto)psychists, as Lewtas has proposed an actual serious model of the micro-phenomenal properties. The crucial point in Lewtas's account is the assumption that a certain quality, if it's a complex quality, can only be intelligibly explained (i.e., without "the explanatory gap") by basic qualities of the same type, and the typology of qualities is based on familiar human phenomenology. As a result, Lewtas's account contains a large number of types of basic qualities, presumably more than the types of basic physical entities/ultimates. Therefore some ultimates have to "simultaneously instantiate a plurality of strictly-basic conscious properties." (p. 57) But ultimates are supposed to be simple, and the qualities they instantiate must also be simple. The way Lewtas accommodates multiple simple qualities into one ultimate (quark) is to posit that an ultimate has multiple separate conscious fields or "subjects", "as many completely disconnected subjects as instantiated strictly-basic conscious properties." (p. 58) Thus, Lewtas's account of "quark consciousness" offers a pool of micro qualities that *significantly* resemble the qualities that we know, many of which are found in ordinary human consciousness.

Although the picture has some parts that we can imagine with some ease, metaphysically speaking this multiplicity of subjects in one ultimate seems implausible. Even though a conscious subject is not necessarily a psychological subject with thoughts and attitudes, claiming that conscious subjects are not metaphysically "heavy" (pp. 59-60) and therefore OK to be multiplied is like cheating. Epistemically, Lewtas's proposal also has issues. It still posits many items that we may never understand, esp. the basic qualities in animal and alien consciousness. Furthermore, of the sort of basic qualities that we seem to understand (e.g., simple color qualities in visual experience), Lewtas's account seems suspiciously easy. Why think that our ordinary consciousness reveals some aspects of micro consciousness? Lewtas makes this move because he rejects the "explanatory gap", and he believes that there is "explanatory gap" between apparently different types/modalities of qualities. Therefore, to explain macro consciousness from micro consciousness without the gap, Lewtas concludes that micro consciousness must include types/modalities of qualities similar to those in macro consciousness. But this understanding of the "explanatory gap" between apparently heterogeneous qualities is controversial, as many philosophers do consider it possible that our understanding of the dimensions of phenomenal qualities is insufficient. Also, this reasoning from macro qualities directly to the "atomic" phenomenal qualities proceeds without a clear concept of how qualities "combine" to explain qualities in a higher-level system. Therefore, whereas most other panpsychists recognize the difficulty of studying micro consciousness but don't tackle it, Lewtas (2013) may have tackled the topic with an accessible but mistakenly easy answer.

5.2.2 About the properties of systems in pan(proto)psychism

Thus pan(proto)psychism has little to say on the (proto)phenomenal properties of the units. Now recall, from section 1.2 of this chapter, that a theory on the combination of micro-level entities should also account for the systems of units. If explanations of the features and changes in observable entities can be informed at all by this theory, the theory should have something to say about what happens when multiple units compose a system (since an observable phenomenon is in virtue of the composition of the units).

In this section I will argue that, as it turns out, pan(proto)psychists have little to say about facts concerning the systems of units, about how the composition of micro (proto)phenomenal units determines the (proto)phenomenal properties of systems. Pan(proto)psychists tend to agree that macro-level consciousness manifests certain qualities and structures, which belong to systems of units rather than to the units themselves. But their proposals on these matters are meager and speculative.

To recognize the unsatisfying status quo of pan(proto)psychist account of systems of multiple units, one should check pan(proto)psychists' attempts to answering the "palette problem" and the "structure mismatch problem" (Ch. 1, section 3.2), since answering these two problems just is to explain how the combination of micro (proto)phenomenal units gives rise to the (proto)phenomenal properties of the systems of micro units for pan(proto)psychism.

This section can be regarded as a followup on the introduction of the palette problem and the structure mismatch problem introduced in Ch. 1, section 3.2, with

emphasis on exposing the current unsatisfying state of knowledge concerning these problems.

5.2.2.1 Of The Palette problem

About the “palette problem”, recall that it is about the question of how the varieties of macro-qualities can be explained in terms of (probably) a small varieties of micro-qualities in a (probably) small varieties of ultimates (Ch. 1, section 3.2). I will consider Roelofs (2014), Seager (2017), and Coleman (2017), who have offered some sketches of conjectures on the palette problem.

Roelofs. Roelofs (2014) has a “phenomenal blending” proposal to answer the palette problem, according to which “There are a small number of basic qualities” such that all phenomenal qualities are results of “blending” when some of the basic qualities are experienced together (p. 61). Roelofs argues that there are examples of phenomenal blending, in experiences of colors and flavors. Roelofs also argues that phenomenal blending could be ubiquitous; some apparently simple phenomenal qualities may be results of blending more basic qualities, only that we cannot tell. However, given that phenomenal qualities seem extremely diverse and heterogeneous, what basic qualities could be the fundamental phenomenal ingredients for all of them? About this, Roelofs says that,

[W]e cannot reasonably hope to get all qualities from any small set of known qualities, but the defender of blending need not think that the basic ingredients are known to us. Instead, the basic ingredients may be ‘alien’, unimaginable but not inconceivable. It is commonplace that there are such qualities: just as a

human born anosmic cannot imagine olfactory qualities, we are all similarly limited regarding the qualities of the many sensory modalities that humans lack. We can accept the existence of such qualities, but we cannot ‘know what they are like’. (p. 65)

I think this is a reasonable move to defend a panpsychism that acknowledges only a small set of basic phenomenal properties at the micro level. An alien property is acceptable as far as the concern is only to make the case for a metaphysical hypothesis. But this is also conceding ignorance regarding the details of the “phenomenal blending” proposal. We are offered the possibility that the palette problem might be solved by the “blending” of certain basic qualities, but we cannot know what these basic qualities are like. We cannot know, according to Roelofs’s proposal.¹¹³

Seager. Seager (2017) has two comments on the palette problem. One idea seems to be increasing the number of micro phenomenal properties at the micro level, by positing that

[E]ach kind of fundamental physical entity [host] a constellation of fundamental micropsychic features which then combine according to combinatorial infusion as dictated by the laws of infusion and the variegated associated physical systems. (2017, p. 244)

As for what this “constellation” is, Seager’s following passage suggests that it might be “a highly disunited consciousness in the sense that a number of incompatible basic micropsychic features belong to it”. (2017, p. 244)

¹¹³ This ignorance doesn’t bother Roelofs (2014). Indeed, although Roelofs (2014) sketches a proposal to address the palette problem, Roelofs’s goal in that work is to defend constitutive panpsychism, for which a sketch of a conceivable answer to the palette problem might suffice.

Seager hasn't given an example of such a constellation. Perhaps the suggestion is something like the following: electrons bear basic color qualia, photons bear basic taste qualia, etc. But in this proposal it's still unclear whether one particle bears only one qualia (e.g., an electron experiencing either red or green, but not both) or a number of different qualia (e.g., a photon bearing a "disunited consciousness" of bitter and sweet). Seager also hasn't explained what "basic micropsychic features" are incompatible.

Anyway, whatever Seager means, there is no specification of micro phenomenal properties or their kinds, no account of what kind of micro physical particles bears what "constellation" of micro phenomenal properties, and no account of "the laws of infusion".

Seager's other answer to the palette problem is that a small palette might be able to make a lot. By analogy, "it takes but three fundamental hues to generate all possible colors that humans can experience." (2017, p. 244) Again, this idea leaves out all details. We know that red, yellow, and blue pigments suffice to create all hues. But Seager has said nothing about what micro phenomenal properties have the similar effect of creating all experiential qualities.

Coleman. Coleman (2017) has considered two aspects of the palette problem.

On the one hand, the palette problem can be considered a "numerical challenge" (Coleman 2017, 10.3.iv), a task of finding enough number of basic micro phenomenal properties for making macro-qualities. Coleman doesn't regard this task as particularly difficult. Assuming that microphysical differences associate with differences in micro (proto)phenomenal properties, one may expand the stock of micro

(proto)phenomenal properties by finding enough number of states in microphysical entities.

Thus Coleman suggests that “Given that we’re likely to discover more particles, we can expect that number [of distinctive kind of qualities] to increase somewhat.” (Ibid., 10.3.iv) Coleman also conjectures that if one adopts string theory or “something like Bohm’s [...] of a much finer-grained fundament of entities and properties beneath the quantum mechanical level currently considered basic” (Endnote #43, Ibid., 10.3.iv),¹¹⁴ then there might be enough microphysical states to mirror a great number of micro (proto)phenomenal properties, because “Strings can potentially vibrate in infinitely many ways, and perhaps each vibration corresponds to a different quality.” (Ibid., 10.3.iv)

Perhaps one can raise metaphysical challenges to this proposal. For instance, apparently a macro experience of seeing red associates with certain activities of certain brain areas. But if, as Coleman suggests, that macro experience is determined by the combination of micro (proto)phenomenal properties, which in turn associates with various microphysical states, then it follows that one’s experience of seeing red is determined by certain micro (proto)phenomenal properties (say, micro-(proto)reddishness), which in turn associates with certain vibrations of strings in neurons. Unless specific types of neural activity patterns associate with specific types of string vibrations, I don’t know how Coleman’s proposal can cohere with available empirical observations on consciousness.

¹¹⁴ Coleman refers to Bohm (1980).

But, even without such challenges on the coherence of this proposal, like Seager's response to the palette problem, Coleman's proposal also leaves out the details. There is no account of micro-level qualities or their kinds, no account of what sort of particles or string vibrations correspond to what micro qualities, and no account of how micro-level qualities figure into macro phenomenal properties.

On the other hand, Coleman holds that the palette problem can be about the "incommensurability" of different sorts of experiences. For example, visual experience and tactile experience seem so different that it's hard to consider them as variations based on the same sort of micro qualities, like variable ways of string vibrations.

In response, Coleman proposes to adopt Hartshorne's "continuum hypothesis"¹¹⁵ by hypothesizing that all qualities are on a continuum. If "two sets of our qualia stubbornly appear absolutely different", this could be because there is "a missing 'intermediary' modality that we lack", which we miss due to our own biological contingency (Coleman 2017, 10.3.v). Thus, perhaps visual qualities and tactile qualities are on a same "spectrum," but we fail to understand it because there are some unknown qualities in between visual and tactile qualities on the spectrum. To make sense of this, Coleman presents the following example of a "color-experiences-only" species, who experience colors as if they are different modalities:

Imagine a creature whose qualia-space featured only (what we would call) colors. Though possessing our five external senses, their qualitative products in its consciousness would just be colors of different sorts, with no color that features in

¹¹⁵ Hartshorne, C. (1934). *The philosophy and psychology of sensation*. Univ. Chicago Press.

‘audition’ featuring also in ‘vision,’ and so on. We can get some grip on this being’s mental life by thinking about the way that heat—thought of primarily as a tactile quality—can feature in visual experience as red or orange. Now we just have to imagine that the creature, in touching a warm surface, experiences these visual qualia *only*, instead of the tactile ones we feel.

Plausibly this creature would come to conceive of the qualia corresponding to its various senses, what for us would just be different kinds of colors, as qualia belonging to irreducibly different ‘modal spaces.’ [...] Perhaps a being with a qualia-space correspondingly greater than ours as ours is greater than the color-only creature, would conceive of human qualia as belonging to a single ‘modality.’ (Ibid., 10.3.v)

I admire the imagination that renders sensible the metaphysical possibility of a single continuum of all qualities, although I should point out that Coleman’s hypothetical being who “would conceive of human qualia as belonging to a single ‘modality’” might nevertheless experience their own consciousness as belonging to radically different modalities. They may regard human qualia as all in one super-modality, and still have other super-modalities for themselves. But, besides the challenge just mentioned, obviously Coleman has left out the details. There is no account of what the continuum of all human qualia is like, since there is no account of the experience of that hypothetical being who has that greater qualia-space. And we don’t know if there is such a being of a greater qualia-space.

5.2.2.2 Of the Structure Mismatch Problem

About the “structure mismatch problem”, recall that it is about how micro-experiences of the ultimates combine to make macro-experiences of a certain structure that we seem to observe in our consciousness (Ch. 1, section 3.2). Since panpsychism

takes macro consciousness to result from micro consciousness organized in certain structures, and the neurophysiological structures are supposed to be involved in this, one might have some expectations about what one observes in macro consciousness.

However these expectations are unsatisfied. For example, one might expect to find micro consciousness in macro consciousness, just as one expects to find pigments and strokes in a painting. But no one can claim to have found micro consciousness. Or, one might expect to find neurophysiological structures mirrored by macro consciousness, just as one expects to find a scene on the painting to be structurally similar to the paints on the canvas. (For example, if at the place of a face there are many pigments, there one can see a complex expression rather than a blank face.) However one doesn't find neurophysiological structures in macro consciousness. For instance, if I enjoy a piece of homogeneous experience of greenishness, the structure of the macro consciousness I have doesn't seem to mirror the neural structure behind it.

This question about "structure" is admittedly obscure, as I have noted (Ch.1 section 3.2). Answers to this question often don't seem much clearer. In what follows, just to give a sense of what pan(proto)psychists might say about the structure mismatch problem, I will present Seager's (2017) and Coleman's (2017) answers.

Seager. Recall that Seager argues for an "infusion" view of panpsychism, according to which macro consciousness is a result of the fusion of micro consciousness (Ch. 2, section 1). Below is what Seager has said about the structure mismatch problem:

Certain versions of what might be called purely constitutive or structural panpsychism could fall victim to this objection. But [Seager's panpsychist

infusion] will not [fall victim] because (1) the postfusion state will not reveal prefusion constituents, and (2) the physical state that occasions mental fusion need not have its structure duplicated in the fused mental state. The example of quantum entanglement might be useful here. There is no way to tell by looking at the entangled constituents of such a state whether they are entangled or not. Only by looking at the system as a whole will this be revealed. We can interpret this as a denial that the constituting structure must be duplicated or reflected in the resulting states. (2017, p. 245)

To the extent that some sense can be made out of this response, Seager's point is that it's not a problem for his infusionist panpsychism that macro consciousness doesn't show the structures and components that generate it, because fusion is a process where the components and their structures are "lost" into one whole (Ch. 2, section 1). But clearly, Seager's response only tells what will *not* follow from the infusionist view: the "prefusion constituents" (presumably the micro consciousness before fusion) and "the physical state that occasions mental fusion" (presumably including neurophysiological states) don't show up in the structure in macro consciousness. Insofar as the structure mismatch problem is about how to accommodate the apparent absence of some structures in macro consciousness, Seager's answers have done the job. But, if one believes that macro consciousness has certain structures which should be explained by the arrangement of micro consciousness, then Seager has offered nothing to explain the generation of the structures in macro consciousness. After all, Seager has not attempted to specify the laws of panpsychist infusion.

Coleman. As for Coleman's reply to the structure mismatch problem, recall that Coleman holds "panqualityism" according to which what one experiences depends on higher-order thoughts (HOT) that represent the qualities (which results from

combinations of micro-level qualities.) For this reason, panqualityism can separate the structures that are involved in making consciousness (such as neurophysiological structures) and the apparent structures presented in consciousness. What structure is experienced in macro consciousness depends on HOT's task and target. HOT isn't supposed to represent physical structures of the brain:

[O]n HOT panqualityism, the requisite mirroring is not between what we are aware of and brain structure *simpliciter* [...] It's no surprise, given an environment of significant medium-scale dry goods (threats, food, etc.), that the HOT system should have evolved to 'gloss over' [...] the complexity, and amend the structure, of the pre-conscious raw qualitative feed; all in the cause of helpful isomorphism between the conscious percept and the subject's *environs*, not her brain. (Coleman 2017, 10.3.ix)

For example, concerning "the grain problem" (Ch. 1, section 3.2), Coleman says,

Why are we not conscious of [a given macrophysical brain portion's] microqualitative texture? Perhaps this is a matter of the relatively low 'resolution' of HOTs. [...] we can on HOT panqualityism in fact stand to the brain's structure (in consciousness) somewhat as one does in viewing a newspaper photograph: we perceive not the 'dots,' only the image. (Ibid., 10.3.viii)

Coleman's suggestion on the structure mismatch problem seems clearer than Seager's: HOT has many important targets to draw our attention to (whose importance is explained by plausible, well-understood biological concerns), and so neurophysiological structures are not to be mirrored in consciousness. However, like Seager's response, Coleman's answer only tells us what does *not* follow from panqualityism: i.e., it doesn't follow from panqualityism that we should be consciously aware of brain structures.

