



Panpsychism and the Combination Problem

Supervisor: **Prof. dr. Roger Vergauwen**

A thesis presented in partial fulfillment
of the requirements for the degree of
Master of Philosophy (MA)

by **Rodrigo Coin Curvo**

Leuven, 2020

Table of Contents

Acknowledgments.....	3
1. Introduction.....	4
2. Panpsychism.....	5
2.1. The hard problem of consciousness.....	5
2.2. The panpsychist thesis.....	6
2.3. The physical in panpsychism.....	9
2.4. Arguments for panpsychism.....	11
2.4.1. Interlocking arguments.....	12
2.4.2. A general argument for panpsychism.....	15
2.5. Types of panpsychism.....	15
2.5.1. Constitutive and non-constitutive panpsychism.....	16
2.5.2. Russellian and non-Russellian panpsychism.....	17
2.5.3. Panprotopsychism.....	18
3. The combination problem.....	18
3.1. The hard problem and the combination problem.....	19
3.2. Aspects of the combination problem.....	21
3.2.1. Subjects combination problem.....	22
3.2.2. Qualities combination problem.....	24
3.2.3. Structures combination problem.....	25
4. A holarchical solution to the combination problem.....	26
4.1. Ken Wilber's panpsychism.....	28
4.1.1. Holons and holarchy.....	28
4.1.2. Interiors and exteriors.....	29
4.1.3. Panpsychism or pan-depthism.....	29
4.2. Solving the combination problem.....	31
4.2.1. Setting the stage.....	31
4.2.2. Reinterpreting the problem.....	32
4.2.3. Subject layering, not subject summing.....	32
4.2.4. Qualities are not pixels.....	36
4.2.5. Structures from the inside and from the outside.....	37
4.3. Russellian or not?.....	38
4.4. The holon generation problem: constitutive or not?.....	40
4.4.1. Causal efficacy and causal closure.....	41
4.4.2. Individual and collective: which systems are conscious?.....	43
4.5. Integrated Information Theory, externalism and vanishing subjects.....	47
4.6. What lies ahead.....	49
5. Conclusion.....	51
6. Bibliography.....	52
Abstract.....	55

Acknowledgments

Even though this thesis was partially written in a time of social distancing, I still have many people to thank, those who were present during this process in one way or another. First, I would like to thank my supervisor, Prof. dr. Roger Vergauwen, not only for his valuable feedback—even at a distance—, but also for being open to the directions and subjects I wanted to explore. This is specially important to me, given that I proposed to include a discussion about an author that is not very well known in the field of Philosophy of Mind.

I also would like to express my gratitude for the support of my family and friends who are in Brazil, and who played a huge role in this process, even when moving abroad was barely an idea. I thank my mother and my siblings for their help with all possible aspects that I could need, specially during changing and uncertain circumstances. I thank my friends for the many times they had to endure me talking about my plans, and also for their emotional and philosophical feedback. Thank you all for your understanding and for being such an important part of my life. I am especially thankful for and to you, Thiesa, as this project and the past year are irremediably entangled with you. Thank you for your support, for your companionship, for your love.

Thank you also to the friends I've made while in Leuven. A big thank you to Caner for being such a great 'buddy' and friend through all the changes during this period of my life. And a final special thank you to Dea, who quickly became a friend for deep discussions and stupid jokes, for simple happiness and muddy commiserations.

1. Introduction

The hard problem of consciousness is the problem of explaining the existence of phenomenal qualities in a world that is purely physical. It is possibly the most famous problem in contemporary Philosophy of Mind, and it is often referred even in other fields. But despite receiving so much attention, it is not clear how close we are to actually solving it. Its fame is probably proportional to how hard it is; and it is *hard*.

Faced with such difficulty, many philosophers have started to explore different approaches that could help unravel the problem. One of the possibilities that has gained a lot of traction in recent years is the view of panpsychism., which is the idea that fundamental entities of reality have very simple forms of consciousness. The hope is that, by establishing consciousness as something fundamental at the base of reality, it could be much easier to explain the conscious experience of complex beings like humans.

However, as panpsychism starts solving one problem, it creates another one: the combination problem. Even if we assume that panpsychism is correct and that fundamental entities are conscious, it seems that we have no straightforward way of conceptualizing how simple conscious beings—microsubjects—could combine into more complex ones—macrosubjects. It is common to characterize the combination problem by saying: subjects don't sum. This is the major problem that contemporary panpsychism has been facing.

In this thesis I will examine panpsychism and the combination problem. I will also discuss a solution that, as far as I can tell, has received little to no attention in the philosophical literature. It comes from the philosopher Ken Wilber, and uses the concept of holons—wholes/parts—to think about questions of ontology and consciousness. Unfortunately, even Wilber himself discusses his solution very briefly. Therefore, part of my goal is to expand its application to try and deal with the different aspects of the combination problem.

In Chapter 2, I discuss panpsychism in connection to the hard problem of consciousness. I will elaborate on what is precisely the thesis of panpsychism, as well as some of the issues associated with its definition. Next, I will discuss some of the arguments that have been made to defend panpsychism, and try to summarize them in a general argument. I will also talk about some of the classification that different types of panpsychism can have.

Chapter 3 contains an examination of the combination problem, beginning with a connection and a parallel with the hard problem of consciousness. Then, I discuss various ways in which the

combination problem has been presented, depending on what aspect is taken as the focus. This will include distinct combination problems related to subjects, to qualities, and to structures.

In Chapter 4 I explore Ken Wilber's view and solution to the combination problem. I then apply it to the different versions of the problem and discuss some of its implications. Next, I discuss how Wilber's view can be seen in terms of panpsychism classification, and examine some sources of problems, as well as potential responses. Finally, I close the chapter by talking about a few paths that could be fruitful for further exploration.

2. Panpsychism

In this chapter we will examine panpsychism, which is the idea that the ultimate constituents of reality have some form of consciousness. I will begin by discussing the hard problem of consciousness, followed by a definition of the panpsychist thesis, the arguments for its defense, and a classification of the different types of panpsychism.

2.1. *The hard problem of consciousness*

The hard problem of consciousness is the problem of explaining the existence of phenomenal qualities in a world that is purely physical. In this context, when we speak of phenomenal qualities—or simply of *qualia*—we refer to the qualitative aspects of subjective experience, such as the redness of red, for example. Coined by the philosopher David Chalmers (1995), the expression 'hard problem of consciousness' became the contemporary representative of the mind-body problem.

It is said to be 'hard' in order to distinguish it from what Chalmers (1995) would call the 'easy problems of consciousness'. The latter are the ones that can be solved using the standard methods of cognitive sciences. It is not that they are actually easy—it could take us a long time to solve them, if ever—but that they are in principle solvable by continuing to apply similar methods to what we have already been using. Chalmers' seminal article sparked a lot of discussion and publications by other philosophers. These in turn lead Chalmers (1997) to further develop his argument as he attempted to respond to support and criticism alike.

The hard problem of consciousness is a problem that comes from trying to see how qualia would fit into a physicalist view. To understand why, on a first approximation let us consider physicalism as the view that all truths about reality are grounded in truths described by physics. Physics can give structural and functional explanations about reality. But experience *qua* experience

is neither functional nor structural. Therefore, physicalism can not explain subjective experience, because truths about qualia can not be grounded in truths given by physics.

Put differently, all truths given by physics are compatible with the absence of subjective experience. It is *a priori* conceivable that there could exist a world that is exactly like ours, but where subjective experience does not exist at all. However, since the physicalist description of that world would remain the same as in ours, this means that there is something that the physicalist description is missing.

Those arguments for the hard problem were given by Chalmers (1995), but there had been other previous arguments about the limits of a physicalist explanation for subjective experience, such as the ones given by Nagel (1974), Jackson (1982) or Levine (1983). The point here is that, if we think that those arguments are strong reasons to question the possibility of a purely physicalist view of the world, then we are urged to find alternatives.