Insofar as the structure mismatch problem is about explaining the lack of presentation of certain brain structures in macro consciousness, Coleman's answer may suffice. But if there is a question about explaining the apparent structures of macro consciousness, then HOT panqualityism must tell us what sort of first-order states are targeted and represented by HOT, what sort of structures exist in those first-order states (which are represented by HOT, and so should be discernible in macro consciousness), and how micro qualities contribute to these structures. Panqualityism seems not ready to answer these questions.

Section 1 and Section 2 should suffice to show that pan(proto)psychism is very far from being a theory about the micro-level components of nature that could inform us about consciousness, unlike our knowledge of atoms that tells plenty about the composition of the physical world. One may think that this is a temporary state of lack of knowledge, that pan(proto)psychists would gain more knowledge about micro (proto)phenomenal properties in the future. However, I suspect that pan(proto)psychists will have a hard time reaching this knowledge. Thus, now that I have summarized what I perceive as the current epistemic record of pan(proto)psychism, in the rest of this chapter I will assess its epistemic future.

5.3 How is the study of the basic units possible: Transdiction

Physics is close to being a fundamental science about the physical world. This is because physics is the science that has established many facts about the micro-level components of the world, about the properties of atoms and other subatomic particles and

their relations. Similarly, if pan(proto)psychism is to suggest a science of consciousness about the fundamental components (basically by attributing (proto)consciousness to certain micro-level entities), such a pan(proto)psychist theory should also establish facts about micro-level entities. A pan(proto)psychist theory should tell us about the (proto)phenomenal properties (and relations) of micro-level entities.

How to achieve such knowledge about micro-level things, which goes beyond everyday knowledge and observation? Given the track record of physical sciences in prying facts about the micro world, it's helpful to delineate the logic that guides the studying and theorizing about the microphysical world, and then consider how the same logic can be transferred to the study of micro (proto)consciousness. In this section, I will explain this logic, with the study of atoms as an example.

There is a problem of “transdiction”, of studying the features of the micro-level, unobservable things by studying what is empirically more available. What we know today about atoms is a result of wrestling with the transdiction problem in physical science. Mandelbaum has described the problem of transdiction as concerning the justification of theories, esp. of corpuscularianism, “concerning the ultimate constituents of the physical world” while relying on experience as the ultimate source of human knowledge about the physical world, thus making it possible to use experiential data to

“say something meaningful and true about what lay beyond the boundaries of possible experience.” (1964, p. 61)¹¹⁶

The transdiction problem must be addressed if the study of micro-level facts is to be possible. In the study of the micro-level facts that cannot be directly observed, transdiction is necessary for generating hypothesis that can be formulated and tested.

- **Concepts.** A theory which posits certain general types of properties (such as micro phenomenal properties) should have the potential to become concrete, its detail able to be specified. Otherwise, the theory would be merely an abstract proposal. Regarding micro-level facts, our commonsense knowledge, before scientific studies or reflection, does not supply us with any concepts to describe them, because we have no direct observation of them. We therefore need a source of concepts. Macro-level models offer a stock of such concepts to formulate initial hypotheses about the micro.
- **Initial justification.** You might think that one may posit wholly new hypotheses about the micro-level world, and thus transdiction is not necessary for generating hypotheses about the micro-level. The problem with this proposal is that there can be multiple competing hypotheses about the micro-level unobservable world, none of which is directly confirmable empirically. Thus, at the beginning stage of the study of the micro-world, one can hardly select one theory from them on empirical grounds, unless one

¹¹⁶ Similarly, McGuire (1970) puts the question (which he calls “the problem of transduction”) as

[A]re there justificatory principles by which to sanction inferences from what is observable to what in principle is unobservable? If the basis of natural knowledge is strictly confined to what is actually or possibly observable, how can any claims be established about the nature and existence of atoms? (p. 3)

can find support for a theory from what one can observe, i.e., the macro world. For this reason, to study the micro-level world to begin with, a theory about the micro-level that has *some* empirical support is more justifiable than those that do not. Some empirical support is needed to motivate a theory about micro-level facts.

- **Connecting the micro and the macro.** Ultimately, the study of the micro-level facts should explain what one observes at the macro-level, and transdiction is about bridging the micro and the macro levels. The macro-level behavior of a collection and the relation between the macro-level system and its macro-level individual components can be understood with observation. Thus, through transdiction, if one considers a macro-level phenomenon as the behavior of a system of micro-level things, one can use the system-individual relation observed and understood at the macro level to understand the system-individual relation at the micro level. Then, one can understand and infer properties of the micro-level components of a system of micro-level things (the system itself being observable at the macro level).

In sum, transdiction makes further study of the micro level possible. An empirical theory should be formulated with specific concepts of items and properties that can be characterized in empirical terms, in terms of what can be observed and measured. We cannot directly observe the micro level. Therefore, studies of the micro-level facts must obtain justification through observable facts, which are at the macro level. Therefore, transdiction—to find the connection between the micro and the macro—is necessary.

Unsurprisingly, in general, transdiction is justified through analogy. The guideline, in short, is this: similar systems, whether at the micro or the macro level,

manifest similar properties. One begins the study of the micro world by considering micro-level systems (under certain descriptions) and macro-level systems (under the same terms of descriptions) that are structurally similar as behaving similarly, only differing in size. That initial modeling allows one to infer micro-level facts from observable facts at the macro level, if we can establish (or temporarily hypothesize) that they are similar.

Descartes has explained this rationale when discussing the methodology toward his atomic theory, where he considered that similar rules discovered in macro-level observable things may apply regardless of scales:

No one who uses his reason will, I think, deny the advantage of using what happens in large bodies, as perceived by our senses, as a model for our ideas about what happens in tiny bodies which elude our senses merely because of their small size. This is much better than explaining matters by inventing all sorts of strange objects which have no resemblance to what is perceived by the senses... (1644/1985, *Principles of Philosophy*, Bk. IV, section 201.)¹¹⁷

Transduction has been used to argue for the view that the physical world is made of small corpuscles and to infer the properties of the corpuscles and their systems. Even when atoms are largely out of the reach of empirical method, hypotheses in terms of corpuscles

¹¹⁷ Consider also Boyle, in a more vivid passage:

[T]o say that, though in natural bodies whose bulk is manifest and their structure visible the Mechanical principles may be usefully admitted, they are not to be extended to such portions of matter whose parts and texture are invisible, may perhaps look to some as if a man should allow that the laws of mechanism may take place in a town clock, but cannot in a pocket-watch ... ("Excellency of the Mechanical Hypothesis", p. 143)

may offer decent (though incomplete) explanations of many phenomena, giving the corpuscularian view a decent amount of credibility. Such explanations can be found in Robert Boyle, an early-modern advocate of corpuscularian philosophy.

According to Boyle, there are small pieces of matter, each has size, shape, and motion/rest (“Origin of Forms and Qualities”, p. 20; “Excellency of the Mechanical Hypothesis”, p. 141). And Boyle held that we could explain diverse phenomena just with these small corpuscles.¹¹⁸ One of Boyle’s main strategies to support atomism proceeds by arguing that empirically observable phenomena can be modeled by analogy with the behaviors of collections of atoms with these properties.

For example, as Mandelbaum (1964, pp. 107-112) and Newman (2006, pp. 204-206) have recognized, in “The History of Fluidity and Firmness” Boyle used analogies to explain the fluid and solid states of things in terms of corpuscles. Boyle held that liquids are made of tiny corpuscles that “glide” in the spaces between their surfaces, “touching one another in some parts only of their surfaces [...] and separately agitated to and fro” (“History of Fluidity and Firmness,” Part I, “Of Fluidity”, section II, p. 241). In contrast with fluidity, according to Boyle firmness manifests in a body if its consisting corpuscles are stably connected because “there is among them a mutual cohesion

¹¹⁸ “For if, with the same bricks, diversly put together and ranged, several walls, houses, furnaces, and other structures, as vaults, bridges, pyramids, &c., may be built, merely by a various contrivement of parts of the same kind, how much more may great variety of ingredients be produced by — or, according to the institution of nature, result from — the various coalitions and contextures of corpuscles, that need not be supposed, like bricks, all of the same or near the same size and shape, but may have amongst them, both of the one and the other, as great a variety as need be wished for, and indeed a greater than can easily be so much as imagined?” (“Excellency of the Mechanical Hypothesis”, p. 150)

whereby they are rendered unapt to flow or diffuse themselves every way” (“History of Fluidity and Firmness,” Part II, “Of Firmness”, p. 254). With an analogy, Boyle argued that liquids are collections of *very* small corpuscles:

[O]f bodies, that consist of incoherent parts, and are made up, as it were, by aggregation [...] in their being poured out, most resemble liquors, that are the smallest; as would appear upon the emptying of several sacks, the one of apples, the other of walnuts, the third of filberts, the fourth of corn, the fifth of sand, and the sixth of flour. (“History of Fluidity and Firmness,” Part I, “Of Fluidity”, section XI, p. 244)

That is, one can observe that the outpouring of an aggregation of homogeneous sphere-like items creates a liquid-like flowing gathering, from which one can infer Boyle’s general model of fluids as made of corpuscles that “glide” along one another. But furthermore, as one observes apples, walnuts, filberts, corn, and sand pouring out respectively, one observes that as the grain of outpouring items getting smaller, the result behaves more liquid-like. From this comparison of observable cases, one can infer that the corpuscles that compose liquid must be very small.

The analogy which explains the apparent fluidity in aggregations of tiny corpuscles is also utilized to explain the change of solid things into liquid state under high heat:

[I]f [nitre] be beaten into an impalpable powder, this powder, when it is poured out, will emulate a liquor [...] But this is but an imperfect fluidity, both because the little grains or corpuscles of salt, though easily enough moveable, are not always in actual motion; and because they continue yet so big, that both they and the spaces intercepted betwixt them are, near at hand, perceivable by sense. But if with a strong fire you melt this powdered nitre, then each of the saline corpuscles being sub-divided into I know not how many others, and these insensible parts

being variously agitated by the same heat [...] the whole body will appear a perfect liquor ... (Ibid., section XVI, p. 246)

That is, given powders which “emulate” liquids due to their small corpuscles, heat further divides the corpuscles into yet smaller ones. Plus the fact that heat adds more motion to corpuscles, we end up with liquid—a body of very small corpuscles moving along each other’s surfaces.¹¹⁹

Boyle used the same strategy to offer an account of colors. For example, Boyle considered white color as caused by the surface of an entity reflecting most of the light rays to the observer (“The Experimental History of Colours”, Part II, “Of the Nature of Whiteness and Blackness”, Ch. 1, p. 24).¹²⁰ To argue for this account of white color, Boyle used multiple observations, including the following analogy which compares a white surface made of light-reflecting corpuscles with the milky way that consists of many stars:

[W]hereas [diverse] parts of the sky, and especially the milky way, do to the naked eye appear white [...] yet the galaxy looked upon through the telescope does not shew white, but appears to be made up of a vast multitude of little stars;

¹¹⁹ See also Newman 2006, pp. 204-206 for a similar analysis of the analogies cited here. Newman holds that the case of melted nitre is not merely using analogy, because “Here we pass from analogy at different levels—as in the case of the bags filled with discrete items—to identity at different levels.” (p. 206.)

¹²⁰ Boyle identified whiteness of a body with the fact that

[T]he superficies of the body [...] is asperated by almost innumerable small surfaces; which being of an almost specular nature, are also so placed, that some looking this way, and some that way, they yet reflect the rays of light that fall on them, not towards one another, but outwards towards the spectator’s eye. (“The Experimental History of Colours”, Part II, “Of the Nature of Whiteness and Blackness”, Ch. 1, p. 24)

so that a multitude of lucid bodies, if they be so small, that they cannot singly or apart be discerned by the eye, and if they be sufficiently thick set by one another, may by their confused beams appear to the eye one white body. And why is it not possible, that the like may be done, when a multitude of bright and little corpuscles being [crowded] together, are made to send together vivid beams to the eye, though they shine but, as the planets, by a borrowed light? (Ibid., p. 26)

From these several examples in physical science, it should be clear that drawing analogy from the behaviors of observable systems can offer significant empirical insight into corpuscularian philosophy, since it offers valuable clues on how a system of corpuscles may behave. It can also offer illuminating sketches for understanding the properties of observable things as consequences of corpuscles in composition.

But such accounts of phenomena in atomic terms inevitably leave out many details about atoms. In fact, despite the popularity of corpuscularian philosophy, ideas of atoms largely failed to help the practice of chemistry in 17th and 18th century, during which the empirical connection between corpuscles and the observed chemical

phenomena was obscure.¹²¹ The sort of transdiction that empirically recognized atoms as we know today was realized by Dalton through a much more subtle reasoning than the Boyle's examples that I have shown.¹²²

Dalton's reasoning was more subtle, in the first place, because Dalton used a more developed (and empirically informed) concept of corpuscles. That is, Dalton considered atoms as belonging to different "elements" as defined by Lavoisier, which gives elements or kinds of atoms an *empirical* meaning, as substances that cannot be separated by experimental methods (e.g., dissolving, burning, mixing with other substances, etc.).¹²³ Given how such elements are defined, aggregations of particles of such elements can be empirically identified. Furthermore, the property that Dalton relied on to infer about atoms of different elements is a measurable quantitative property (i.e., relative weight.)

¹²¹ Lavoisier, the "father of modern chemistry," said in 1789 that

All that can be said upon the number and nature of elements [...] is, in my opinion, confined to discussions entirely of a metaphysical nature. [...] I shall be content, therefore, in saying that if by the term *elements* we mean to express those simple and indivisible atoms of which matter is composed, it seems extremely probable we know nothing at all about them; however, if instead we apply the term *elements* or *principles of bodies*, to express our idea of the last point which analysis is capable of reaching, we must admit as elements, all the substances into which we are capable, by any means, to reduce bodies during decomposition. Not that we can be certain that these substances we consider as simple may not be compounded of two, or even a greater number of principles; but, since these principles cannot be separated, or rather since we have not hitherto discovered the means of separating them, they act with regard to us as simple substances, and we ought never to suppose them compounded until experiment and observation has proved them to be so. (quoted from Brock 1993, p. 129.)

¹²² On Dalton's atomism, see Brock (1993), Ch. 4; Siegfried (2002), Ch. 14.

¹²³ Lavoisier's definition of "elements" can be clearly recognized in the passage quoted in Footnote #121.

Around 1800, chemists observed that substances often combine with fixed proportions. For example, the ratio of weight between an acid and an alkali when they combine into a neutral substance (a salt) is a constant, suggesting a fixed proportion of elements in the salt.¹²⁴ Dalton realized that an atomic hypothesis can explain such weight relations among substances, including the fixed proportions: If each atom of an element has a fixed weight, assuming that atoms combine in relatively small integer numbers (such as 1+1, 1+2, 2+3, etc.), then the elements in a compound will have a fixed weight proportion. Thus, if element A has atomic weight X and element B has atomic weight Y, then a (pure) compound of A and B will manifest a fixed weight ratio such as X:Y, or X:2Y, 2X:3Y, etc. This ratio in atomic compounds is empirically accessible, because a pure sample of substance is just an aggregation of many such compounds which all have the same fixed proportion. Dalton thus discovered a way to empirically account for *what* are combined: what's combined are atoms of various chemical elements, each characterized by its atomic weight; how many units of an element might exist in a compound can be learned by calculation from the fixed weight proportion and the atomic weights.

Dalton's reasoning consists of a *transdiction* between the macro and the micro levels, through the idea that if substances are made of small atoms with fixed characteristic weights combined in fixed ways in compounds, then the weight proportion of the elements (from which the relative weights of elements can be derived) would be a constant, whether in an unseen unit of this compound or in an observable amount of it. The empirical success of the atomic weight theory confirms that the relative weight is

¹²⁴ See Brock (1993), pp. 130-131; Siegfried (2002), pp. 224-32.

such a property that is the same at both the macro and the micro levels. Given a measurable property that is the same at both the micro and the macro level, empirical investigation of the micro level becomes possible. Thus, like Boyle, Dalton also used transdiction. But unlike Boyle, the “analogy” between the micro/macro level used by Dalton offers empirical clues to the properties of atoms, and is conceptually more developed and precise.

The foregoing examples suffice to show what’s needed for transdiction from the macro to the micro. Some form of *analogies* between macro-level observable states and hypothesized micro-level states are needed. To explain, or even gesture toward an explanation of a phenomenon in terms of the combination of micro-level entities, we

need analogies to model the micro world.¹²⁵ There are preconditions for such analogies to be possible.

(a) First, we need plausible assumptions about similarities between big and small things or “units” that shall figure into the systems under study. If a model with an arrangement of macro-level entities can illustrate an arrangement of micro-level entities, and if the intrinsic properties of the macro-level things is part of what allows this macro-level arrangement, then in the micro-level states modeled by the macro-level

¹²⁵ The use of analogies in transdiction doesn't mean that the end product of the investigation of the micro entities will yield a picture of entities that look similar to things in ordinary life. The application of analogy is a complex matter, especially as the research advances to a level far from ordinary experience. Quantum mechanics may illustrate this complexity. (1) In the first place, far before quantum mechanics appeared in physics/natural philosophy, the idea of particles or micro-level entities and events started with analogy. The knowledge toward micro-level entities has to start with analogies with macro-level things. Physicists could not simply begin with the unintuitive notions of quantum mechanics. If a physicist in 17th century started with a theory like quantum mechanics, given the state of empirical knowledge at that time, that theory would be entirely without ground. (2) Second, the unusual behaviors of micro-level entities described in quantum mechanics doesn't show that the methodology of analogy is rejected in quantum mechanics. (2.1) In the first place, understanding of the quantum behaviors of particles *is* achieved through analogy with familiar models. For example, consider the double-slot experiments that demonstrate the wave-particle duality. It is the application of analogy that allows the characterization of electrons as wave-like when interference is observed. Likewise, the application of analogy makes possible the characterization of electrons as particle-like when no interference is observed. (2.2) Furthermore, the reason that the picture of quantum mechanics seems unlike the familiar macro-level world is precisely that, *after characterizing the observed behaviors of micro-level particles by analogy with familiar models*, we find that the story defies our expectations of macro-level phenomena. To cite the earlier example of the double-slot experiments: we don't know anything that manifests the wave-particle duality in our ordinary experience. Thus, quantum mechanical models are still developed by analogy with models that have been better understood. Without analogies that make possible the description of particle behaviors, about the particles we would only have a vague sense of ignorance but not the articulated puzzles in quantum mechanics. (Thanks to Joseph Levine for raising quantum mechanics as a potential counterexample to my account of transdiction in terms of analogy.)

arrangement, we should assume that the intrinsic properties of the micro-level things are *similar* to (although maybe “simpler” or in a smaller quantity than) those of the macro-level things.

(b) Moreover, we need assumptions about similarities in the arrangement between the micro and the macro levels. The relations among the micro-level things should be similar to the relations among the observable macro-level things.

(c) To complete a micro/macro analogy, a system composed of micro-level things should manifest a property (supposedly the observed property to be explained) that is similar to the corresponding property of the system composed of macro-level things in this analogy. Or, conversely, the macro-level system should manifest a property that is similar to the property that it models, i.e., the property of (hypothetically) a system of micro units. This is particularly important, because it is the goal of transduction to use properties of macro-level things/systems to guide the study of micro-level things/systems, to support or test micro-level accounts.

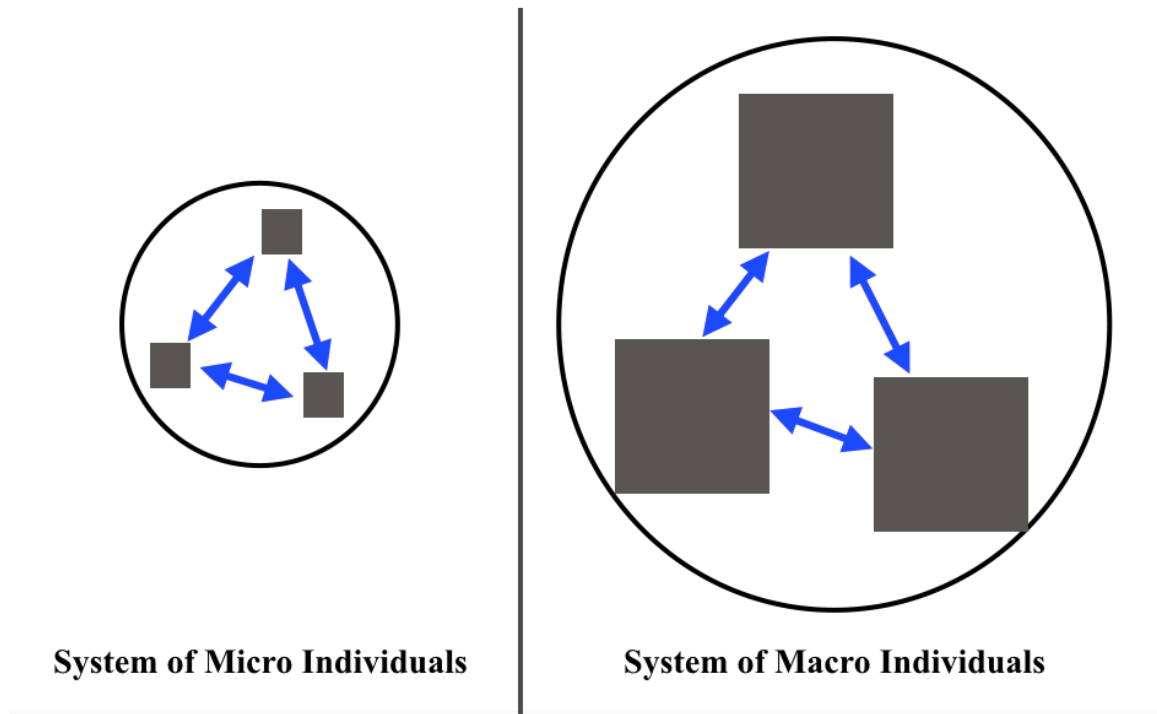


Figure 2. Micro and macro systems in transdiction. Using analogy with a macro-level system to model a micro-level system requires: (a) One can reasonably hypothesize that micro-level individuals have similar properties as those of the macro-level individuals. (b) One can reasonably hypothesize that the relations among micro-level individuals are similar to those among the macro-level individuals. (c) The macro-level system in this analogy does manifest the property of the micro-level system that is modeled.

The aspects of micro/macro analogy in (a)-(c) (as shown figuratively in Figure 2) can be found in the examples of transduction. In Boyle's corpuscularian conjectures, **(a)** the properties of corpuscles are hypothesized (including size, shape, motion), even though Boyle didn't have the concept or method to study such micro-level properties precisely. And these are properties shared by macro-level objects. For example, without assuming sizes and motions of corpuscles and the influence on motions by factors such as heat, Boyle cannot support his account of fluidity with analogies of rolling apples or walnuts pouring out from sacks. **(b)** Boyle's analogies also presuppose certain relations among corpuscles. For example, to compare corpuscles in liquid with a pile of rolling apples or walnuts, Boyle should assume that, like the apples or walnuts, corpuscles are individual entities in space and can touch against each other. **(c)** The success of Boyle's analogies are measured by the similarity between the (to be modeled) properties of systems of micro corpuscles and the properties of the analogous systems of macro objects. For example, to exploit the analogies on fluidity once more: Boyle's corpuscular model of fluidity could make sense, thanks to the fact that piles of moving objects (such as rolling apples, walnuts, or corns) do behave like fluids to some extent.

These aspects of micro/macro analogy are also present in Dalton's atomism. **(a)** Concerning properties of units, at both the micro and the macro level Dalton chose weight. Both micro-level atoms and macro-level samples of elements have weights. **(b)** The relations among units in this case are simple: the units can be put together or separated. At the micro level, different kinds of atoms combine or separate, just as at the

macro level, samples of different elements combine or separate. This simple notion of relations can be put to practice, first because of the empirical definition of elements as that which cannot be further divided with experimental means (and thus in experiments we do not expect elements to change their identity), and second because of the assumption of conservation that the amount of an element is constant before or after experiments (and so an element doesn't increase or decrease in an experiment.) The result is that, in a system, all that is left for the units to do is to gather or separate (as far as the interest is simply about chemistry.) **(c)** The success of Dalton's atomism is tested by the precise closeness between a property of systems of micro-level atoms (which is theoretically derivable according to Dalton's hypothesis) and a property of the comparable systems of macro-level substances. That is, the weight ratios of atoms in a

compound derived according to Dalton's atomism is consistent with the weight ratios of elements obtained in empirical analyses.¹²⁶

If it is correct that the study of micro-level world (transdiction) proceeds through analogies across the micro/macro levels, then concerning the study of micro (proto)consciousness in pan(proto)psychism, it's also plausible that analogies across the micro and macro levels are necessary. Thus, in the rest of this chapter, I will examine whether pan(proto)psychism has good analogies to facilitate studies of micro (proto)consciousness.

5.4 Transdiction in Panpsychism?

In this section, I will argue that panpsychism doesn't have good analogies to facilitate studies of micro-level consciousness. The epistemology of panpsychism, of the

¹²⁶ Let me note that besides having a micro/macro analogy, more is required if the transdiction is to become more precise, as close to the precision in accounts of chemical composition in physical sciences. Analogies across micro/macro levels are still required for more precise use of transdiction, as in the case of Dalton, whose atomism of chemical compounds is based on the analogy through the constancy of *relative weight* at both micro and macro levels. But notice what's present in Dalton's transdiction that is lacking in Boyle's: Dalton's account analyzes physical things in terms of "elements" understood as the simplest substances obtainable through empirical means, whose status as *fundamental* components was reasonably suspended. The assumption of empirically identifiable elements is crucial, since it allows Dalton to state *what* empirically identifiable components are involved. The caution against treating chemical elements as fundamental is also correct, because atoms of chemical elements are not physically fundamental. The atoms of chemical elements are *intermediate structures* identified through chemical methods, which leave open the possibility of yet smaller sub-atomic structures. Thus, it's plausible that one crucial aspect in which Dalton's transdiction differs from Boyle's is that Dalton's account includes an *intermediate level*. Therefore if pan(proto)psychism is to become a theory of "mental chemistry," we should expect pan(proto)psychism to have some empirically-definable notions about the intermediate-level (proto)phenomenal properties. But this is a topic for another paper.

possibility of studying micro-level consciousness, will be examined following the logic of transdiction, as explained in the previous section (section 3).¹²⁷

One announcement before I proceed. In what follows I will focus on the prospect of transdiction in panpsychism, leaving out panprotopsychism. The reason, as is explained in the previous section and will be illustrated further in this section, is that transdiction requires assumptions about micro-level properties that are somehow similar to macro-level properties. Since proto-consciousness or protophenomenal properties are neither consciousness nor paradigmatic physical properties, it's hard to say what protophenomenal properties are at all, not mentioning what they are at the micro level. This also explains the lack of development of panprotopsychism (compared to panpsychism). Since little can be said about protophenomenal properties, analogies about them can hardly begin.

Consider analogies in transdiction. Recall that, for an atomic hypothesis (à la Boyle's, see section 3) in terms of the analogies across the micro/macro levels to have any explanatory value at all, one should spot the properties, relations, and behavioral

¹²⁷ Section 3 of this chapter delineates a general methodology for the study of micro-level facts. Section 4 applies this methodology to panpsychism by examining the possibility of the knowledge of micro-consciousness with the general methodology from section 3. Although sections 3 and 4 of this chapter discuss the possibility of knowing micro-consciousness self-consciously in terms of transdiction, other philosophers writing on panpsychism may have offered pieces of similar discussions without self-conscious attention to the general methodology for the study of micro-level facts. It is to be expected that the general methodology and its application to panpsychism are being followed by some current approaches to panpsychism in the literature, if only implicitly. For this reason, some discussion in this section (in section 4.1, 4.2) can relate to existing ideas in panpsychist literature. In some of the upcoming footnotes (Footnote #128, #131, #133, #136), I will point out when an attempt to apply the methodology of transdiction in a specific way is exemplified in recent literature.

patterns in macro-level things/systems that can be justifiably applied to the micro level. Simply put, an analogy should first find certain models in the macro world, and then extend it to the micro world. For example, in Boyle's analogies, macro objects have motion, size, shape, which are then extended to Boyle's atoms by analogy. These properties of physical bodies provide ample space for elaboration, as can be illustrated by what can be made out of legos or billiard balls. Already we understand plenty of possible combinations of macro physical objects. Thus, if the analogy with the microphysical should be made, many analogies can be formulated.

So, to think about physical properties of micro entities, analogies are initiated by thinking about the physical properties, relations, and systems of macro things, which are then extended to micro-level things. Similarly, to make analogies that lead toward micro phenomenal properties, it makes sense to first think about the phenomenal properties, relations, and systems of macro-level things, and then consider if one can posit similar phenomenal states in micro-level things by analogy. In other words, to understand the combination of micro-level consciousness, one should first have accounts of the combination of macro-level phenomenal properties, and then extend them to the micro level by analogy.

So, what accounts of macro-level phenomenal combinations are available? There seem to be two ways to think about macro phenomenal combinations (Figure 3):

- **Analogy (i).** The macro-level phenomenal things we select for combination are conscious subjects of individual persons, which may correspond to activities of individual heads or brains as associated physical substrates.
- **Analogy (ii).** The macro-level phenomenal things in combination are token experiences such as *J's current experience of seeing whiteness* or *J's current experience of seeing redness*, which may correspond to activities of brain parts as associated physical bases.