In the same article where he discusses the hard problem of consciousness, Chalmers defends a possible solution:

I suggest that a theory of consciousness should take experience as fundamental. We know that a theory of consciousness requires the addition of something fundamental to our ontology, as everything in physical theory is compatible with the absence of consciousness. We might add some entirely new nonphysical feature, from which experience can be derived, but it is hard to see what such a feature would be like. More likely, we will take experience itself as a fundamental feature of the world, alongside mass, charge, and space-time. If we take experience as fundamental, then we can go about the business of constructing a theory of experience. (Chalmers 1995, 210)

Even though in that article Chalmers does not talk explicitly about panpsychism as he does in other texts, the idea of taking experience as a fundamental part of the world is a core tenet of panpsychism. Instead of positing that consciousness is something completely new that ‘lights up’ at some moment, and instead of saying that it is another substance distinct from everything else, we simply accept consciousness as a fundamental and natural part of reality.

In the next section I will discuss in more details what is included in the panpsychist view. In truth, the basic idea of panpsychism is much older than Chalmers’ article, and the same is true for the mind-body problem, which has a particular expression in the hard problem of consciousness. But it’s interesting to see that both the problem and a possible solution have a kind of a concomitant resurgence.

2.2. *The panpsychist thesis*

Panpsychism is actually not one single view, but a group of views, and each has a specific take on the nature of the mind. For that reason, David Skrbina (2005, 2) refers to panpsychism in general as a meta-theory. We will detail some of the different types of panpsychism in following sections.

However, in order to discuss the general arguments, we need to establish a working definition, one that clearly states the thesis to be defended. That in itself is not a simple task, because of the variety of views that can be encompassed. In addition, as Skrbina (2005) points out, the different definitions of panpsychism rely on other terms that are often ambiguous and not always interchangeable, such as ‘mind’, ‘consciousness’ and ‘experience’. Still, he identifies three characteristics as being essential to the concept, and from them he forms a definition:

(1) Objects have experiences for themselves; that is, the mind-like quality is something internal to or inherent in the object. (2) There is a sense in which this experience is singular; to the extent that a structure of matter and energy that we call an object is one thing, this oneness is reflected in a kind of unitary mental experience. (3) An object is a particular configuration of mass/energy, and therefore any configuration or system of mass/energy should qualify in the same sense. Thus, a functional definition of panpsychism might be “All objects, or systems of objects, possess a singular inner experience of the world around them.” (Skrbina 2005, 16)

We can discuss each of these characteristics as playing a different role in the definition. The first one is the basis of the panpsychist view, since the idea is to make the mental phenomena an intrinsic part of the world. And because the mental features are taken to be intrinsic to the objects, this distinguishes the position from other ones, such as substance dualism, for example.

The second characteristic represents two different but related questions. One is about the unity of consciousness, that is, the fact that our experience of consciousness is always given as a unified phenomenon. The other is that each of us has their own, unified, distinct consciousness. It makes sense that those two features are present in his definitions since, as we will discuss in other sections, the arguments for panpsychism usually begin with our own human experience, from which these question are taken.

While (1) and (2) relate to the ‘psychism’ part of panpsychism, characteristic (3) relates to ‘pan’, i.e., it defines the extension to which the mental features apply in the world. Skrbina himself admits that this third characteristic is the most contentious one. Still, he uses it for his definition because *a priori* he sees no reason to exclude any system from qualifying for the attribution of mental features. However, allowing any system to be conscious could lead to other theoretical problems, including conflicts with the second characteristic.

But most importantly, it is not even necessary for panpsychism to state that all systems are conscious. To include this characteristic in a general definition seems to be an overstep, contradicting certain versions of panpsychism. Strawson hints at the issue of the extension to which mental features should be applied in a panpsychist view:

we will also need to apply our minds to the question whether the class of subjects of experience contains only ultimates, on one hand, and things like ourselves and other whole animals, on the other hand, or whether there are other subjects in between, such as living cells. Panpsychism certainly does not require one to hold the view that things like stones and tables are subjects of experience (Strawson 2006, 26)

Skrbina's definition does seem to be general enough to encompass different types of panpsychism. Nonetheless, it is a strong definition, and not all types of panpsychism would be in agreement with it. For that reason, it would be better to use a different definition, one that is compatible with Skrbina's, but has a weaker claim. That would also mean it is more open to different sets of assumptions.

One possibility would be to use a definition such as the one David Chalmers proposes when giving an overview of panpsychism:

we can understand panpsychism as the thesis that some fundamental physical entities have mental states. For example, if quarks or photons have mental states, that suffices for panpsychism to be true, even if rocks and numbers do not have mental states. Perhaps it would not suffice for just one photon to have mental states. The line here is blurry, but we can read the definition as requiring that all members of some fundamental physical types (all photons, for example) have mental states. (Chalmers 2013, 1)

This definition differs from Skrbina's because it is restricted in two ways. First, it is limited to ultimate physical entities, instead of applying to all objects and systems of objects. Second, Chalmers restricts it even more by specifying it explicitly in terms of "some fundamental physical entities".

Strictly speaking, it is true that it is not necessary for panpsychism that all fundamental physical entities have some form of consciousness. Restricting it to some of them would be enough. This is because, as we will see in the next section, panpsychism is trying to solve the hard problem of consciousness by putting consciousness itself at the bottom level, so that we can get it when more complex systems arise.

Even if we restrict consciousness to some but not all of the fundamental entities, it could be enough to explain the existence of more complex forms of consciousness, as long as the relevant

type of fundamental entities are present in conscious macroentities such as humans. So since the possibility of the existence of other non-conscious ultimate entities does not change the general arguments for panpsychism, Chalmers' definition is unnecessarily restricted. Therefore, using this second restriction would only add an unnecessary complexity to the discussion.

An additional issue in these definitions is about what exactly are the features that are being claimed to be present in fundamental entities. Skrbina uses "inner experience", while Chalmers talks about "mental states". As mentioned before, the use of these terms is often ambiguous, and the precise manner in which they appear in the definition will depend on each version of panpsychism. That is an important and extensive discussion in itself, but for the moment we need to accept that any choice we make will not be able to cover all the different positions. Therefore, we will use these terms with a broader general meaning, unless otherwise specified.

With that in mind, we establish a working definition: panpsychism is the view that fundamental physical entities have some form of experience.

2.3. *The physical in panpsychism*

There is one remaining question regarding the panpsychist thesis we have just established. In contemporary Philosophy of Mind, the revival of panpsychism happened mainly as a possible solution for the hard problem of consciousness which, as explained, is the problem of accounting for the existence of subjective experience in a wholly physical world.

Because of that starting point, many contemporary panpsychist views have a close relationship with physicalist ideas. In this case it can be seen in our definition above, in which we speak in terms of fundamental *physical* entities. The remaining question then pertains to what is the role of the word 'physical' here.

To answer that question, some distinctions are useful. Strawson (2006) distinguishes physicalism from what he calls *physicSalism*. For him, physicalism is the view that every concrete phenomena is physical, while *physicSalism* is the idea that "the nature or essence of all concrete reality can in principle be fully captured in the terms of physics" (Strawson 2006, 4). Strawson also says that, since most people use the term 'physicalism' as meaning what he calls *physicSalism*, he needs to use other names for his position, such as 'real physicalism', 'realistic physicalism' or simply 'realistic monism'.

Chalmers (2013) makes a related point, based on the distinction between *narrowly* physical properties and *broadly* physical properties:

We can say that *narrowly physical* properties are microphysical role properties, such as the dispositional property associated with having a certain mass, or the second-order property of having a property that plays the mass role. We can say that *broadly physical* properties are physical role properties along with any properties that realize the relevant roles: categorical bases for the mass dispositions, first-order properties that play the mass role. (Chalmers 2013, 11, emphasis in the original)

Through that distinction, he differentiates two notions of physicalism: “narrow physicalism, which holds that phenomenal truths are grounded in narrowly physical truths, from broad physicalism, which holds that phenomenal truths are grounded in broadly physical truths” (Chalmers 2013, 12). By ‘phenomenal truths’ he means truths about phenomenal properties, i.e., the properties that characterize subjective experience.

Chalmers’ concept of *narrow physicalism* relates closely with Strawson’s idea of *physicSalism*. Physics is taken to cover at least¹ truths about narrowly physical properties. Therefore, to say that *physicSalism* is true (physics can capture all truths about concrete reality) is roughly equivalent to saying that narrow physicalism is true (phenomenal truths are grounded in narrowly physical truths).

With that in mind, we can now discuss the role the term ‘physical’ plays in our definition of panpsychism. The question we face with that usage is twofold: first, what we mean by ‘fundamental physical entities’; second, whether we would be better off simply talking about ‘fundamental entities’.