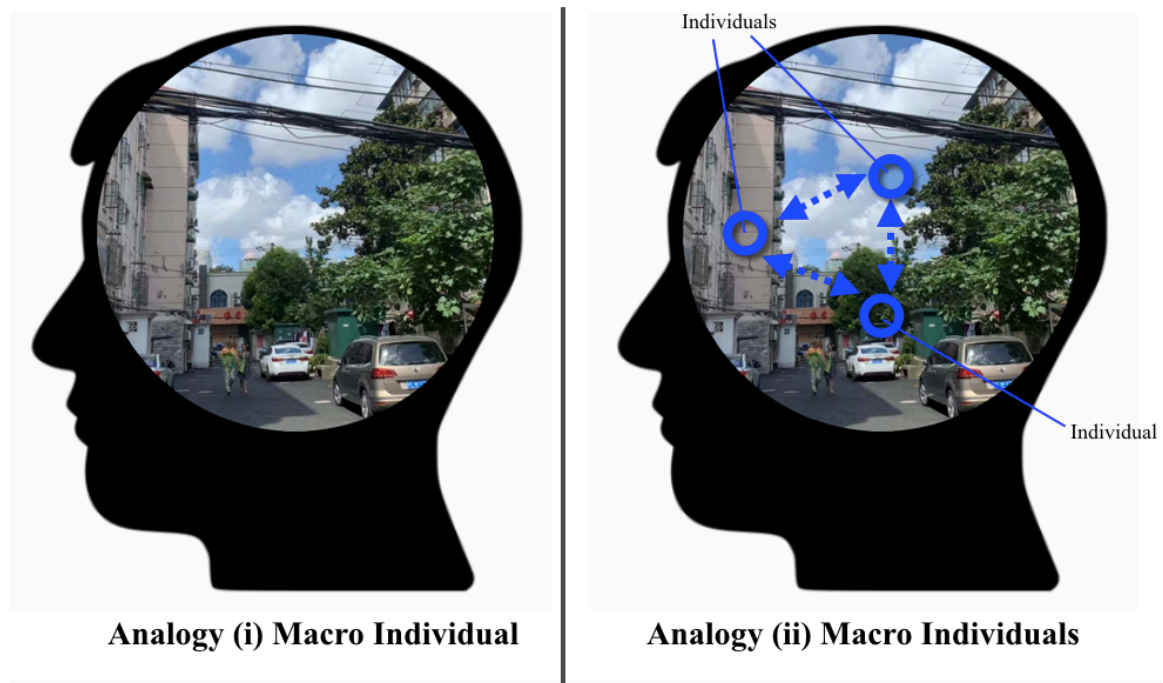


Figure 3. Analogy (i) and Analogy (ii). In analogy (i), an entire head is a phenomenal individual. In analogy (ii), a token experience is a phenomenal individual.

I want to argue that, given current state of knowledge, it seems that neither analogy can provide much insight into the details of how the micro-level phenomenal world works. We lack an analogy in which a macro-level model can be extended to the micro level.

5.4.1 Analogy (i)

Suppose, following analogy (i), we consider the (combined) macro-level units with phenomenal properties as entire brains with consciousness. What macro-level account can feed into an analogy with micro phenomenal combination? Recall (from section 3) that, for such an analogy to be possible, we need, at both the micro and the macro levels

- (a) properties of the units to be combined,
- (b) relations among the units (that organize them into a system), and
- (c) the system-level properties.

Thus, if we are to understand micro consciousness combination through an analogy with the combination of macro-level “mental legos” conceived as consciousness born by brains, we shall ask for a sensible macro-level story that may extend to the micro level. This macro-level story should cover (as illustrated in Figure 4)

- (a) (phenomenal) properties of individual heads/brains, as the properties of macro units,

- (b) relations among the brains (that may combine into a system), and
- (c) the (phenomenal) properties of the system of heads.

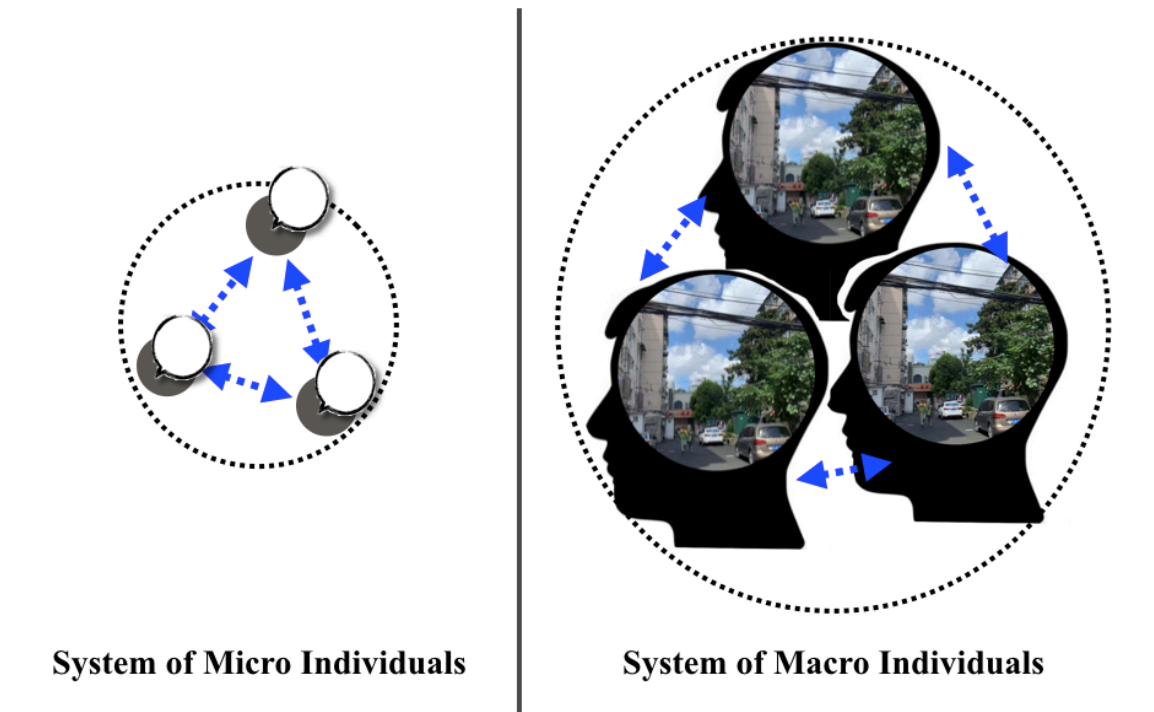


Figure 4. Analogy (i). In Analogy (i), the macro-level individuals are entire conscious heads/brains.

So let's consider what sensible things we could hold about (a)-(c).

(a) If macro units are consciousness of entire brains, then we seem to have a good grasp of the phenomenal properties of the units at the macro level. The question is, under this condition, what might be the phenomenal properties of the micro-level units?

Not much. It may be plausible to suppose that the phenomenal properties of the micro units are extremely "simple," but it seems that nothing more can be said with confidence. One may attempt to identify the phenomenal properties of the micro units by considering what's in common in all phenomenal properties that one knows, that is, in all macro-level consciousness one could imagine. The micro phenomenal property could be a property of this common dimension, and other varieties of phenomenal properties in macro consciousness might be variations of this common dimension. But it's hard to identify the common phenomenal factor in all forms of experiences. Touch, vision, and audition provide different modalities of conscious experience. Is there a common dimension of experience among these modalities, or is there one modality to which other modalities can be reduced? Furthermore, there are sense modalities such as magnetoreception (the sense of the magnetic fields in animals such as pigeons), which seems to have little in common with human senses. It's also hard to see what common phenomenal dimensions can be found among sense experiences and the experiences of expectation, longing, feeling of pleasure, etc. So, despite our apparent knowledge of macro-level consciousness, little of this knowledge can inform us about the phenomenal properties of the micro units.

(b) If macro units are consciousness of entire brains, what relations can be recognized among them, and which of these relations can be extended to the micro level?

Apparently the relations among heads/brains are not especially different from those among discrete physical objects, like several billiard balls. We may include spatial relations among the heads, and perhaps other physical relations such as forces of attraction or repulsion (e.g., through universal gravitation). And, it seems plausible to extend spatial relations and physical relations such as forces to micro-level units.

One may suppose that there are “forces” from one head to another due to phenomenal instead of physical properties. For example, maybe I feel pain, which (through empathy) can cause you to feel pain (for me.) However, this “force” is not a nonphysical effect of my brain over yours, because it is exerted by the physical process of your perception of me or the context. In general, human heads/brains do have “forces” that are absent in billiard balls, but (unless one believes in spurious parapsychological force among individual heads) such effects are mediated through physical means of communication.

Therefore, based on what we can observe, the relations among macro units as conscious brains seem largely physical, in the forms of spatial relations and physical forces, which can be plausibly assumed to exist as among the micro units.

(c) However, if macro units are consciousness of entire brains, what might be the macro-level phenomenal properties of the systems? This is crucial for panpsychists, because one fact that panpsychists must respect is that there is macro-level consciousness, which presumably is the phenomenal property of a system of micro units.

Thus, if an analogy with macro-level mental combination is to shed light on this phenomenon, we need the combination of macro units to provide a system that has its own consciousness. In the current analogy where macro units are consciousness of entire brains, we want a system of macro units (as models of micro units) to have a system-level consciousness (that models the consciousness of a system of micro units.)

But this is where the analogy breaks down, because a collection of multiple conscious heads does not seem to readily generate consciousness. This potential breakdown is the source of the classic form of the combination problem, the problem of subject summing (Ch. 1, section 3.1) In the classic formulation of this problem, vividly illustrated by William James (see Ch. 1, section 3.1), an analogy is drawn between micro-consciousness and individual personal consciousness, and the lack of micro phenomenal combination is concluded by analogy with the lack of combination of macro consciousness (as shown in Figure 5).

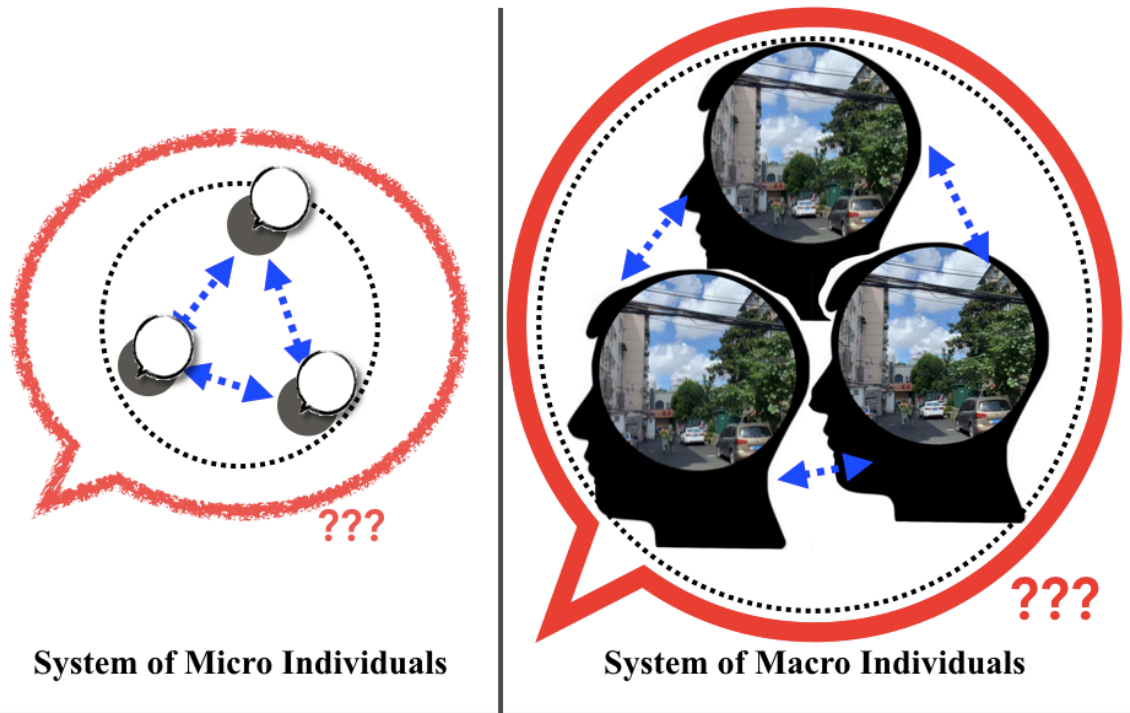


Figure 5. Issue with analogy (i), (c). To account for the combination of micro-consciousness into a system that has a further consciousness, the analogy in terms of a macro-level system in Analogy (i) requires that the macro-level system has a system-level consciousness. But we have no knowledge of this system-level consciousness in such a macro-level system.

Attempts of solving the subject summing problem often proceed as responses to this potential breakdown of analogy. **Maneuver on (c).** Thus, one may argue that, contrary to intuition, a system of individual conscious heads might have system-level phenomenal properties. **Maneuver on (b).** Or, one may argue that perhaps the current analogy is not entirely apt because it has missed something. In particular, there are special relations among the micro units that make consciousness in a system of micro units, which are apparently missing in a macro-level system of heads/brains that includes only well-known physical relations.¹²⁸

Issues with analogy (i). Thus, if in formulating the micro/macro analogy of panpsychism, one considers the macro units as individual heads/brains with consciousness, the analogy might collapse due to the apparent lack of macro-level combination of consciousness. To maintain the analogy, one needs additional metaphysical maneuver. The macro-level story itself, as can be understood most naturally without additional metaphysical maneuver, isn't apt for providing an analogy.

But the difficulty cannot be fixed simply with some metaphysical moves, if panpsychists intend their hypotheses to be conducive to further studies of consciousness. In this aspect, it seems that no fix can help.

¹²⁸ Goff's "phenomenal bonding" proposal is an attempt of this sort, as it hypothesizes that there are special relations (i.e., phenomenal bonding) such that micro-subjects in these relations would produce a composite that has consciousness. See Ch. 2, section 2. (The emergentist approach in terms of special nature laws that relate micro and macro consciousness may also belong to this sort of attempts. See Ch. 2, section 1.)

Of the Maneuver on (c). Concerning the proposal that a system of conscious heads/brains can have system-level consciousness, the problem is that we don't know what such system-level consciousness is. For example, we don't know *what it's like to be a parliament*. (This is different from *what it's like to be in a parliament*.) It's also unclear who can verify or falsify a claim about system-level consciousness. The ensuing analogy is therefore uninformative, since in the macro-level story, which is supposed to serve as the model of the micro-level story, the phenomenal properties of a system of conscious heads/brains are merely hypothetical.

Of the Maneuver on (b). As for the proposal of special relations among micro units (e.g., phenomenal bonding) that make system-level consciousness, if similar special relations are supposed to exist among macro units such that conscious brains can be related in similar ways to produce a system-level consciousness, then this proposal suffers a similar difficulty as the one just mentioned. Since no one knows what it is like to be a system of conscious brains, the ensuing analogy between the micro and macro levels would be uninformative. But, if the special conscious-system-building relations are restricted to micro units, then we have to abandon the current analogy, since micro units are held together with conscious-system-building relations whereas macro units are not. Though logically coherent, this move effectively kills the attempt of transduction through analogy with macro-level systems.

To summarize, one may propose an analogy for micro-level mental combination in terms of the combination of macro phenomenal units, where the macro phenomenal units are considered as conscious heads/brains. But the macro-level story of this analogy

cannot be completed informatively. **(a)** In the first place, an analogy with individual personal consciousness suggests little about micro consciousness. **(c)** But more importantly, the analogy requires that a system of conscious heads/brains has consciousness, and there should also be clues about the character of this system-level consciousness. However, as far as our experience can tell, systems of conscious heads/brains have no consciousness, or we cannot tell what sort of consciousness it is like.

5.4.2 Analogy (ii)

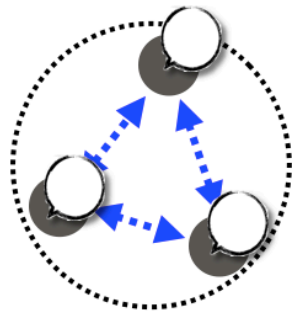
Suppose, following analogy (ii), we consider the (combined) macro-level phenomenal things as token experiences, or that which bear token experiences as their phenomenal properties. That is, we may consider the many experiences in one head as indicating the presence of many macro phenomenal units, instead of treating an entire conscious head as a macro unit. If so, what macro-level phenomenal stories can be told to feed into an analogy for micro consciousness combination?

Again, to formulate a micro/macro analogy, we need, at both the micro and the macro levels

- (a)** properties of the units to be combined,
- (b)** relations among the units (that organize them into a system), and
- (c)** the system-level properties.