Panpsychism is a possible solution for the hard problem of consciousness. However, if *physicSalism*—or narrow physicalism—is true then panpsychism is unnecessary. Therefore, when we talk about ‘physical’ in the context of panpsychism, we can not be simply referring to what physics tells us about the fundamental nature of reality. What then?

The term ‘physical’ here is playing two roles. First, it is establishing a type of monism. Just as in the argument above, if substance dualism² is true, then panpsychism is unnecessary—even though the two are not strictly incompatible. So, when we speak of physical entities for panpsychism, we are implying that there is only one type of *stuff* in the world, that is, we are

1 I say ‘at least’ because there is a parallel question here on whether or not physics even require broadly physical properties, i.e., *quiddities*. However, that does not affect the discussion at this point, because we are only establishing a link between the concepts of ‘narrow physicalism’ and ‘*physicSalism*’.

2 Substance dualism is understood here as the general claim that mind and matter are two distinct substances.

implying monism. Second, we are saying that those entities are of the same nature as to what we usually refer when we point to chairs and tables, for example³.

But if that is all we mean by ‘physical’, would it be better if we just used a more general wording, as when Strawson talks about ‘realistic monism’? If we just talked about fundamental entities without any mention of ‘physical’, our definition of panpsychism would be simplified. The drawback would be that we would risk losing some clarity for the definition. Without a reference to the physical, we could be getting too close to idealism⁴, for example.

In a sense, the choice is not that important, as long as we are clear about what is implied. However, there is one last reason that makes me believe it is better to not use ‘physical’ in our definition. In contemporary panpsychism we are often coming from a physicalist point of view and bringing with us plenty of assumptions about its ontology, and those can get mixed up easily. That will become clearer when we start discussing the combination problem.

Besides, it is common to refer to atoms as examples of fundamental entities in the discussion, and for the sake of simplicity we will do the same. So even if we remove ‘physical’ from the definition, it will never be far from the discussion. But here atoms will merely serve as placeholders. What we actually refer to are the fundamental entities—the ultimate constituents or simply ‘ultimates’—, that is, whatever entities that happen to be fundamental in the structure of reality.

The result, then, is that we can update our definition: panpsychism is the view that fundamental entities have some form of experience.

2.4. Arguments for panpsychism

We have discussed how the hard problem of consciousness might motivate us to look for alternative solutions such as panpsychism. An important issue to keep in mind is that, as mentioned previously, panpsychism is not a single theory, but can be viewed as a meta-theory. Looking at the many versions of panpsychism that appeared throughout the history of philosophy, Skrbina (2005) identifies over 10 distinct types of arguments that have been given in its defense. Although many of

3 We are assuming, as discussed in our definition, that all ultimate entities have some form of consciousness. It could be the case that only some of the ultimate entities are conscious, and that the entities that form tables and chairs are not of that type. In that case, we would need extra examples to include conscious entities. The point here is simply to say that we are not talking about some out of the ordinary type of entity.

4 This, of course, is not necessarily something bad in itself. Chalmers tells the story about how he heard someone saying that “One starts as a materialist, then one becomes a dualist, then a panpsychist, and one ends up as an idealist” (forthcoming, 1). Maybe this is part of that path. However, it’s important not to blur the lines too much if we hope to limit our discussion around panpsychism.

those arguments are not necessarily convincing, especially in the context of contemporary philosophy, there are some that deserve special attention.

From the arguments Skrbina (2005, sec. 10.1) presents, the ones that can be highlighted as the most relevant are: *the argument from continuity*⁵, which relies on the idea of a common principle that exists in all things; *the argument from non-emergence*, which rejects the possibility of mental features emerging from non-mental ones; *the naturalized mind argument*, which emphasizes panpsychism's advantage to integrate mind into the natural world; and *the last man standing argument*, which is a type of a negative argument, stating that because all other alternative theories have failed, panpsychism gains more credibility.

2.4.1. Interlocking arguments

The types of arguments mentioned above are stronger when given in combination with each other, in interlocking and more fully developed arguments. One interesting feature of such interlocking arguments is that many of them seem to work through the rejection of either other positions or routes that seem troublesome. Chalmers (2013), for example, gives what he calls a dialectical argument, where the thesis is materialism—which we will call ‘physicalism’, for consistency—the antithesis is dualism, and the synthesis is panpsychism.

For the rejection of the thesis of physicalism, Chalmers applies a conceivability argument similar to what we briefly discussed for the hard problem of consciousness. The basic idea is as follow. Let us say that we have all truths given by physics on one hand, and some truth about mental phenomena on the other. It is conceivable that the first could exist while the latter is absent at the same time. Because it is conceivable, one can argue it is metaphysically possible, which in turn would entail that physicalism is false.

This leads Chalmers to the antithesis of dualism. Against dualism and in favor of physicalism, he uses what he calls the causal argument. This argument relies on two ideas: first, that everything that is causally relevant for a physical event can be fully explained in physical terms; second, that mental phenomena is causally relevant for physical events. It follows, then, that mental phenomena can be fully explained in physical terms, which means that physicalism is true and, as a corollary, that dualism is false.

With the rejection of both physicalism and dualism, Chalmers defends that panpsychism can then serve as the synthesis, “as a view that captures the virtues of both views and the vices of

⁵ Skrbina names this ‘the argument by continuity’ when enumerating all of the arguments, but in other places he calls it ‘the argument from continuity’. I chose to list it as the latter.

neither” (2013, 2). Specifically, he believes that a particular type of panpsychism, *constitutive Russellian panpsychism*, is the view that can deal with the problems faced by physicalism and dualism. In the next section we will discuss the different types of panpsychism.

Arguments against physicalism and dualism have been debated at length in the philosophical literature, and it is not in the scope of this work to rehash them. Each has multiple premises, explicit as well as hidden ones, and the discussion around those premises lead to many forking paths. What is important here is that, by rejecting both physicalism and dualism, Chalmers arrives at panpsychism as the possible alternative. This displays features of the arguments Skrbina mentions.

The rejection of physicalism has implicitly rejected a version of emergence of the mental from non-mental, and this of course is the argument from non-emergence. The rejection of dualism, on the other hand, relates indirectly to the argument from continuity, for it avoids creating a split between the mental and the physical. Finally, the establishment of panpsychism as the synthesis ties back to both the naturalized mind argument and the last man standing argument.

Strawson (2006) also gives a defense of panpsychism that displays some of these types of arguments, especially the argument from non-emergence. As explained previously, Strawson distinguishes between physicalism and real physicalism. Physicalism is the idea the physics covers all truths about reality, while real physicalism (RP) states that every concrete phenomenon is physical.

Strawson begins with the idea that consciousness is a concrete phenomena, which in RP means it is physical. We are all aware of our own human consciousness, so he takes experience to be an indubitable starting point in RP, for “experience is itself the fundamental given natural fact” (Strawson 2006, 4). By saying that, he is doing two things: one, he is avoiding some paths in which consciousness is excessively deflated or even dismissed as unreal; two, he is saying that consciousness has to be put in the natural order of the world as any other phenomena. Here we can see the basic assumptions for both the argument from continuity and the naturalized mind argument.

However, the core of Strawson’s defense is related to the argument from non-emergence. First, he states that in what people usually call physicalism there is an assumption that is unwarranted, namely that the fundamental basis of the physical world is completely non-experiential (NX⁶). He believes that this assumption is not only unnecessary, but actually not defensible if one accepts the basic ideas in real physicalism.

6 Strawson names the non-experiential assumption as ‘NE’, but I’m going to use ‘NX’ to avoid confusing ‘NE’ with non-emergence

Strawson defends that if one tries to put the assumptions of NX and RP together, the only way to get consciousness is to appeal to a sort of emergence. However, the type of emergence required to get the experiential—i.e., consciousness—from the non-experiential would be too big a leap. In weaker cases of emergence—such as the liquidity of water, for example—, there is a graspable and intuitive sense of what is happening, as well as a type of reduction without remainder. For the case of consciousness, though, that would not be the case, because what would be required is a type of brute emergence, which is unacceptable.

We can now summarize his entire argument as follows. From RP we have that, because consciousness is a concrete phenomena, it is therefore physical. If NX is true, we need a type of brute emergence in order to explain consciousness. But emergence can not be brute. The only option, then, is to reject the premise of NX. Therefore, panpsychism is true.