Thus, if we are to understand micro consciousness combination through an analogy with the combination of macro-level units conceived in terms of token experiences, we shall ask for a sensible macro-level story that may extend to the micro level, which should cover (as illustrated in Figure 6)

- (a) (phenomenal) properties of macro-level units that bear token experiences,
- (b) relations among these units that bear token experiences, and
- (c) the (phenomenal) properties of the system of such units.



System of Micro Individuals



System of Macro Individuals

Figure 6: Analogy (ii). In Analogy (ii), the macro-level phenomenal individuals are token experiences.

So let's consider what sensible ideas could be held about (a)-(c).

5.4.2.1 Possible macro phenomenal elements in (ii)

(a) If the phenomenal properties of macro units are token experiences, what could we know about micro-level unit phenomenal properties?

Not much, I suspect. Transdiction from the phenomenal properties of token experiences to those of micro-level consciousness is hard, for reasons similar to those of why phenomenal properties of conscious heads (section 4.1) suggest little about micro-level consciousness. Macro-level phenomenal qualities could be visual, auditory, pain, pleasure, etc., in different varieties.¹²⁹ To determine micro phenomenal properties, one may consider what's common in all experiences, and if possible, consider micro phenomenal properties as having the "simplest" or "smallest" share of that common factor, just as one considers material objects as all having some size, shape, and motion, and then infer that micro corpuscles also have size, shape, and motion, only "smaller" or "simpler." However, we have little idea about what's in common in these varieties of phenomenal properties. The character of micro consciousness is hard to learn through transdiction, whether one considers a phenomenal macro unit as an entire conscious head/brain or as a part of it that produces certain token experiences.

¹²⁹ For a more exhaustive typology of phenomenology, see Kriegel 2015.

How about (b) and (c) in analogy (ii)? It seems that considering macro phenomenal units in terms of token experiences offers more to say about (b) and (c) than considering macro phenomenal units as conscious heads.

(b) If the phenomenal properties of macro units are token experiences, it is possible to recognize relations among the units that are *phenomenal* in the sense that the token experiences of these units are *phenomenally related*. **Phenomenal Space.** Perhaps the token experiences have different apparent locations in a “phenomenal space.” For example, I may experience a patch of yellow shape as appearing in the left side of my visual field, and a patch of blue as in the right side, etc. Thus if my entire consciousness is a space, different token experiences may occupy different (apparent) locations in this space, and thus having “phenomenally spatial” relations amongst them. Also, token experiences of different modalities may show up together, sometimes without clear relations in the phenomenal space. For example, I might see a patch of whiteness in front of me and hear a melody playing, without clear idea where the sound comes from. Or I may be overwhelmed by a mood, in which case the entire phenomenal space seems “immersed” in this mood.

Phenomenal Blending. Also, these token experiences may appear to influence the character of each other and even induce new qualities, a phenomenon which can be called “phenomenal blending” (Roelofs 2014). We know from experience that token experiences of the same sensory modality could have this effect. Seeing a color individually is different from seeing it with other colors. Hearing only a single sound differs from discerning it in a harmony. There is also evidence of phenomenal blending

from multimodal integration where experiences of different modalities interact. The experience of flavor, which is determined by multiple modalities such as taste and smell, is a good example.¹³⁰

Therefore, if one considers macro phenomenal units as the structures that produce token experiences, then besides physical relations among these units, we can also speak of the more distinctively phenomenal relations such that if the macro units are in such relations, then their token experiences are phenomenally related in ways we can observe in our experience. Such observable phenomenal relations among macro units are unavailable if we consider macro units as conscious heads/brains.¹³¹

(c) The attraction of considering macro units as token experiences is more obvious when thinking about the phenomenal properties of a system of such macro phenomenal units. We have seen that if macro units are conscious heads/brains (as in analogy (i)), there is little clue on whether a system of conscious heads has consciousness or what phenomenal properties would be in this system-level consciousness. In contrast, if we take macro phenomenal units as that which produce token experiences (as in analogy (ii)) rather than entire conscious heads, then it's easy to tell if such system-level consciousness may exist or what it is like. A system of macro units, in which each unit produces a token experience, could just be a brain with an individual, personal consciousness that consists

¹³⁰ Concerning the experience of flavor, see Auvray & Spence (2008). For other examples of multimodal integration, see Alais & Burr (2004), McGurk & MacDonald (1976), and Shams et al. (2000). These three additional examples all involve audio-visual integration.

¹³¹ This could be an attraction of Miller's (2017) strategy of considering Goff's "phenomenal bonding relation" as "co-consciousness", a relation that unifies token experiences into a conjoined experience (Ch. 2, section 2.2).

of all token experiences of the macro units. Thus we have a clear candidate for the system-level consciousness in this macro-level story of mental combination, a system-level consciousness which we actually encounter in experience.

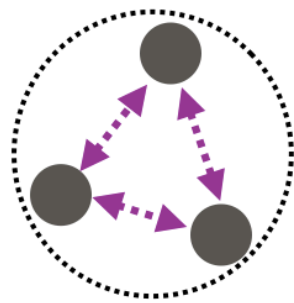
Thus, if in developing the micro/macro analogy we consider the macro phenomenal units (to be combined) as that which produce token experiences, we seem to have a complete macro-level side of the analogy, whose elements (a-c), including the phenomenal relations and the system-level phenomenal properties, are all more or less available for observation in experience.

But how far can a panpsychist proceed with this analogy? Earlier I have argued that this analogy cannot help with (a) the transduction from macro unit consciousness to micro unit consciousness. In what follows I argue that there is a further difficulty with the extension of the macro mental combination in this analogy to an account of micro-level mental combination. I will focus on (b) the relations among phenomenal units. The question is, Can the relations among macro phenomenal units in this analogy be extended to the micro units?

5.4.2.2 Reconsider (b) the relations among macro units in (ii)

There is an issue with the approach of understanding macro phenomenal units in terms of token experiences, which may undercut the advantage of having observable phenomenal relations among token experiences. Recall that, the analogy requires that the macro units are related in a way similar to how the micro units are related. That is, the macro-level relations and the micro-level relations should be similar. However, the

macro-level relations among token experiences (which belong to part **(b)** of analogy (ii)) seems unlike the micro-level relations among the micro phenomenal units. How so? Because, as I will argue in this section, the observable macro-level relations among token experiences are in the “*phenomenal space*,” whereas the micro-level relations among micro conscious entities are relations in a *physical space* (Figure 7).



Items in a Physical Space



Items in a Phenomenal Space

Figure 7. Physical Space vs. Phenomenal Space. The micro-level relations are in a physical space. The macro-level relations among token experiences are in a phenomenal space. In this figure (and later figures), relations in a physical space are shown in purple arrows, while relations in a phenomenal space are shown in light-green arrows.

Phenomenal space, as I have mentioned earlier, is a space in which different token experiences can occupy apparent locations. But we can understand it more generally as a realm in which different phenomenal items coexist or cohabit. Thus, although it's unclear *where* my experience of mood appears to be with respect to my experience of seeing a white patch, my mood experience and seeing-white-patch experience are in the same phenomenal space. But suppose that at the same time you are having an experience of hearing birds chirping. Your experience is not in the same phenomenal space as my experiences. **Physical space**, then, can be understood as a realm where different physical items cohabit. Thus our heads and brains are in one physical space.

Now that we understand phenomenal space and physical space, I will argue that the macro-level side of the analogy *requires*

the macro phenomenal units being such that when they exist and relate *in physical space* in a certain way, there are (macro) phenomenal systems at a larger scale.

This is because, as I will argue in this section, the desired micro-level side of the analogy, *to begin with*, should be

the micro phenomenal units being such that when they exist and relate *in physical space* in a certain way, there are phenomenal systems at a larger scale.¹³²

¹³² Ideally, the systems should be macro-level consciousness in creatures.

Why should the micro-level relation of the micro phenomenal units be *initially considered* as “in physical space” but not “in phenomenal space”? In the first place, we should not assume that the micro units are related in a phenomenal space, since we cannot assume that they are in the same phenomenal space (I will return to this point later.) Second, given our current knowledge and technique, humans will most likely have a better grasp of micro-level physical relations (such as spatial relations) than any phenomenal relations among the token experiences of a micro system (if such micro phenomenal relations exist). Therefore, given that the transdiction requires an analogy between a system of related micro units and a system of similarly-related macro units, and because we have a better prospect of knowing the micro (and macro) physical-spatial relations, our best clue to begin to identify similar systems at the micro and macro levels is the similarity in micro and macro *physical* relations. If a number of quarks with micro consciousness are considered a system, we don’t know about the phenomenal relations among the phenomenal properties of these quarks, but we may characterize the micro physical/spatial relations among them. Following the logic of transdiction, we should expect to infer the (phenomenal) properties of this micro system from a similar macro system. To determine if a macro-level system is similar to this micro system (which consists of quarks as micro units), we should tell if the units of the macro system are related in a way similar to the quarks in the micro system, whose relations can be known in physical/spatial terms. This is analogous to how transdiction works in physical science: to learn about a micro-level system, schematically characterizable as

micro physical units, *so-and-so* related in physical space, make a system of

One uses a transdiction from an observable macro-level account, schematically characterizable as

macro physical units, *so-and-so* related in physical space, make a system of *so-and-so system-level property*

Thus, the analogy needs macro-level units as characterized as related in physical space.

However, if the relations among macro phenomenal units are understood as the relations among token experiences, as described in (b), then the relations among macro units seem to exist in a phenomenal space. For consider the experience of yellow patch, which appears to be to the left of the experience of a blue patch. The apparent spatial relation, as it appears, exists in a phenomenal space. So is the simultaneous presence of an experience of a sound and a mood. All these relations among the token experiences presuppose that these token experiences exist in one phenomenal space. For if they are in different phenomenal spaces (e.g., the experience of yellow patch is mine while the experience of mood is yours), then they are not in any phenomenal relation.

Thus, the transdiction needs the micro units characterized as related in a physical space, and consequently needs the macro-level units characterized as related in a physical

space. But if macro phenomenal units are marked by their token experiences, whose relations are only characterized as in a phenomenal space, then we lack the required macro-level relations for the transdiction to proceed. Here is a “mismatch” between the sort of macro-level relations that are required for the analogy, and the sort of macro-level relations that are available.¹³³

¹³³ Thus there might be a confusion in Miller’s (2017) claim that co-consciousness is a relation that relates micro-subjects (Ch. 2 section 2.2). Co-consciousness is a relation among *experiences* of subjects, not exactly a relation among *subjects* of experiences (especially if we consider a subject as an entity that bears the experience, such as a brain, or a quark in a panpsychist world.) Concerning micro-subjects, such as conscious quarks, we may observe or understand their (micro)physical relations. But it’s hardly known just from the existence of micro-subjects that their experiences are in the phenomenal relation of co-consciousness. Nor does the (micro)physical relations straightforwardly entail the phenomenal relation.

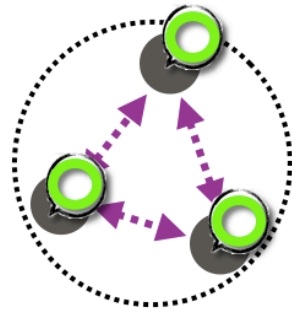
Miller sometimes seems to recognize the distinction between experiences and their subjects, for he says,

[T]he phenomenal bonding panpsychist should aim to show that the co-consciousness relation that holds within a subject’s conscious field can hold between the conscious fields of distinct subjects. (p. 554)

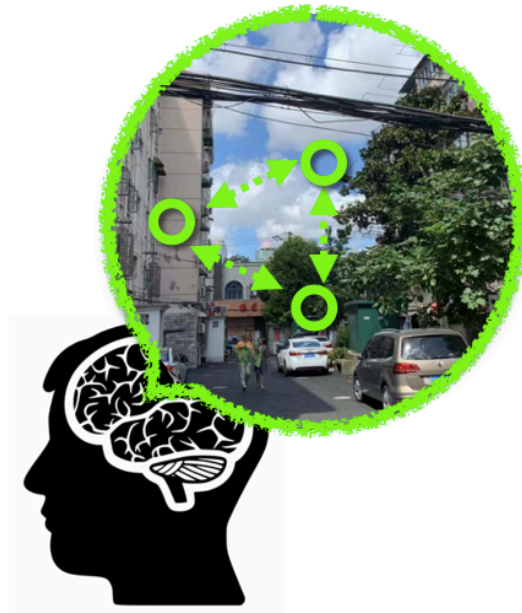
In this statement, Miller considers that co-consciousness may “hold between the conscious fields of distinct subjects”. He doesn’t simply say that co-consciousness “hold between distinct subjects”, which may indicate an awareness of a difference between subjects and their “conscious fields.” However, mostly Miller seems to hold that co-consciousness relation can hold among experiences *and* subjects:

[W]e can say that phenomenal relations are relations that hold *between subjects and/or experiences*. In other words, ‘*conscious things*’ like *subjects or experiences are the relata of phenomenal relations*. (Ibid., pp. 547-48; my emphasis)

Once *each member of the set of subjects and their experiences becomes bonded by co-consciousness*, there exists an experience that corresponds to the set and the co-consciousness relations between them ... (Ibid., p. 557; my emphasis)



System of Micro Individuals



System of Macro Individuals

Figure 8. Issue with Analogy (ii), (b). A mismatch between the micro-level relations and the macro-level relations in the analogy. In analogy (ii), the micro-level relations are in physical space (shown in purple arrows), whereas the macro-level relations are in phenomenal space (shown in light-green arrows).

If this mismatch (as shown in Figure 8) cannot be fixed, then there is no guarantee that the sort of phenomenal relations among token experiences we can observe in our consciousness can be inferred to exist in a system of micro units. For, if the phenomenal relations among token experiences are to exist in a system of micro units (most directly, as phenomenal relations among the phenomenal properties of the micro units), then the phenomenal properties of the micro units should be in a common phenomenal space. Otherwise, the micro phenomenal units would simply have no phenomenal relations. But how to determine that a system of micro units has a common phenomenal space for the units' phenomenal properties? We cannot observe or simply assume that any system of micro units has such a phenomenal space. Thus, a step of transdiction is needed: we determine if a system of micro units has such a phenomenal space by asking if a *similar* macro-level system has it, i.e., if a similar macro system is such that the phenomenal properties of its macro units are in a common phenomenal space—whether these token experiences are in the same consciousness. But how to determine that a macro-level system is similar to a chosen micro-level system? We need to know the organization of the micro-level system and select a similarly-organized macro-level system. Thus, we need some characterization of the relations among the micro-level units. But, as I've mentioned, we don't have initial idea of the phenomenal relations among the micro-level units or if such relations exist among them. The only available characterization of such micro-level relations are micro physical/spatial relations, understood as relations in a physical space. Therefore, to find a similarly-organized macro-level system is to identify

a system whose macro units are (similarly) related in a way characterized as such, as (similar) relations in a physical space.

Therefore, if the current analogy is only able to offer relations characterized as in a phenomenal space but not relations in a physical space, then we cannot tell if a micro system can be analogous to any macro system that has a common phenomenal space for the phenomenal properties of its macro units. Hence there is no way to determine if a micro system also has a common phenomenal space for the micro phenomenal properties of its units.