Coleman (2006) defends that in order to arrive at the conclusion that Strawson does, there are two important assumptions that are needed, which he calls *smallism* and *perspicuity*. Smallism is the idea that all facts about reality are grounded in facts about the smallest entities, i.e., the fundamental entities. Perspicuity, in turn, is the view that the determination of facts of higher-level entities from the facts of lower-level ones is, in principle, knowable *a priori*.

Smallism is, according to Coleman, first a metaphysical issue, and only afterwards it becomes an epistemological one. He says that this is the basis where many philosophers apply the idea of supervenience. Coleman also believes it is understandable that Strawson assumes it, given how pervasive it is in the philosophy of mind, especially among physicalists. This assumption shows up particularly in the way Strawson argues about how we would need a type of brute emergence in order to get experiential features from non-experiential ones.

However, Coleman defends that it is possible to hold both NX and RP if one rejects smallism, for example in a *macro non-reductive physicalism*. Such a view rejects the idea that higher levels are determined by the lowest level, even if there are in fact levels of reality. In this case, each level could be autonomous with regards to how its properties are determined. For consciousness, this would mean that maybe it accompanies a certain level of complexity of reality, but does not emerge out of it. Coleman says it would be a type of brute accompaniment and not of brute emergence. Because of that, it would avoid this specific objection from Strawson.

This leads us to the second assumption, perspicuity. While smallism is more of a metaphysical assumption, perspicuity is an epistemological one. It is not enough to be committed to smallism in order to apply Strawson's argument, because one could say that even though the facts about the

lowest levels determine the facts about the higher levels, it would not be possible to see *a priori* how that determination works. This means that NP and NX could work together if one gives up on the idea of knowing—even from ‘God’s point of view’—how the experiential emerges out of the non-experiential.

This, of course, does not mean that Strawson’s argument does not hold. It simply means that there are more assumptions needed, and that there is no need to commit to panpsychism if you are willing to give up on one of those assumptions. Coleman himself thinks that these assumptions are true, but also that they can not be taken for granted without argument.

2.4.2. A general argument for panpsychism

The exposed arguments are a few examples of how panpsychism can and has been defended, and how the interlocking parts—assumptions and arguments—play a role in this. With those in mind, we can now summarize a general version of an argument for panpsychism:

- (1) Consciousness is real.
 - (2) Consciousness is causally relevant.
 - (3) Reality is made of one type of stuff.
 - (4) The type of stuff reality is made of is physical, i.e., the same as tables and chairs.
 - (5) Facts about reality are determined by facts about the smallest entities.
 - (6) To get experience out of non-experience it would require a type of brute emergence.
 - (7) Emergence can’t be brute.
-
- (8) From (1)–(7), panpsychism is true.

Let us go through each one: (1) is what is assumed in real physicalism, and a rejection of eliminativism; (2) is required for the causal argument, and a rejection of epiphenomenalism; (3) is related to the argument from continuity, but also comes from the rejection of dualism given by the causal argument; (4) is also a basis for real physicalism and, as explained in the section about the role of the physical in panpsychism, a type of monism and a rejection of strict idealism; (5) is smallism, also indirectly related to the argument from continuity; (6) and (7) are both perspicuity and non-emergence, and ultimately a rejection of physicalism. If all these assumptions are accepted, we conclude that panpsychism is true.

2.5. Types of panpsychism

As we have reiterated, panpsychism is not a single theory, but can be thought of as a group of theories or even as a meta-theory. This means that its common and general features are not enough to pin down the view. So far we have been dealing with panpsychism without discussing some specific claims, particularly about the nature of the mind.

As with many philosophical taxonomies, the different views do not always fit perfectly within the bounds of some categories. However, it is possible to use axes with which the theories can be roughly classified. The two axes we will discuss are: *constitutive vs non-constitutive* panpsychism; and *Russellian vs non-Russellian* panpsychism.

2.5.1. Constitutive and non-constitutive panpsychism

The categorization of a panpsychist theory as constitutive or non-constitutive depends on how it understands the relation between *microexperiences* and *macroexperiences*. A *microexperience* is the experience that a fundamental entity has, that is, the experience of microentities. A macroexperience, on the other hand, is the type of experience that macroscopic entities have, entities such as human beings and any other entity that is not a fundamental one.

Constitutive panpsychism is the view that microexperiences constitute macroexperiences. This view relates to the smallism assumption that Coleman (2006) describes, in which the macroproperties are grounded in the microproperties. Chalmers (2013) states that the relation of constitution can be one of partial grounding—as opposed to complete grounding—, if one believes that only certain aspects of macroexperiences are grounded in microexperiences, for example regarding structural or functional properties. The constitutive type of panpsychism can also be related to a more mereological discussion, for example in the version that Luke Roelofs (2019) calls combinationism.

It is commonly accepted that macrophysical properties are grounded in microphysical ones. In this type of panpsychism, the same thing is thought about the relation between macroexperiences and microexperiences. In this sense, this view keeps the features that one would expect, for example, in a defense of panpsychism that uses arguments such as the one from continuity and the naturalized mind argument.

In contrast to the constitutive type there is also *non-constitutive panpsychism*. For non-constitutive panpsychism the relation between macro and microexperiences is not one of constitution. Usually the proposed relation in this case is one of emergence, which is why it is sometimes called emergentist panpsychism.

The appearance of emergence in panpsychism could create some worry, especially given that the argument from non-emergence is sometimes central in the defense of panpsychism. The stronger the emergence required for the theory to work, the less one can rely on the argument of non-emergence. However, even if we talk about emergence, there is still leeway in the definition of the type of emergence, which means it does not need to be the same type of brute emergence that Strawson (2006) criticizes.

2.5.2. Russellian and non-Russellian panpsychism

For the second axis of classification we have the contrast between Russellian and non-Russellian panpsychism. This is related to a view about the intrinsic nature of matter, which is usually associated with the philosopher Bertrand Russell—hence the name. The idea is that although physics can tell us what matter *does*, it is unable to tell us what matter *is*. In other words, although we can know empirically about the extrinsic nature of reality, our theories seem to have a hole where the intrinsic natures—the fundamental categorical properties, the *quiddities*—should be.

Russellian panpsychism is a way of trying to solve the problem of intrinsic nature at the same time as it solves the hard problem of consciousness. Because of this feature, it can be used as an argument, what Roelofs (2019) calls *the intrinsic natures argument for panpsychism*. On one hand, panpsychism is trying to find a suitable place for macroexperiences, which in turn are connected to microexperiences, that is, the basic phenomenal qualities. On the other hand, there is a philosophical problem with the hole where intrinsic natures should be. In Russellian panpsychism, microexperiences are taken to be precisely that intrinsic nature. Here, the basic phenomenal qualities—attributed to fundamental entities—are actually the intrinsic nature that plays the role of the properties that physics investigates. Put differently, in Russellian panpsychism the microexperiences are what matter *is*; physics then investigates what it *does*.

Finally, there is non-Russellian panpsychism. For this type of panpsychism, the microexperiences can be seen as being side by side with physical properties. This can be interpreted as a type of property dualism, which is why Roelofs (2019) calls it ‘dualistic’ panpsychism. Many philosophers (Chalmers 2013; Roelofs 2019; Goff 2007) believe that this is a more problematic type of panpsychism, which is why they usually prefer to focus on the Russellian type. One of the reasons is that non-Russellian panpsychism seems to be particularly vulnerable to objections such as the causal argument, which we explained as one of the arguments used to defend panpsychism. However, a non-Russellian type of panpsychism could still be defensible if one finds alternative arguments to defend panpsychism, rather than relying on the causal argument.

2.5.3. Panprotopsychism

Panprotopsychism is a view that is usually discussed together with panpsychism. Even though it will not be a part of the focus of this thesis, I would like to mention it for the sake of completeness. Goff (2009) describes the view:

The panprotopsychist adds protophenomenal properties to the physical facts. Protophenomenal properties are ‘hidden’ properties of physical ultimates – hidden in the sense of being empirically indiscernible – which, although not themselves phenomenal properties, in certain combinations come together to constitute conscious experience. (Goff 2009, 292)

As can be seen, panprotopsychism takes an approach similar to that of panpsychism, but instead of establishing phenomenal properties as fundamental, it uses protophenomenal ones. We can raise the question of whether this would actually be a distinct view, or if it would simply collapse into either panpsychism or physicalism. Chalmers (2013, 16) defends that there are enough reasons to think that panprotopsychism can be seen as a distinct view, even if it is susceptible to some of the same criticisms that are directed towards panpsychism and physicalism.