Compare with transdiction in physical science. This damage of mismatch shows the complexity in the attempt to understand micro-level consciousness via transdiction, in contrast with studies of microphysical properties. In physical science, it's plausible to assume only one physical space in which every entity, whether at the macro or the micro level, exists somewhere in this physical space and has spatial relations to other entities. Thus spatial relations in a physical space is basic and ubiquitous. If macro physical units are in certain spatial relations in a physical space, micro physical units are also in spatial relations in the same physical space. Transdiction of a system-level physical property with a micro/macro analogy is therefore possible. With panpsychism, however, the foregoing reflection shows that phenomenal space and phenomenal relations in a phenomenal space don't seem ubiquitous. That is an important aspect in which a phenomenal space seems different from the physical space. Unless a common phenomenal space for all phenomenal existence is the default, one cannot say that since some token experiences are known to be so-and-so related in a phenomenal space in a

macro-level system, the phenomenal properties of micro units in a system should also be related likewise in a phenomenal space. No. Without independent ways to tell the similarity between the macro-level and the micro-level systems, nothing guarantees that there is a phenomenal space in which multiple token phenomenal states of micro units cohabit.

Is it possible to fix the mismatch? Two fixes might be available.

Fix 1. One might insist that while micro phenomenal units in a system are known to exist in a physical space, their token phenomenal properties are also in a phenomenal space that allows them to relate phenomenally, *because all token phenomenal properties are in one phenomenal space*. Thus phenomenal relations among token experiences exist by default.

The problem with this approach is that such posit of a common phenomenal space for all experiences seems merely speculative. No one experiences others' experiences. For example, if I am in pain at the bottom of a swimming pool, you will not feel it when you are enjoying a landscape on a mountain top. One may insist that there is a cosmic consciousness that encompasses all experiences, including ours. But I seriously doubt if anyone can confirm the existence of a cosmic consciousness or characterize it. And, we need analogies based on the macro-level facts *that we can know and not just speculatively conceive* in order to form the non-speculative concepts to couch our theory of the micro-level facts. The difficulty here is analogous to the earlier difficulty with making a larger conscious system from the combination of multiple conscious heads/brain, albeit in a more extreme form. The consciousness in a system of conscious heads may be beyond

empirical confirmation, but positing such system-level consciousness allows setting boundaries for how “big” consciousness could be. The present suggestion of a universally encompassing consciousness, in an all-encompassing phenomenal space, abandons the boundaries of consciousness.

Fix 2. The other, less extreme fix, is to insist that while micro phenomenal units in a system are known to exist in a physical space, their token phenomenal properties are also in a phenomenal space, *because this micro-level system is similar, in the right way, to a macro-level system that has a phenomenal space of its system*. Exactly what the “right way” of similarity is is the target of study. But, at least, if the method of inferring a system-level phenomenal space in a micro-level system is based on the similarity between this micro-level system and a macro-level system, some reason can be given for attributing a phenomenal space to this particular micro-level system, without allowing all systems of phenomenal units to have a phenomenal space for all units.

How to tell that a micro-level system is similar to a macro-level system? As I have argued, we cannot rely on characterizations of phenomenal relations among token experiences, for we have no idea of phenomenal relations among the phenomenal properties of micro units, and we shouldn’t presume that such relations exist. We need to use those relations among micro units that are characterizable as relations in physical space, and those among macro units that are characterizable as in physical space. We need macro-level physical relations, such as the physical spatial relations among the physical realizers of the individual phenomenal experiences (e.g., the physical relations among parts of brain, as shown in Figure 9), not the phenomenal relations among token

experiences. The macro physical relations are no different from the paradigmatic physical relations such as forces and spatial relations.

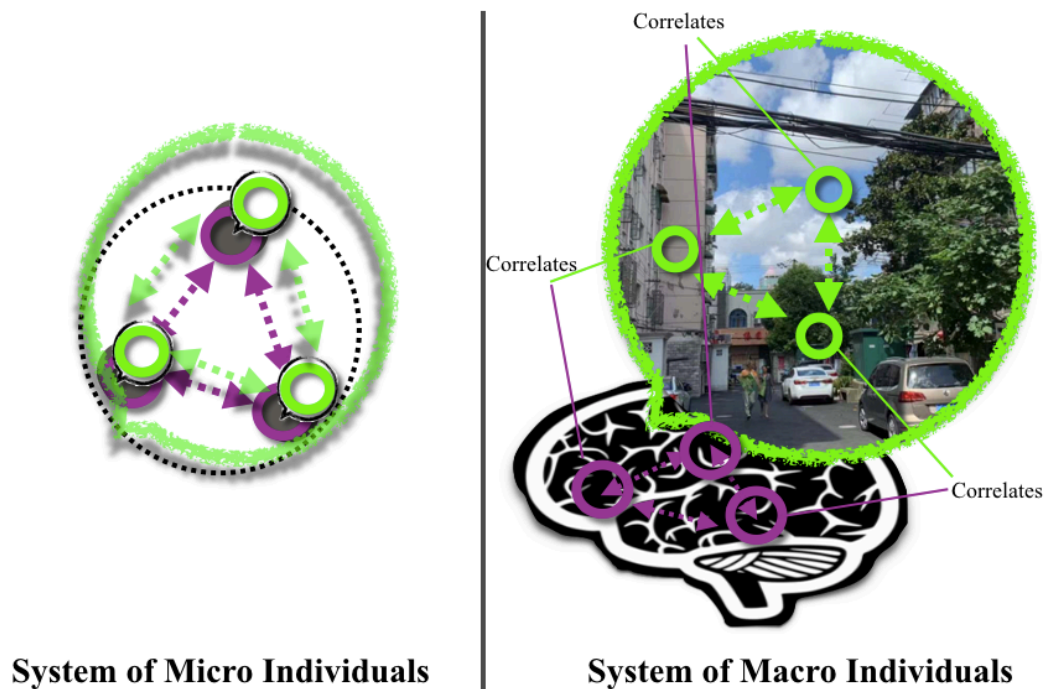


Figure 9. Analogy (ii), Fix 2 to the “mismatch” issue with (b). To infer that a micro-level system of conscious individuals have a system-level consciousness, one may seek analogy with a macro-level system of macro-level conscious individuals which is known to have a system-level consciousness. To establish the similarity between the two systems without prior knowledge of phenomenal relations at the micro level, the similarity between the relations within the two systems should be determined based on the physical relations. The physical relations in the micro and the macro systems (all marked in purple arrows) are, respectively, the microphysical relations among the ultimates, and the physical relations among the brain parts (neural correlates of token experiences). If enough physical similarity can be identified between the micro- and the macro-level systems, one may infer the phenomenal relations and the system-level consciousness in the micro system (marked by the light-green arrows and large circle with shadows).

But, to find these relations, we need to first figure out what they relate, i.e., what the macro-level units that bear the token experiences are.

Recall that for the micro/macro analogy to be effective, the macro-level account and the micro-level account should be sufficiently similar. Now notice that the micro-level units in panpsychism are assumed to be micro-subjects, i.e., each micro unit is supposed to have its own consciousness as its intrinsic property, independent of other micro units. Thus, the micro-level relations are supposed to relate *conscious* units to make a conscious system. Therefore, to find a macro-level story that serves as an analogy, the macro-level units that model the micro units should also have consciousness as an intrinsic property of each of them, independent of other macro-level units.

The difficulty is that we might not find such macro-level units if, as in the current analogy, we consider macro units as that which bear the token experiences in our consciousness. Presumably if macro phenomenal units are characterizable in terms of token experiences, then macro units should be parts of brain. Thus the analogy requires that portions of brain be able to have their own consciousness. However, current sciences of consciousness mostly indicate that the consciousness of a brain isn't produced by "stitching" together token experiences that exist in bits-and-pieces of a brain. There is evidence suggesting that *consciousness* (not merely *consciousness of a particular content*), with its complex arrangement of token experiences, results from complex connections across brain regions.¹³⁴ Also, there is evidence that the neural activities that make a creature's consciousness possible require some common structures, such as

¹³⁴ See Dehaene & Naccache (2001), Dehaene & Changeux (2011).

reticular formation in brainstem and connections between thalamus and cerebral cortex.¹³⁵ Thus it is possible that the current analogy cannot proceed because we cannot find macro units that can model the micro units, since perhaps no macro units in the form of a brain part has its own consciousness.

In response, one might argue that we can obtain consciousness with only parts of the brain. Perhaps each cerebral hemisphere of a brain has its own consciousness, so that we may derive two streams of consciousness if we separate them. This seems to be an example of brain parts with their own consciousness.¹³⁶ For another example, one might imagine individuals with severe brain damage, with the result that their consciousness are severely impoverished. One individual might only have visual experience, another only auditory experience, yet another only somatosensory experiences. This also seems to be an example of brain parts with their own consciousness, if in each case the brain lesion leaves only parts of the brain intact.

However, these examples do not present macro-level brain parts that can be conscious while being totally independent. The consciousness of two hemispheres depends on a common structure to which both hemispheres must be connected, e.g., the brainstem and the thalamus. Similarly for the impoverished consciousness of the unfortunate individuals. Since their visual consciousness, auditory consciousness, and

¹³⁵ On the role of brain stem for consciousness, see Parvizi & Damasio (2001), Parvizi & Damasio (2003). On the role of thalamocortical connection, see Tononi & Edelman (1998), Alkire et al. (2000), Alkire & Miller (2005), and White & Alkire (2003).

¹³⁶ Miller (2017) has suggested a similar case of “the co-consciousness relation holding between the conscious fields of different subjects” with a split-brain scenario, in which each hemisphere is considered a conscious subject (p. 554).

somatosensory consciousness all depend on some common structures which are vital for consciousness to exist, it is not the case that normal consciousness is due to the token experiences in the consciousness of multiple independent, non-overlapping brain parts. Therefore, in the case of human consciousness and its physical substrate, we might not find the sort of macro-level units that can feed into an analogy to the micro units in panpsychism.

Thus, although we may have knowledge of **(b)** the phenomenal relations in token experiences at the macro level, we cannot easily infer by analogy that the phenomenal properties in micro-level units bear similar phenomenal relations. Phenomenal relations presuppose a phenomenal space in which the related phenomenal properties cohabit, and such a phenomenal space may be lacking in a micro-level system. If one chooses not to assume a universal phenomenal space that includes all existing phenomenal properties, to establish the existence of phenomenal relations among the phenomenal properties of micro units in a system, one has to show that this micro-level system is sufficiently similar to a macro-level system that is known to have a phenomenal space that contains the phenomenal properties of its macro units. However, current knowledge about the physical substrate of macro consciousness doesn't support the existence of such a macro-level system in analogy (ii), since it's hard to see how the physical system that has a system of token experiences (i.e., the conscious brain) is a system of macro units (brain parts) such that each macro unit has consciousness independently of other units, as micro units in panpsychism are supposed to do.

An advantage of analogy (i) over analogy (ii). Here one may find an advantage of analogy (i) over analogy (ii). In analogy (i), which considers macro units as simply conscious heads/brains, macro units are known to independently have consciousness, whereas in analogy (ii) which considers macro units as that which have token experiences, the existence of macro units that have consciousness independently is far from being guaranteed. For this reason, it is easier to see what could model (a) the micro units and their systems in analogy (i) than in analogy (ii), even though analogy (ii) provides relatively straightforward examples of (b) phenomenal relations and (c) system-level phenomenal properties.

(c) The difficulty of transduction of (b) phenomenal relations in the macro-level token experiences to the micro-level system also weakens the use of knowing (c) the system-level phenomenal properties in macro consciousness. Admittedly, unlike analogy (i), analogy (ii) has a macro-level story that gives an idea of system-level phenomenal properties in the form of our personal consciousness, which we can know. However, if we cannot establish that in a micro-level system, the micro units and their relations are sufficiently similar to the units and relations in a macro-level system that has a certain form of system-level consciousness, then we cannot infer by analogy that the micro-level system has a similar system-level consciousness, or that it has system-level phenomenal properties at all. Therefore, since in analogy (ii) we cannot establish that in a micro-level system, the micro units and their relations are sufficiently similar to the units and relations in a macro-level conscious system, we cannot infer that the micro-level system has a world of consciousness similar in some way to the one we have.

5.4.2.3 Further reflection on (b) of analogy (ii)

Thus, in analogy (ii), the major difficulty with extending phenomenal relations in the macro level to the micro level is the possible lack of a phenomenal space of a micro-level system, or more precisely, our lack of means to rule out this possibility. But there are more difficulties with the transdiction of (b) phenomenal relations among token experiences, which has to do with the difficulty in the transdiction of (a) the phenomenal properties of the units.

Difficulty with phenomenal space: What sort of space? In the first place, if a micro-level system does have a phenomenal space, what sort of phenomenal space is it? What could be the “phenomenal spatial relations” among the token experiences of the micro units? Perhaps it’s like co-presence of color patches, if micro phenomenal properties are color-like. But what if the micro phenomenal properties resemble sounds? Or, what if the most primitive experiences are simply pain and pleasure? Apparently these experiences arrange in different ways. The experiences of color patches, sounds, and pain/pleasure seem to occupy the phenomenal space in different ways.¹³⁷ If we need accounts of the making of phenomenal space in a system of micro units, we need to figure out what sort of phenomenal spatial relations or existence is in the system.

¹³⁷ By “phenomenal space,” I do *not* mean “quality spaces” in which one can quantify different qualities and find some dimensions in them. (For an example of “quality spaces,” see Clark 1993.) The “quality spaces” are theoretical representations of qualities in a systematic way, not really a “space” where things happen (or appears to happen). By “phenomenal space” I mean a perspective of experiences, the “Cartesian theater” where experiences show up.

However, because (as I have argued) we have no clue about what micro phenomenal properties are like, whether they are like colors or sounds or pain/pleasure, we also have no clue about what sort of phenomenal spatial relations would micro experiences be in.

Difficulty on phenomenal blending: what sort of blending? There are also questions about phenomenal blending at the micro level, i.e., the modification of phenomenal properties when multiple phenomenal properties are co-present in the system-level phenomenal space of a micro system. The issue is that, apparently there are many sorts of phenomenal blending. Phenomenal blending can happen among experiences within the same modality, and we cannot assume that phenomenal blending in different modalities happen in the same way. Phenomenal blending also happens across different modalities, which creates more types of phenomenal blending. And, again, because (as I have argued) we have no clue about what micro phenomenal properties are like, we have no clue about what sort of phenomenal properties are “blended”, and therefore no clue about the sort of phenomenal blending that operates in the combination of micro experiences. To put it more vividly, since we don’t know if micro phenomenal properties are like sounds or tastes/smells or other qualities, we don’t know if their phenomenal blending (if phenomenal blending happens at all among micro consciousness) generates something like harmonies or flavors.

One may try to address the ignorance about **(a)** the phenomenal character of micro consciousness and the ensuing ignorance about **(b)** the character of the micro-level phenomenal relations by looking for the *common factors* among various kinds of experiences or various sorts of phenomenal relations, including the apparently different

sorts of phenomenally-spatial relations and phenomenal blending, within the same modality or across different modalities. Here, again, we are mired by the difficulty of discerning the commonalities across different sorts of experiences, esp. experiences of different modalities.

Therefore, even if we are somehow justified to infer that a micro-level system has a phenomenal space that includes the phenomenal properties of its micro units, due to the difficulty with the transdiction of **(a)** phenomenal properties of the units at the micro level, it's also difficult to tell **(b)** what phenomenal relations could hold among the micro-level phenomenal properties.

Let me summarize the examination of analogy (ii), in which one may propose an analogy for micro-level mental combination in terms of the combination of macro units, where the macro units are considered as the things that bear token experiences. Analogy (ii) may offer a clearer idea of **(b)** the phenomenal relations and **(c)** the phenomenal properties of a system of units at the macro level, given our acquaintance with the relations between token experiences in our consciousness and our knowledge of our consciousness as a system of token experiences. However, although analogy (ii) has a more informative macro-level story, the macro-level story of this analogy cannot be easily extended to the micro level. **(a)** In the first place, like in the case of analogy (i), considering the phenomenal properties of macro units as token experiences suggests little about the phenomenal character of micro consciousness. **(b)** But more importantly, there is little guarantee that the phenomenal relations among token experiences that one

observes in macro consciousness can be inferred to exist among the micro experiences in the micro units.

5.5 Concluding Remarks

To conclude, one primary cause of the difficulty of developing panpsychism is the difficulty in constructing a reasoning of transduction to learn about the micro-level phenomenal through the macro-level phenomenal facts, which is because of the difficulty in formulating an informative analogy between the micro and the macro levels.

Several contributing factors to the difficulty can be observed.

- First, it's unclear how to attribute phenomenal properties to things outside of the level of creatures/organisms, whether to groups of creatures, to parts of a creature (or its brain), or to systems of micro particles. We have little idea what is to empirically confirm a hypothesis of this sort.
- Second, related to the first problem, we don't have an established concept of a phenomenal space in which any two tokens of phenomenal properties of two distinct entities can be unproblematically related. (Unlike, in physics, for any intrinsic properties of two distinct entities, it's possible to relate them in the same physical space.)

- Thirdly, we have no idea of the common factors/dimensions among all phenomenal properties which may inform us about the candidate phenomenal properties of the micro units. (Unlike, in physics, we have managed to identify some commonalities among the properties of all objects, such as mass, size/spatial distance, charge, and motion.)

How strong is this criticism to pan(proto)psychism? Obviously I am not refuting the metaphysical possibilities of various forms of panpsychism. Panpsychists are free to use their metaphysicians' conscience to argue that some sort of beings exist for certain metaphysical reasons, without feeling bad about not being able to know or explore more closely the things that they believe in. Neither am I knocking down completely the epistemic future of pan(proto)psychism. Maybe I have, in my critique, pointed to something that pan(proto)psychists should focus on, or illustrated some approach that pan(proto)psychists should avoid. Maybe, to push forward the transdiction for pan(proto)psychism, someone will find the common factors among all phenomenal properties that we know of and systematize them into a form similar to the Periodic table of chemical elements; or, someone will develop an empirically respectable method of attributing phenomenal properties beyond the creatures or determining the existence of phenomenal spaces outside of brain or neural structures. If so, that's fine.

Nonetheless, I want to make clear that these epistemic challenges to panpsychism are real. The need to answer them should be acknowledged by everyone who contemplates on panpsychism. After all, theories of metaphysics can always be offered

like sketches of grand castles. However, if the authors of these theories are perpetually unable to specify the elements in their theories, then what they are doing is similar to offering sketches of castles that never receive an engineer's blueprint. Thus, we will never see a real castle based on these sketches. Nor will anyone be able to live in such castles.

One theme of this dissertation is that pan(proto)psychists cannot account for micro (proto)phenomenal properties and how these properties combine to form macro phenomenal properties. Regarding this ignorance, Coleman says:

What are the rules of phenomenal combination? That is, what is the logic that takes qualitative instances x and y and produces qualitatively-different instance z, for example? I simply don't know. But I do not need to know. My present task [of arguing for panqualityism] is only to show how phenomenal combination might be possible. The combinatory rulebook is something that remains to be investigated. (Coleman 2012, p. 158, fn.37)

Also, recall that (section 2.1, this chapter) regarding the ignorance of micro phenomenal properties, Seager says

The main point is that while it seems hopelessly mysterious to point that phenomenal experience is generated by a system of entities possessed only of physical properties such as mass, spin, charge, energy and a few more abstruse quantum mechanical properties, it does not seem similarly paradoxical that a system of entities possessed of elementary phenomenal features could generate complex states of consciousness... (2017, p. 240)

In my experience with the literature of panpsychism, I find this attitude prevalent: people admit that they have little idea about micro consciousness, but most of them don't seem to

regard it as an urgent task to figure out what micro phenomenal properties are like or to figure out a path to identify such micro properties. Their confidence in the power of panpsychism to explain consciousness relies largely on the fact that they can have something “phenomenal” that shows up in their explanans of consciousness (even though they cannot specify *what phenomenal thing or property* that is), while physicalists have nothing “phenomenal” to appeal to in their explanations. It’s as if using some merely “phenomenal” notion is enough to treat the uneasiness that one might feel with an “all-physical” conception of the world, despite the lack of determinate ideas in this “phenomenal” notion.

This *prima facie* satisfaction with appealing to the unspecified “phenomenal” is not limited to panpsychism. A dualist might insist that there are fundamentally nonphysical, phenomenal properties, which would instantiate with the neural activity of intelligent creatures. Dualism also opens up many questions about what the fundamental phenomenal properties are and what the physical-phenomenal laws are. These questions are not easy. Answering these questions about the specifics of phenomenal properties might take more work than articulating a metaphysics of dualism. Therefore, dualists should feel some urgency to answer these questions. However, the philosophical literature on dualism doesn’t seem to manifest enough sense of urgency toward this question. The focus of dualists seems to be arguing that something phenomenal is necessary for explaining consciousness; besides, little has been said to describe the fundamental phenomenal properties and physical-phenomenal laws in dualism.

Thus, dualists and panpsychists seem largely complacent with their metaphysical positions that contain something phenomenal in its fundamental ontology, so that they can expect that their metaphysics may offer something phenomenal at the creature level. This apparent complacency in dualism and panpsychism without strong motivation to specify their accounts of (micro-level) phenomenal properties is something worth discussion, although it seems hardly noticed. In the rest of this section I will speculate on why there is this complacency.

One explanation may be that dualism and panpsychism are at their early stages where theorizing has to leave out many details for further study. However, this should only justify a strong sense of urgency toward the need to theorize about specific (micro-level) phenomenal properties. Also it doesn't explain why the question of what specific (micro) phenomenal properties are is much less discussed than the puzzling relationship between consciousness and the physical world.

Another explanation is that philosophers, especially metaphysicians who focus on the most general ontological categories, tend to focus on characterizing the categories in general terms but not the specific properties in these categories. Thus dualists or panpsychists who posit (micro)phenomenal properties as a nonphysical category of being may find questions about specific (micro)phenomenal properties less interesting. Perhaps this observation of a professional complacency is right, but this doesn't mean that questions about specific (micro)phenomenal properties are not important questions, that we philosophers should be glad to stay silent or plead ignorance on such matters and get

away with it. Also it doesn't explain this complacency. We need an explanation of why dualists or panpsychists can be comfortable in this state of professional complacency.

The best explanation, I think, is that about the relationship between consciousness and the physical, what people *really care about* has only been the apparent distinctness between consciousness and physical phenomena. The central motivation for philosophers to argue about consciousness is to come to terms with the intuition that consciousness doesn't seem to be a kind of physical things or processes. Therefore, even though it is a question for dualism/panpsychism what the fundamental (micro)phenomenal properties are, this question doesn't seem bothersome, because this question about the specifics of (micro)phenomenal properties doesn't seem to have a direct bearing on why consciousness seems different from physical things or processes. To come to terms with the apparent difference between consciousness and the physical in dualism/panpsychism, some abstract and unspecified notion of the "phenomenal" seems sufficient.¹³⁸

Thus, consider a schematic dualist/panpsychist account of consciousness, in a form like the following:

¹³⁸ People do think about phenomenal qualities in general terms or without sufficient determination of particular qualities. For example, one might talk about "the experience of color" generally and distinguish the experience of color (spoken generally) from the experience of pitch (also spoken generally). Also one might talk about "the experience of smelling skunk" without knowing specifically what it's like (not even knowing that it's unpleasant). Such talking of phenomenal qualities without knowing or using the specific qualities doesn't prevent one from understanding such terms as about phenomenal qualities in consciousness. I have briefly explained this way of thinking about phenomenal properties in Liu (2020b).

(micro)Phenomenal fact1, (micro)Phenomenal fact 2, ... (Auxiliary conditions) |=
Phenomenal fact (e.g., my current experience of *this white color*)

It is right that fully specifying this schematic account requires answering questions of what the (micro)Phenomenal facts are and how (micro)phenomenal properties work. However, just to mention these (micro)Phenomenal facts without specifying them could seem satisfying. Since both sides of this schema are understood as about phenomenal items, one might feel that there is no strong discrepancy in supposing that the left side of the schema (“(micro)Phenomenal fact1, (micro)Phenomenal fact 2, ... (Auxiliary conditions)”) can explain the other. Now contrast this with the following physicalist schematic account of consciousness:

Physical fact1, Physical fact 2, ... (Auxiliary conditions) |= Phenomenal fact (e.g., my current experience of *this white color*)

In this physicalist schema without using any explanatory items in phenomenal terms, the left side (i.e., “Physical fact1, Physical fact 2, ... (Auxiliary conditions)”) is not conceptualized as phenomenal, whereas the other side (the explanandum, “Phenomenal fact”) is conceptualized as phenomenal. Thus, unlike the dualist/panpsychist schema, the physicalist schema leaves the impression that its explanan cannot account for the explanandum, because they seem to be about essentially different kinds of things.

Therefore, by merely using phenomenal terms without specification, a schematic dualist/panpsychist account may seem sufficient for a dualist/panpsychist who wants to feel that her account is *about consciousness*. Once this bit of satisfaction is achieved, it seems to her that enough justice has been done to the core intuition that consciousness is not some physical process, and the central motivation for philosophizing about consciousness is satisfied. Hence little further motivation to investigate the phenomenal world of dualism/panpsychism that is posited to exist and to be studied.

BIBLIOGRAPHY

- Aaronson, S. (2014) Why I Am Not An Integrated Information Theorist (Or, The Unconscious Expander), <http://www.scottaaronson.com/blog/?p=1799>
- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, 14(3), 257-262.
- Alkire, M. T., Haier, R. J., & Fallon, J. H. (2000). Toward a unified theory of narcosis: brain imaging evidence for a thalamocortical switch as the neurophysiologic basis of anesthetic-induced unconsciousness. *Consciousness and cognition*, 9(3), 370-386.
- Alkire, M. T., & Miller, J. (2005). General anesthesia and the neural correlates of consciousness. *Progress in brain research*, 150, 229-597.
- 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*. Oxford University Press.
- Aquinas. *On the Principles of Nature*. In Klima, G., Allhoff, F., & Vaidya, A. (Ed.) (2007). *Medieval philosophy: essential readings with commentary*. Blackwell Pub.
- Aquinas, *Summa Theologiae*, Part I, Questions 75-76. [Accessed on Feb. 13, 2018] <http://spot.colorado.edu/~pasnau/westview/st1a75-76.htm>
- Aquinas, *Summa Theologiae*, Part I, Questions 78-79, 84-86. [Accessed on Feb. 13, 2018] <http://spot.colorado.edu/~pasnau/westview/st1a78-7984-86.htm>
- Auvray, M. and C. Spence. (2008). The Multisensory Perception of Flavour. *Consciousness and Cognition* 17: 1016–31.
- Barnes, J. (1984). *The Complete Works of Aristotle: The Revised Oxford Translation*. Volume 1 & 2. Princeton University Press.
- Basile, P. (2010). It must be true—but how can it be? Some remarks on panpsychism and mental composition. *Royal Institute of Philosophy Supplement*, 67, 93-112.
- Blamauer, M. (2011). Taking the Hard Problem of Consciousness Seriously: Dualism, Panpsychism and the Origin of the Combination Problem. *The Mental as*

Fundamental: New Perspectives on Panpsychism, 99-116.

Blamauer, M. (2013). Panpsychism without Subjectivity? A Brief Commentary on Sam Coleman's 'Mental Chemistry' and 'The Real Combination Problem'. *International Journal of Philosophy*, 4(34), 299-309.

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

Bohm, D. (1980). *Wholeness and the Implicate Order*. Routledge.

Bohm, D., & Hiley, B. J. (2006). *The undivided universe: An ontological interpretation of quantum theory*. Routledge.

Boyle, R. (1744). "The History of Fluidity and Firmness". In Boyle, R. *The works of the Honourable Robert Boyle*. In five volumes. To which is prefixed the life of the author. Volume 1. London, MDCCXLIV. Eighteenth Century Collections Online. Gale. Univ of Massachusetts Amherst. [Retrieved on 25 Mar. 2019] http://find.galegroup.com.silk.library.umass.edu/ecco/infomark.do?&source=gale&prodId=ECCO&userGroupName=mhlin_w_umassamh&tabID=T001&docId=CW113168800&type=multipage&contentSet=ECCOArticles&version=1.0&docLevel=FASCIMILE

Boyle, R. (1744). "The Experimental History of Colours". In Boyle, R. *The works of the Honourable Robert Boyle*. In five volumes. To which is prefixed the life of the author. Volume 2. London, MDCCXLIV. Eighteenth Century Collections Online. Gale. Univ of Massachusetts Amherst. [Retrieved on 25 Mar. 2019] http://find.galegroup.com.silk.library.umass.edu/ecco/infomark.do?&source=gale&prodId=ECCO&userGroupName=mhlin_w_umassamh&tabID=T001&docId=CW113169546&type=multipage&contentSet=ECCOArticles&version=1.0&docLevel=FASCIMILE

Brock, W. H. (1993). *Norton history of chemistry*. WW Norton.

Brogaard, B. (2017). In Search of Mentons, in Bruntrup, G. & Jaskolla, L. (eds.) *Panpsychism: Contemporary Perspectives*, pp.130-152, New York: Oxford University Press.

Brower, J. E. (2014). *Aquinas's Ontology of the Material World: Change, Hylomorphism, and Material Objects*. Oxford University Press.

Brüntrup, G. (2017). Emergentist panpsychism. In Brüntrup, G. & Jaskolla, L. (Eds.). *Panpsychism: Contemporary Perspectives*. Oxford University Press.

- Brüntrup, G. & Jaskolla, L. (Eds.). (2017). *Panpsychism: Contemporary Perspectives*. Oxford University Press.
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of consciousness studies*, 2(3), 200-219.
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- Chalmers, D. (2003). Consciousness and its place in nature. In Stephen P. Stich & Ted A. Warfield (eds.), *Blackwell Guide to the Philosophy of Mind*. Blackwell. 102-142.
- Chalmers, D. (2006). Strong and weak emergence. In Clayton, P. & Davies, P. (eds.). *The Re-Emergence of Emergence*, pp. 244-256.
- Chalmers, D. (2007). Phenomenal concepts and the explanatory gap. In Alter, T. & Walter, S. (eds.). *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.
- Chalmers, D. (2009). The two-dimensional argument against materialism. In Chalmers, D., *The character of consciousness*, 141-191.
- Chalmers, D. (2015). Panpsychism and panprotopsyism. In Alter and Nagasawa (eds.), *Consciousness in the physical world: Perspectives on Russellian monism*. Oxford University Press. 246-76.
- Chalmers, D. (2017). The combination problem for panpsychism. In Brüntrup, G. & Jaskolla, L. (Eds.). *Panpsychism: Contemporary Perspectives*. Oxford University Press.
- Chalmers, D. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies*, 25(9-10), 6-61.
- Clark, A. (1993). *Sensory Qualities*. Oxford University Press.
- Clarke, D. S. (Ed.). (2004). *Panpsychism: Past and recent selected readings*. SUNY Press.
- Coleman, S. (2012). Mental Chemistry: Combination for Panpsychists. *Dialectica*, 66(1), 137-166.
- Coleman, S. (2014). The Real Combination Problem: Panpsychism, Micro-Subjects, and Emergence. *Erkenntnis* 79 (1), 19-44.