3. The combination problem

Arguments such as the ones explained in the previous sections might suffice for us to take panpsychism into serious consideration as a solution to the hard problem of consciousness. But even though panpsychism has its promises, it also faces a very specific roadblock: the combination problem. The combination problem is the problem of explaining the existence of higher order consciousness, if one accepts that panpsychism is true. It has been the focal point of many of the contemporary research about panpsychism, for example with Goff (2009), Mørch (2014), Chalmers (2016), Miller (2018), and Roelofs (2019).

The name ‘combination problem’ was coined by William Seager (1995), but the problem seems to have been first pointed out by William James:

Where the elemental units are supposed to be feelings, the case is in no wise altered. Take a hundred of them, shuffle them and pack them as close together as you can (whatever that may mean); still each remains the same feeling it always was, shut in its own skin, windowless, ignorant of what the other feelings are and mean. There would be a hundred-and-first feeling there, if, when a group or series of such feelings were set up, a consciousness *belonging to the group as such* should emerge. And this 101st feeling would be a totally new fact; the 100 original feelings might, by a curious physical law, be a signal for its *creation*, when they came together; but they would have no substantial identity with it, nor it with them, and one could never deduce the one

from the others, or (in any intelligible sense) say that they *evolved* it. (James 1890, 160, italics original)

James also characterizes the problem in an example where you suppose that there is a group of people, each thinking of a word. No matter how you put those people together or how hard each one thinks of the word, James says, the combination of the consciousness of the words will never form the consciousness of a sentence. In the book, this is discussed in a section called ‘*Self-compounding of mental facts is inadmissible*’, which indicates how strongly he believed the idea was problematic. In his own words, “the private minds do not agglomerate into a higher compound mind” (James 1890, 160).

As mentioned, the combination problem is the problem of explaining the existence of higher order consciousness, even if we are willing to accept that panpsychism is true. If panpsychism is true, the fundamental entities of reality have some form of consciousness. But the starting point and one of the eventual goals of the philosophical discussion about consciousness is to explain our own, human consciousness. As Philip Goff (2009) puts it:

Whilst it is of course true that panpsychism gives an account of the presence of consciousness in reality, in the sense that it supposes that physical ultimates are conscious, what we require out of a theory of consciousness is not just a theory that explains the presence of *any old conscious experience* in our world. What we want explained is *our* conscious experience, the conscious experience that corresponds to humans and other organisms. (Goff 2009, emphasis original)

So the question is: how is it that we can get higher-order consciousness—such as that of a human—from conscious fundamental entities? To accept that fundamental entities are conscious does not seem to immediately imply that, when those entities are put together, they would combine to form a unified higher-order consciousness.

3.1. *The hard problem and the combination problem*

The combination problem shares a feature with the hard problem of consciousness, and understanding that similarity can be helpful. As previously mentioned, one of the arguments behind the hard problem of consciousness is based on conceivability, as pointed out by Chalmers (1995). It is conceivable, the argument goes, that there could be a universe in which all truths given by physics are the same as in ours, but where no consciousness exists. Because this is conceivable, it seems that something is missing in the physicalist picture of reality.

A similar argument can be created against panpsychism, which is what Seager (1995) uses when defining the combination problem. Against physicalism, we underline the conceivability of physics without consciousness. Now let us assume that panpsychism is true. Even if every fundamental particle is conscious, it is still conceivable that no combination of particles would lead to a higher-order consciousness, *including the combination that is found in human brains*. In other words, just as we can not seem to get consciousness from physics, we seem unable to get macro-consciousness from micro-consciousness.

Seager (1995) refers to the hard problem of consciousness as ‘the generation problem’, since it is about “explaining why and how experience is generated by certain particular configuration of physical stuff” (W. Seager 1995, 272). He explains that panpsychism suffers from the same issue:

the combination problem points to a distinctive generation problem in panpsychism which is formally analogous to the problem of generating consciousness out of matter. Panpsychism will have no advantage over physicalism if essentially the same problem lurks at its heart, and, of course, it faces the intrinsic implausibility of asserting that atoms are conscious (W. Seager 1995, 281)

A similar point is made by Chalmers (2016, sec. 7.4.3) in what he calls *the conceivability argument*. For the argument, let ‘PP’ be all facts about reality at the microscopic scale, that is, microphysical and microphenomenal facts, and let ‘Q’ be a macrophenomenal fact, such as ‘some human being is conscious’. Chalmers then sets up these premises: “(1) PP&~Q is conceivable. (2) If PP&~Q is conceivable, it is metaphysically possible. (3) If PP&~Q is metaphysically possible, constitutive panpsychism is false” (Chalmers 2016, 187). The conclusion, from (1)-(3), is that constitutive panpsychism is false.

In that argument, Chalmers (2016) focuses on constitutive panpsychism—the idea that microconsciousness constitute macroconsciousness—, which he takes to be the most important type of panpsychism. Such a focus is also understandable because it can be harder to construct arguments against the theory without being specific. But as he alludes (Chalmers 2016, 185), the argument could in principle be applied generally against panpsychism.

Chalmers also makes the connection with the argument for the hard problem of consciousness, saying that “premises (2) and (3) are parallel to corresponding premises in the familiar conceivability argument against physicalism” (Chalmers 2016, 187). It is premise (1) that is different. However, there is still a connection that can be seen in it.

In the conceivability argument against physicalism one can use the idea of *experiential zombies*. An experiential zombie is a perfect *microphysical duplicate* of a human being, for example, but one that does not have any experience at all. The behavior of an experiential zombie might be exactly the same as the human—since it is physically identical—, but there is no *what it is like to be* the zombie, that is, ‘the lights are out’. And in a sense it is the conceivability of experiential zombies that is at the core of such argument against physicalism.

Similarly, one could talk about zombies in the context of panpsychism. But this would be a different type of zombie, one that Goff (2009) calls *microexperiential zombies* and Chalmers (2016) calls *panpsychist zombies*. A panpsychist zombie is a higher-order zombie. On one hand, it is not only a perfect microphysical duplicate, but also a microexperiential duplicate. Put differently, both the microphysical and the microexperiential facts are identical to those of a regular human being. On the other hand, the panpsychist zombie lacks a higher-order consciousness. All its microphysical constituents might be conscious as per panpsychism, but it is conceivable that the entity as a whole could be non-conscious.

Chalmers (2016, 188) says that the panpsychist zombie is what is being asserted in premise (1) discussed above, where $PP \& \sim Q$ is conceivable. And just as in the argument against physicalism, because this possibility is conceivable even if one accepts panpsychism, it indicates that there is something that the theory is missing.

3.2. Aspects of the combination problem

So far we have been discussing the general version of the combination problem. There are, however, aspects of it that can be distinguished as specific problems in themselves. In a sense, even the conceivability argument can be seen as one of those that can be classified as combination problems. Chalmers (2016) gives an extended description of multiple versions of the combination problem. He describes three general aspects:

The combination problem can be broken down into at least three subproblems, reflecting three different aspects of phenomenal states: their *subjective character* (they are always had by a subject), their *qualitative character* (they involve distinctive qualities), and their *structural character* (they have a certain complex structure). (Chalmers 2016, 182, emphasis mine)

These aspects—as well as the further subdivisions of them—can also be turned into more formal arguments against panpsychism, something that is done both by Chalmers (2016) and Roelofs (2019) in their exposition. On top of Chalmers’ grouping and description of these aspects, Coleman (2016, sec. 10.1) classifies the different combination problems in two ways:

Objections to a theory of mind that flow from its fit with our best science I call ‘**bridging problems**’— they are difficulties around our building a bridge between a given theory of mentality and our scientific theories of the brain and physical world. The [other] [...] sort of objection are ‘**internal problems**’— they are difficulties alleged to afflict a theory of mind taken on its own terms, or within the field of theories of mind. (Coleman 2016, 251, bold mine)

What follows is a discussion—using Chalmers’ grouping and Coleman’s categorization—of the different aspects of the combination problem, as well as how they highlight, within the combination problem, the issues that a panpsychist theory needs to be able to handle in order to be successful.