- Coleman, S. (2017). Panpsychism and Neutral Monism. In Brüntrup, G. & Jaskolla, L. (Eds.). *Panpsychism: Contemporary Perspectives*. Oxford University Press.
- Dasgupta, S. (2015). Inexpressible ignorance. *Philosophical Review*, 124(4), 441-480.
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, 79(1-2), 1-37.
- Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200-227.
- Descartes, R. (1642/1991), "Letter to Regius". In Cottingham, J., Murdoch, D., Stoothoff, R., & Kenny, A. (trans. & eds.). *The Philosophical Writings of Descartes: Volume 3, the Correspondence*. Cambridge University Press.
- Descartes, R. (1644/1985). *Principles of Philosophy*. In Cottingham, J., Stoothoff, R., & Murdoch, D. (trans. & eds.). *The philosophical Writings of Descartes: Volume 1*. Cambridge University Press.
- Descartes, R. (1664/1985), "The World or Treatise on Light". In Cottingham, J., Stoothoff, R., & Murdoch, D. (trans. & eds.). *The philosophical Writings of Descartes: Volume 1*. Cambridge University Press.
- Feigl, H. (1958). The 'mental' and the 'physical'. *Minnesota Studies in the Philosophy of Science* 2: 370-497.
- Feynman, R. P., Leighton, R. B., & Sands, M. (2013). The Feynman lectures on physics, Vol. I: mainly mechanics, radiation, and heat. Retrieved on March 6th, 2020, from http://www.feynmanlectures.caltech.edu/I_toc.html
- Fine, K. (1999). Things and their parts. *Midwest Studies in Philosophy*, 23(1), 61-74.
- Gardner, E. J. (1965). *History of biology*. Minneapolis: Burgess Pub. Co.
- Gassendi, P. (1624/1972). *Exercises against the Aristotelians*. In Brush, C. B. (trans. & eds.). *The selected works of Pierre Gassendi*. New York: Johnson Reprint Corp.
- Goff, P. (2006). Experiences don't sum. *Journal of Consciousness Studies* 13 (10-11): 53-61.
- Goff, P. (2009a). Why Panpsychism doesn't Help Us Explain Consciousness. *Dialectica*, 63(3), 289-311.

- Goff, P. (2009b). Can the Panpsychist get around the Combination Problem?. In Skrbina, D. (ed.). *Mind that Abides: Panpsychism in the New Millennium*, Amsterdam: Benjamins, 129-136.
- Goff, P. (2011). A posteriori physicalists get our phenomenal concepts wrong. *Australasian Journal of Philosophy*, 89(2), 191-209.
- Goff, P. (2017a). The phenomenal bonding solution to the combination problem. In Brüntrup, G. & Jaskolla, L. (Eds.). *Panpsychism: Contemporary Perspectives*. Oxford University Press.
- Goff, P. (2017b). *Consciousness and Fundamental Reality*, Oxford University Press.
- Haldane, J. (2000). The state and fate of contemporary philosophy of mind. *American Philosophical Quarterly*, 37(3), 301-311.
- Haldane, J. (1998). A Return to Form in the Philosophy of Mind. *Ratio*, 11(3), 253-277.
- Hartshorne, C. (1934). *The Philosophy and Psychology of Sensation*. Univ. Chicago Press.
- Hill, C. S., & McLaughlin, B. P. (1999). There are fewer things in reality than are dreamt of in Chalmers's philosophy. *Philosophy and Phenomenological Research*, 59(2), 445-454.
- Hobbes, T. (1651/1962), *Leviathan*. In Molesworth, W., (eds.). *The English works of Thomas Hobbes of Malmesbury*. Vol. III. Aalen [Ger.]: Scientia.
- Holman, E. (2008). Panpsychism, physicalism, neutral monism and the Russellian theory of mind. *Journal of Consciousness Studies*, 15(5), 48-67.
- Humphreys, P. (1997). How properties emerge. *Philosophy of science*, 64(1), 1-17.
- Jackson, F. (1986). What Mary didn't know. *The Journal of Philosophy*, 291-295.
- Jackson, F. (1998). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford University Press.
- James, W. (1891). *The Principles of Psychology*. London: Macmillan.
- Jaworski, W. (2011). *Philosophy of mind: A comprehensive introduction*. John Wiley & Sons.

- Jaworski, W. (2016). *Structure and the metaphysics of mind: How hylomorphism solves the mind-body problem*. Oxford University Press.
- Johnston, M. (2006). Hylomorphism. *The Journal of Philosophy*, 103(12), 652-698.
- Kim, J. (2011). *Philosophy of Mind (Third Edition)*. Boulder: Westview Press.
- Kind, A. (2015). Pessimism about Russellian monism. In Alter, T., & Nagasawa, Y. (Eds.). *Consciousness in the Physical World: Perspectives on Russellian Monism*, 401-21.
- Koons, R. (2014). Staunch vs. faint-hearted hylomorphism: Toward an Aristotelian account of composition. *Res Philosophica*, 91(2), 151-177.
- Koslicki, K. (2008). *The structure of objects*. Oxford University Press.
- Kriegel, U. (2003). Consciousness as sensory quality and as implicit self-awareness. *Phenomenology and the Cognitive Sciences*, 2(1), 1-26.
- Kriegel, U. (2004). Consciousness and self-consciousness. *The Monist*, 87:2 182-205.
- Kriegel, U. (2009). *Subjective Consciousness: A Self-Representational Theory*, New York: Oxford University Press.
- Kriegel, U. (2015). *Varieties of consciousness*. Oxford University Press.
- Kripke, S. (1980). *Naming and Necessity*. Harvard University Press.
- Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific philosophical quarterly*, 64(4), 354-361.
- Levine, J. (1993). On Leaving Out What It's Like. In Martin Davies & Glyn W. Humphreys (eds.), *Consciousness: Psychological and Philosophical Essays*. MIT Press. pp. 543-57.
- Levine, J. (2010). Phenomenal experience: A cartesian theater revival. *Philosophical Issues*, 20, 209-225.
- Levine, J. (2017). Godehard Bruntrup and Ludwig Jaskolla (eds.), *Panpsychism: Contemporary Perspectives*, Bruntrup, G and Jaskolla, L, eds. Notre Dame Philosophical Reviews 2017. <https://ndpr.nd.edu/news/panpsychism-contemporary-perspectives/>.

- Lewis, D. (1970). How to define theoretical terms. *The Journal of Philosophy*, 67(13), 427-46.
- Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50(3), 249-258.
- Lewis, D. (2009). Ramseyan humility. In David Braddon-Mitchell & Robert Nola (eds.), *Conceptual Analysis and Philosophical Naturalism*. MIT Press, 203-22.
- Lewtas, P. (2013). What it is like to be a quark. *Journal of Consciousness Studies*, 20(9-10), 39-64.
- Liu, H. (2020a). Of integrated information theory: a philosophical evaluation. *Philosophical Psychology*, 33(3), 442-468.
- Liu, H. (2020b). On Chalmers on the Meta-Problem. *Journal of Consciousness Studies*, 27(5-6), 91-98.
- Loar, B. (1990). Phenomenal states. *Philosophical perspectives*, 81-108.
- Locke, D. (2009). A partial defense of Ramseyan humility. In Braddon-Mitchell, D. & Nola, R. (eds.), *Conceptual Analysis and Philosophical Naturalism*. MIT Press. pp. 223-242.
- Lockwood, M. (1993). The grain problem. In Robinson, H. M. (ed.). *Objections to Physicalism*. Oxford University Press. pp. 271-91.
- Madden, J. (2013a). *Mind, Matter, and Nature*. Catholic University of America Press.
- Madden, J. (2013b). TH and Philosophy of Mind and Philosophy of Religion. *Philosophy Compass*, 8(7), 664-676.
- Mandelbaum, M. (1964). Newton and Boyle and the Problem of "Transdiction". In Mandelbaum, *Philosophy, Science, and Sense Perception*. The Johns Hopkins Press, Baltimore.
- Marmodoro, A. (2013). Aristotle's hylomorphism without reconditioning. *Philosophical Inquiry*, 37(1/2), 5-22.
- Maxwell, G. (1978). Rigid Designators and Mind-Brain Identity, *Minnesota Studies in the Philosophy of Science*, 9, 365-403.

- McGuire, J. E. (1970). Atoms and the 'analogy of nature': Newton's third rule of philosophizing. *Studies in History and Philosophy of Science Part A*, 1(1), 3-58.
- McGurk, H. and MacDonald, J. (1976). "Hearing lips and seeing voices," *Nature*, 264:746–748.
- McLaughlin, B. (2017). Mind Dust, Magic, or a Conceptual Gap Only?. In Brüntrup, G. & Jaskolla, L. (Eds.). *Panpsychism: Contemporary Perspectives*. Oxford University Press.
- Miller, G. (2017). Forming a Positive Concept of the Phenomenal Bonding Relation for Constitutive Panpsychism. *dialectica*, 71(4), 541-562.
- Montero, B. G. (2015). Russellian physicalism. In Alter, T., & Nagasawa, Y. (Eds.). *Consciousness in the physical world: Perspectives on Russellian monism*, 209-223.
- Morris, K. (2017). The Combination Problem: Subjects and Unity. *Erkenntnis* 82(1): 103-20.
- Mørch, H. H. (2019). Is the Integrated Information Theory of Consciousness Compatible with Russellian Panpsychism?. *Erkenntnis*, 84(5), 1065-1085.
- Nagasawa, Y., & Wager, K. (2017). Panpsychism and priority cosmopsychism. In Bruntrup & Jaskolla (ed.) *Panpsychism. Contemporary perspectives*, 113-129.
- Nagel, T. (1974). What is it like to be a bat?. *The philosophical review*, 83(4), 435-450.
- Nagel, T. (1979). Panpsychism. In *Mortal Questions*. Cambridge University Press.
- Newman, W. R. (2006). *Atoms and alchemy: Chymistry and the experimental origins of the scientific revolution*. University of Chicago Press.
- Ney, A. (2015). A physicalist critique of Russellian monism. In Alter, T., & Nagasawa, Y. (Eds.). *Consciousness in the physical world: Perspectives on Russellian monism*, 346-369.
- Oderberg, D. S. (2005). Hylemorphic dualism. *Social Philosophy and Policy*, 22(2), 70-99.
- Oizumi M., Albantakis L., Tononi G. (2014) From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0., *PLoS Comput Biol* 10(5): e1003588.

- Papineau, D. (2002). *Thinking about consciousness*. Clarendon Press.
- Parvizi, J., & Damasio, A. (2001). Consciousness and the brainstem. *Cognition*, 79(1-2), 135-160.
- Parvizi, J., & Damasio, A. (2003). Neuroanatomical correlates of brainstem coma. *Brain*, 126(7), 1524-1536.
- Pasnau, R. (2011). *Metaphysical Themes 1274-1671*. Oxford University Press.
- Pautz, A. MS. A Dilemma for Russellian Monists about Consciousness. [Accessed on Feb. 13, 2018] <https://philarchive.org/archive/PAUCRM>
- Pereboom, D. (2011). *Consciousness and the Prospects of Physicalism*. Oxford University Press.
- Pereboom, D. (2015). Consciousness, physicalism, and absolutely intrinsic properties. In Alter, T., & Nagasawa, Y. (Eds.). *Consciousness in the Physical World: Perspectives on Russellian Monism*, 300-323.
- Perry, J. (2001). *Knowledge, possibility, and consciousness*. The MIT Press.
- Rea, M. C. (2011). Hylomorphism Reconditioned. *Philosophical Perspectives*, 25(1), 341-58.
- Roelofs, L. (2014). Phenomenal blending and the palette problem. *Thought: A Journal of Philosophy*, 3(1), 59-70.
- Roelofs, L. (2016). The unity of consciousness, within subjects and between subjects. *Philosophical Studies* 173(12): 3199-3221.
- Russell, B. (1921). *The Analysis of Mind*. London: George Allen & Unwin. <https://archive.org/stream/analysisofmind032971mbp#page/n3>.
- Russell, B. (1927). *The Analysis of Matter*. London: Kegan Paul. [accessed on March 6th, 2020] <https://archive.org/stream/in.ernet.dli.2015.221533/2015.221533.The-Analysis#page/n391/mode/2up>
- Schaffer, J. (2005). Quiddistic knowledge. *Philosophical Studies*, 123(1-2), 1-32.
- Scheur, J. K. (2009). Experience and self-consciousness. *Philosophical Studies*, 144(1), 95-105.

- Seager, W. (1995). Consciousness, information and panpsychism. *Journal of Consciousness Studies*, 2(3), 272-288.
- Seager, W. (2006). The 'Intrinsic Nature' argument for panpsychism. *Journal of Consciousness Studies*, 13(10-11), 129-145.
- Seager, W. (2010). Panpsychism, aggregation and combinatorial infusion. *Mind and Matter*, 8(2), 167-184.
- Seager, W. (2012). Emergentist panpsychism. *Journal of Consciousness Studies*, 19(9-10), 19-39.
- Seager, W. (2017). Panpsychist Infusion. In Brüntrup, G. & Jaskolla, L. (Eds.). *Panpsychism: Contemporary Perspectives.*, pp. 229-248, New York: Oxford University Press.
- Shams, L., Kamitani, Y., and Shimojo, S. (2000). "What you see is what you hear," *Nature*, 408:788.
- Shani, I. (2010). Mind Stuffed with Red Herrings: Why William James' Critique of the Mind-Stuff Theory Does not Substantiate a Combination Problem for Panpsychism. *Acta Analytica* 25(4): 413-434.
- Shani, I. (2015). Cosmopsychism: A holistic approach to the metaphysics of experience. *Philosophical Papers*, 44(3), 389-437.
- Shields, C., & Pasnau, R. (2016). *The Philosophy of Aquinas*. Oxford University Press.
- Siegfried, R. (2002). *From elements to atoms: A history of chemical composition* (Vol. 92, No. 4-6). American Philosophical Society.
- Siewert, C. (2013). Phenomenality and Self-Consciousness. In Uriah Kriegel (ed.), *Phenomenal Intentionality*. Oxford University Press. 235-259.
- Skrbina, D. (2005). *Panpsychism in the West*. The MIT Press.
- Skrbina, D. (ed.). (2009). *Mind That Abides: Panpsychism in the New Millennium*. John Benjamins.
- Stoljar, D. (2001). Two conceptions of the physical. *Philosophy and Phenomenological Research*, 62(2), 253-281.

- Stoljar, D. (2014). Four kinds of Russellian monism. In Kriegel, U. (Eds.). *Current controversies in philosophy of mind*, 17-39. Routledge.
- Strawson, G. (1994). *Mental Reality*. The MIT Press.
- Strawson, G. (2006). Realistic monism—why physicalism entails panpsychism. *Journal of Consciousness Studies* 13(10-11): 3-31.
- Stubenberg, L. (2014). Neutral Monism, The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/spr2014/entries/neutral-monism>.
- Stubenberg, L. (2015). Russell, Russellian monism, and panpsychism. In Alter, T. & Nagasawa, Y. (ed.). *Consciousness in the Physical World: Perspectives on Russellian Monism*, pp. 58-90.
- Teller, P. (1986). Relational Holism and Quantum Mechanics. *The British journal for the philosophy of science*, 37(1), 71-81.
- Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *science*, 282(5395), 1846-1851.
- Tononi, G., & Koch, C. (2015) Consciousness: here, there and everywhere?, *Phil. Trans. R. Soc. B*, 370(1668), 20140167.
- Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate (with Supplementary Information). *Nature Reviews Neuroscience*, 17(7), 450-461.
- Van Cleve, J. (1990). Mind-Dust or Magic? Panpsychism Versus Emergence. *Philosophical perspectives*, 4, 215-226.
- White, N. S., & Alkire, M. T. (2003). Impaired thalamocortical connectivity in humans during general-anesthetic-induced unconsciousness. *Neuroimage*, 19(2), 402-411.
- Zahavi, D. (2005). *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge MA: Bradford Book/MIT Press.