3.2.1. Subjects combination problem

The combination problem regarding subjects is possibly the closest to the general version of the combination problem as described above. The subjectivity aspect of the combination problem is the problem of explaining how it is that microsubjects can combine and give rise to a macrosubject. On this, Roelofs (2019) says:

The simplest argument against combinationism [(i.e., Roelofs’ version of panpsychism)] is that “subjects don’t sum”: knowing what it’s like to be each of some collection of subjects seems impotent to tell us anything about what it’s like to be *someone else*. And since the whole they compose is not any one of them, it seems to be an example of “someone else” (Roelofs 2019, 54, italics original)

Chalmers (2016, sec. 7.3) calls this the *subject summing problem*, and says that it can be seen as an extension of William James’ example where 100 subjects do not entail an additional subject. But we can conceive the problem in a slightly different manner, because there is a sense in which it seems that each of the subjects is closed up onto itself, and no matter how close they are to each other, “still each remains the same feeling it always was, shut in its own skin, windowless, ignorant of what the other feelings are and mean” (James 1890, 160).

In Coleman’s (2016) classification, the subject summing problem is internal because “it concerns our concept of a subject and whether subjects could constitute another subject” (Coleman 2016, 251). But within the combination problem with regards to subjects, Coleman makes a further distinction, saying that there is a negative and a positive version of it. The subject summing problem is the negative problem, for it is actually an *explanatory gap* of how to entail macrosubjects from a group of microsubjects (Coleman 2016, 254).

The positive version of the subject combination problem would be what Chalmers (2016) calls the *unity problem*. In Coleman's (2016) view, while the negative version is an explanatory gap, the positive version is "a genuine metaphysical stumbling block or apparent *impossibility*: How could you hope to produce a phenomenally unified, single-perspective, subject by assembling a group of subjects each of which essentially has its own perspective?" (Coleman 2016, 255, italics original).

This problem is also discussed by Roelofs (2019, sec. 2.3.2), in which he points out the possibility of a *fragmented zombie*, that is, one whose experience is not unified. He says:

the experience of a single composite subject must be unified in a way that contrasts with the apparent separateness of experiences had by different subjects. And yet if the whole gets one experience from one part, and another from another part, surely their experiences will be spread out across subjects rather than unified. And on the face of it this will be true whatever the character of those experiences, and however those parts are related. We know conscious unity as a within-subject relation, something connecting the many experiences within one subject's phenomenal field. We do not know how to see it as, or even relate it to, a between-subjects relation, holding between two or more different subjects. Hence, we cannot see what relation among component subjects and their experiences could suffice to unify them. (Roelofs 2019, 55)

In summary, the unity problem highlights the difficulty of establishing a panpsychist view in which the relations between microsubjects and their microexperiences can be thought in a way where the resulting macrosubject would have a unified macroexperience.

A related problem is the *boundary problem*, which Chalmers (2016) says that comes from Rosenberg (1998). The boundary problem is one that is not restricted to panpsychism, but can be generally described as the difficulty of explaining how it is that consciousness can have a clear boundary. In the context of panpsychism, it can be seen as the flip side of the unity problem. While in the unity problem there is a worry about the unification of the experience of microsubjects, in the boundary problem the question is about the relation between bounded microexperiences and their resulting bounded macroexperience.

Roelofs (2019) sees the boundary problem as being connected to another problem, in what he calls the *incompatibility contexts argument*. On example that use uses to explain it is:

consider the experience of seeing blue side by side with red; the red here appears as limited by and contrasting with the blue (and vice versa). If some part of me is experiencing the red alone, does it experience it the same way I do (as limited by and contrasting with another color), or not? If the former, surely that experience implies also experiencing the blue; if the latter, surely that is a

different experience of red from the one I am having. Either way, we do not get substantive experiential combination into a composite subject with unified consciousness. (Roelofs 2019, 58)

We can interpret Roelofs (2019) as pointing to the idea that if two microexperiential parts compose a macroexperiential whole, then the parts either: (1) lose their boundaries, which means that they have the same experience of the whole; (2) keep their boundaries, which means the experiences that the whole inherits from the parts are not the same as the parts themselves experience them. Both options seem to be problematic, which is why Roelofs (2019) describes this as a dilemma.

There is a final and related problem that is discussed by Roelofs (2019) in the context of what he calls the *privacy argument*, and it is the result of putting together some assumptions about composite subjectivity. The first assumption is that at least some experiences had by a macrosubject has to be also present in its component microsubjects. The second, that experiences are essentially subjective. And the third, that an experience that is essentially subjective can only belong to one subject, but not to other, i.e., it is private. The problem, then is explaining how it is that a microexperience can be private to a microsubject, at the same time as it is had by a macrosubject.

3.2.2. Qualities combination problem

This group of problems refers to the combination of phenomenal qualities, that is, of the qualitative aspects of experience. Under panpsychism, we can consider that microsubjects are experiencing microqualities. While the subjects combination problem focuses on the microsubjects, the qualities combination problem focuses on the microqualities. The simplest way of putting it is: how do microqualities combine to create macroqualities?

One specific version of this problem is what Chalmers (2016) call the *palette problem*. This problem is particularly pressing under Russellian versions of panpsychism, because in these the qualities of experience are taken to be the intrinsic nature of matter. But matter at a fundamental level seems to have only a limited set of properties that are expressed, for example, in a limited number of fundamental particles. So for Russellian panpsychism the intrinsic nature of those particles would be equality limited in number. The palette problem is: how is it that a limited set of microqualities can give rise to such a huge array of different macroqualities we experience?

Coleman (2016) considers the palette problem to be a bridging problem, because it only appears given the requirement of connecting microqualities with microphysical properties, which is done under Russellian panpsychism. He explains:

The reason there are only a few microqualities is that Chalmers makes the microqualities isomorphic with microphysical properties, of which there are apparently only a handful. So the problem is: we want to make microqualities isomorphic with microphysical properties; that means only a few microqualities, so how do they generate masses of macroqualities? With the tie to physics removed this problem would be considerably less impressive: without the limited repertoire of microphysical properties we'd have no reason not to indulge in masses of microqualities.

In contrast, there are versions of the problem of qualities combination that are internal. One such problem is what Coleman (2016) calls the *production problem*, and it is the problem of developing a model that explains how two different qualities could in principle combine. Coleman (2016, 252) gives the following example: suppose there are two instances of qualities, one of the color red and one of the color white. How could two separate instances of qualities give rise to a third instance of quality, for example of the color pink?

This also relates to another internal version of the problem, which Coleman (2016) calls the *contamination problem* and which he attributes to Lucretius. Coleman says:

So even if we can get past the production problem, and understand how separately instantiated qualities could interact to combine, we have an arguably more basic problem, of understanding how qualities that can combine actually do so. What does it even mean for two qualities to constitute a quality? Pink isn't red, and it isn't white either. One might expect that for red and white to survive in combination we would get as product *a patch of red alongside a patch of white*. (Coleman 2016, 252, italics original)

A further quality problem discussed by Coleman (2016) is the *incommensurability problem*, which he believes makes the contamination problem even harder. The problem is about explaining what kind of basic set of qualities could be general enough so that in one combination they would give rise to smells, in another to colors, and in yet another to pain. Such experiences seem too different to be just combinations of the same set of microqualities. Although similar to the palette problem, Coleman (2016) classifies the incommensurability problem as an internal problem, since—just as the production problem—it would persist even if we did not have any limitation in terms of the number of types of qualities.

3.2.3. Structures combination problem

The structural aspect of the combination problem is, as with the other aspects, the problem of explaining how the structure of microexperience can give rise to the structures of the

macroexperience. At first, this question might not seem problematic, since we already deal pretty well with the combination of microphysical structures into macrophysical structures.

However, there are two versions of this problem that need to be addressed by any successful panpsychist theory. The first is what Chalmers (2016) calls the *structural mismatch problem*. He says:

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? (2016, 183)

As we can see from the quote above, the structural mismatch problem is tied to the idea of a correspondence between the microphysical and the microexperiential structures. Just as in the case of the palette problem, that is why Coleman (2016) classifies this as a bridging problem. It is also why this problem is particularly pressing for the Russellian panpsychist view. Chalmers says: “Although the structural mismatch argument has received relatively little attention to date, I think it is one of the more powerful arguments against constitutive Russellian versions of panpsychism and panprotopsychism” (Chalmers 2016, 191).

The second version of the structural aspect of the combination problem is the *grain problem*, and it has been discussed both as a general objection to physicalism (Lockwood and Robinson 1993) as well as a specific objection to panpsychism (Goff 2006). In the context of panpsychism, it can be understood as the problem of explaining how an array of *discontinuous* microexperiences can give rise to a seemingly *continuous* macroexperience.

Chalmers (2016, sec. 7.4.6) discusses the grain problem relating it to what he calls the *revelation argument*. He points out the problem that, if the nature of consciousness is revealed to us by introspection, such nature does not seem to be a vast array of microexperiences. Instead, consciousness under introspection seems to be ‘smooth’ in a continuous macroexperience. Coleman (2016) considers the grain problem to be an internal equivalent of the structural mismatch problem, since it is a problem that does not require any tie between phenomenal and physical structures.

4. A holarchical solution to the combination problem

As I mentioned in the previous chapter, the combination problem has been the focal point of most of contemporary debate surrounding panpsychism. There is, however, a possible solution that I

believe has been largely ignored, which is the one discussed by philosopher Ken Wilber (2001; 2000b). Skrbina (2005) says that Wilber belongs to one of the five viable contemporary approaches to panpsychism, that of *part-whole holarchy*. He also says that “Wilber takes the basic system of Cardano/Koestler and marries it with elements of Teilhard, Plotinus, Spinoza, process philosophy, and various Eastern philosophers, producing an ontology which is explicitly panpsychist” (Skrbina 2005, 220).

Besides the possibility of not being that well known, there are two other reasons that could help explain why Wilber’s solution has not been given much attention. The first one is that his body of work is large and his philosophy covers many different areas of knowledge, as his main goal is an explicit attempt at giving rise to the integration of as many fields as he can. Because of this, even though a version of panpsychism is a fundamental part of Wilber’s ontology, it is not always the entry point for his discussions. Therefore, even among those who know about him it might be that his ideas about panpsychism are taken as a background.

The second reason is that, even though Wilber does discuss his version of panpsychism at some length, the discussion about the combination problem is much more restricted. As far as I can tell, there are only two places where Wilber talks about the combination problem: a long endnote in his book *Integral Psychology*, in which he speaks indirectly about it (Wilber 2000b, 176, note 15); and an addendum in his book *The Eye of Spirit*, but one that apparently was only added to later editions (Wilber 2001, 281). This, together with the first reason, may explain why it has not been brought into the contemporary discussion almost at all.

What I will do in this chapter is to try and explain how his views can help solve the combination problem. However, those same two reasons create some restrictions. On one hand, I will attempt to make the most out of what he explicitly says about the problem. On the other, because it is so limited, I will inevitably have to make inferences about questions that are not discussed by him. Those inferences will be informed by his other ideas, but because his work is so broad, it is not clear to what extent Wilber would actually endorse those inferences.

Furthermore, as I discussed in Section 3.2, there are many variations and developments of the combination problem, and they are among the issues that Wilber never discusses. In those cases, I will try to expand on his ideas in order to create a coherent picture of how the different versions could be dealt with based on his theory. This, of course, does not mean that it’s the way Wilber himself would do it. I will do my best to make it clear what is his direct view on the problem, what is being inferred from his view on other issues, and what is being expanded by me. But in the end,

my exposition will be of a position that is derivative of Wilber's, but in which the limits of originality are recognized as unclear.

4.1. Ken Wilber's panpsychism

In order to discuss Wilber's solution to the combination problem, we need to first understand some fundamental ideas of his philosophy. In particular, it is important to know about his ontology based on holons, how he relates subjectivity and objectivity, and how his panpsychism is different from other versions.

4.1.1. Holons and holarchy

A holon is a whole that is also a part. For example, a whole atom is a part of a molecule, and a whole molecule is a part of a cell. Atoms and molecules are both holons. This is a concept that comes from Arthur Koestler (1967), and one that Wilber uses as the basis for his ontology:

Reality as a whole is not composed of things or processes, but of holons. Composed, that is, of wholes that are simultaneously parts of other wholes, with no upward or downward limit. To say that holons are processes instead of things is in some ways true, but misses the essential point that processes themselves exist only within other processes. There are no things or processes, only holons. (Wilber 2000a, 43)

This is, at the same time, an intuitive but also radical view of the world. In the example I gave above, I talked about how atoms and molecules are holons. But in a sense, this is misleading, because it might seem as if holons are just a general name for the same physical entities we already had for example in a physicalist view. But Wilber is actually trying to avoid views such as physicalism, idealism and dualism, which is why he ends up in a type of panpsychism. He is explicit about that, saying: "This approach also undercuts the argument between the materialist and idealist camps. Reality isn't composed of quarks, or bootstrapping hadrons, or subatomic exchange; but neither is it composed of ideas, symbols, or thoughts. It is composed of holons" (Wilber 2000a, 43).

For the sake of simplicity, I will continue to use the physical entities as example of holons. But these should always be taken with the caveat that the basic feature of Wilber's ontology are holons, not physical entities.

Furthermore, holons are wholes that are part of holons, and not just wholes or parts. As Wilber puts it, "there are no wholes, and there are no parts. There are only whole/parts" (Wilber 2000a, 43). This has a nesting effect, where you get a hierarchy of holons, that is, a *holarchy*. And since all

holons are both wholes and parts, the nesting goes all the way up and all the way down⁷ (Wilber 2000a, 47). The importance of this will become clearer when we discuss how it helps solve the combination problem.

4.1.2. Interiors and exteriors

In Wilber’s theory, holons have four fundamental aspects that can not be reduced to each other, which he calls quadrants. These aspects are the result of two axes of differentiation, one of which is the interior-exterior axis, and the other the individual-collective axis. Combining these, we would have the four quadrants: the interior of the individual, the exterior of the individual, the interior of the collective and the exterior of the collective. To Wilber, these differentiations are close to a metaphysical necessity, because you can’t have the concept of an interior without an exterior, nor of an individual without a collective (Wilber 2000a, chap. 4; 2000b, chap. 5).

For now, we can focus on the interior-exterior distinction for the individual, which is the most relevant for the discussion at hand. This division can be roughly equated—but not completely, as we will see—with the distinction between subjectivity and objectivity, or between what is seen in a first person perspective and in a third person perspective.

4.1.3. Panpsychism or pan-depthism

Now, since for Wilber holons go all the way down, and every holon has both an interior and an exterior, the obvious conclusion is that interiors go all the way down, i.e., panpsychism. Wilber embraces the implication, but he prefers not to call his view panpsychism:

For me, consciousness in the broad sense is ultimately unqualifiable (Emptiness), and thus, although interiors go all the way down, no type of interior does. I am a pan-interiorist, not a pan-experientialist, pan-mentalist, pan-feelingist, or pan-soulist. The forms of the interior show developmental unfolding: from a fuzzy something-or-other [...] to prehension to sensation to perception to impulse to image to concept to rules to rationality and so forth, but none of those go all the way down in one specific form. (Wilber 2000b, 276)

Skrbina (2005, 221) says that Wilber misunderstands what is the common view of panpsychism when he says things like “the common panpsychism view [...] is that, for example, rocks have feelings or even souls, which is untenable” (Wilber 2000b, 280). On one hand, we can take at face value that Skrbina is correct in saying that this is not the common view of panpsychism,

⁷ Although to our purposes the most important direction is ‘all the way down’, the ‘all the way up’ direction can lead to some interesting connections and implications to other views such as cosmopsychism, as well to philosophical questions in fields such as logic.

and that maybe Wilber mischaracterizes it. On the other, Wilber seems to be saying two different things at the same time, both of which are more relevant to the discussion.

The first thing is part of the difficulties of panpsychism, and one that I mentioned earlier (Section 2.2): establishing which are the systems that do possess consciousness. The definition of panpsychism that Skrbina himself gives, as I also discussed in Section 2.2, is very open in this regard, saying that “a functional definition of panpsychism might be ‘All objects, or systems of objects, possess a singular inner experience of the world around them.’” (2005, 16).

But regardless of whether that is the common view of panpsychism or not, Wilber is clear when he says that such broad definition does not apply even if we use his understanding of interiors: “*Rocks as heaps have no interiors* (there is the inside of a rock, but that is just more exteriors); rocks, however, do contain atoms, which are holons, and those holons have one of the very lowest types of interiors” (Wilber 2000b, 280). The implication is that rocks are not holons, and that is why they do not have interiors.

This, in turn, raises the important question of how and when do we get a new holon, which is basically the same problem I mentioned previously about how ‘pan’ we want our panpsychism to be. How do we establish which systems are holons and, as a consequence, conscious? I will call this the *holon generation problem*. It is not strictly the combination problem, but it is very closely related to it. Skrbina says that “Wilber [...] draws the burden of explaining just how and when an interior appears in, say, a molecule of salt when one does not exist in the Na and Cl pair just before bonding, or, for that matter, how a new interior of a brain is created from the union of independent interiors of the neurons” (Skrbina 2005, 221). Part of my goal for this chapter is to discuss how Wilber’s view could deal with this issue.

The second thing Wilber seems to be saying is related to how he understands consciousness in the broadest sense:

What most panpsychists mean by consciousness or mind is *not* what I mean by consciousness, which is depth. Because consciousness is depth, it is itself literally *unqualifiable*. It is depth, not any particular, qualifiable level of depth (such as sensation or impulse or perception or intention)—those are all forms of consciousness, not consciousness as such. In other words, depth isn’t a quality, like sensation or impulse or idea, but a relationship (or opening) among holons. [...] I am a pan-depthist, not a pan-psychist, since the psyche itself emerges only at a *particular* level of depth. (Wilber 2000a, 567, chapter 4, note 2, emphasis original)

Here, Wilber uses the notion of depth to describe consciousness, which is why he describes himself as a pan-depthist. Because a holon is a whole/part that is composed of other holons, each

holon has a depth associated with it, in terms of ‘how many’ layers it contains. The point is not to say that each holon has a specific number of layers, but only that each holon has layers and, because of that, it has depth. As Wilber says in the quote above, depth is a relationship among holons and, as a consequence, so is consciousness.

In the earlier quote, Wilber also describes himself as a pan-interiorist. The relationship between depth and interiors is not perfectly clear. It seems to relate to an idea alluded in the quote above, where depth is an opening among holons. Wilber says that “interiors are *ultimately* unqualifiable (in my view, every interior is basically an opening or clearing in which correlative exteriors arise [...])” (Wilber 2000b, 277 italics original). We can say that depth and interiors are at least very closely related. This is an interesting topic to explore, but it is not strictly necessary for us at this point of the discussion. We now have enough to go to the combination problem.

4.2. Solving the combination problem

So far, I have discussed some of Wilber’s ideas that we need in order to tackle the combination problem. Recapitulating: (1) reality is made of holons, which are whole/parts composed of other holons; (2) holons have interiors and exteriors, where interiors are seen from ‘the inside’ and exteriors from ‘the outside’; (3) interiors (and holons) go all the way down, but no specific type of interior does; (4) consciousness or depth is a relationship among holons.

4.2.1. Setting the stage

Just for the sake of clarity, I am going to introduce a simple notation to talk about different aspects of different holons. It might seem an unnecessary complication at first, but I believe it will help make Wilber’s solution clearer and more precise. I will represent a holon by H. Each holon has an interior (I) and an exterior (E). Since we are focusing on interiors, we can make a further distinction within this context, between a subject (S) and an object (O) of experience.

Notice that I am slipping in the word ‘experience’ here, despite the fact that Wilber explicitly rejects the label of pan-experientialism. This is because in his view even experience could be taken as a form of interior and, as stated above, no form of interior goes all the way down. I will still use the word, however, because I believe it also might contribute to the clarity of the explanation.

There is also a subtle issue with the word ‘object’ here. The qualitative aspect of experience is at the core of the hard problem of consciousness, and what panpsychism attempts to deal with in the first place. What is given in the experience of a subject are those qualitative aspects, which here are

equivalent to what I'm calling the object of experience. Therefore, here the word 'object' is taken as having the most basic and fundamental meaning, that is, as the counterpart to the 'subject' of experience.

As a further notation, I will add a number representing the depth of the holon. For example H2 is a holon of depth 2, which means it is composed of holons of depth 1, that is, of holons H1. The number is not supposed to mean anything in particular, it is just a way of making the distinctions clear when we talk about holons at different levels of depth. I will do the same for the other aspects. For example, a holon H1 has an interior I1 and an exterior E1. And within the context of an interior I1, we have a subject S1 and an object O1 of experience.

4.2.2. Reinterpreting the problem

Now, recall that there are multiple versions of the combination problem. Putting in terms of holons, we have multiple holons at one level, let's say H1, that compose a holon at a higher level, H2. Looking from the outside, we could have for example that holons of the type H1 are atoms, and that they form a molecule as a holon H2. The combination problem in general is about relating the respective multiple interiors I1 to the interior I2.

But specifically, the combination problem with regards to subjects could be interpreted as the problem of explaining how the combination of multiple subjects S1 could lead to a subject S2. In turn, the combination problem with regards to qualities is explaining how the qualities of one level combine into the qualities of the level above it. Since here I am taking the qualities as equivalent to the objects of experience, it could be interpreted as the problem of explaining the combination of multiple O1 into O2. However, as I'll discuss in the next section, this way of interpreting the combination problem is the core issue that creates the problem in the first place, and what Wilber's theory tries to change.

4.2.3. Subject layering, not subject summing

Here is where the uniqueness of Wilber's solution comes into play, which he attributes to have its origin in developmental psychology and process philosophy. The solution to the combination problem is to change the way subjects and objects of different levels of depth relate to each other. Multiple H1 compose a H2, but multiple S1 **do not** combine to compose S2. Instead, the multiple S1 are *objects* for S2. That is, S1 has to be understood in terms of O2. In Wilber's words:

the combination problem is actually something that has been successfully handled (on the relative plane) for quite some time by developmental psychology and Whiteheadian process philosophy. In

essence, with each wave of development, the subject of one stage [i.e., level of depth] becomes an object of the next (as Robert Kegan would put it), so that each stage is a prehensive unification of all of its predecessors (Wilber 2001, 283)

The essential phrase is this: “the subject of one stage [i.e., level of depth] becomes the object of the next”. Throughout the whole discussion about the combination problem we have been trying to think about how subjects combine into subjects or how qualities combine into qualities. Some philosophers such as Coleman (2012; 2014; 2016) have defended that we could get macrosubjects directly out of microqualities, without appealing to microsubjects, but that would require a deflation of subjects.

Wilber’s solution, in contrast, keep both subjects and qualities at the macro and the micro level, but changes how they relate to each other. We need to remember that, for him, consciousness is a relation among holons. A holon H2 is conscious of its H1 parts, but in a chained manner. The subject S1 is an object of consciousness to S2. Again, the composition relation is held between holons, not between subjects nor between qualities. When multiple H2 combine to form a further holon H3, the same thing happens. S1 is an object for S2, which in turn is now an object for S3. Each holon keeps its subject, it’s just that this subject becomes an object for a larger holon.

As ingenious as this idea is, there are still many questions to be dealt with. This is where I have to start making significant inferences about Wilber’s solution. I want to make it perfectly clear that what follows is mostly my application of Wilber’s solution, supported by some of his other ideas. Wilber himself does not seem to address any of the specific combination problems, but only a general version as in the context of the quote I gave earlier. I think that his solution is broad enough to be able to encompass all the specific problems, and I am happy to give him due credit for it. But my intention is to try and expand his solution more explicitly, since he does not do so himself.

First, I have tried to be careful with singulars and plurals, but there is here a question of how multiple subjects of one level become a single object for a single subject of the level above. Wilber does not seem to explain exactly how this works, but the quote I gave earlier seems to give us an idea: “the subject of one stage [i.e. level of depth] becomes an object of the next (as Robert Kegan would put it), so that *each stage is a prehensive unification of all of its predecessors*” (Wilber 2001, 283, emphasis mine). This indicates that it would be the subject itself that unifies the multiple objects.

A single subject has a single unified experience, and that experience is a single object only in a broadest sense. Maybe it is possible that a subject could have the unified experience of multiple

objects without those objects having to be further unified into a single object. In fact, this possibility makes sense in this context, because the unification into a single object would actually happen at a further level above. For example, if multiple S1 are objects in an unified experience of S2, then they become a unified object when S2 itself becomes an object for S3. This also means that, at this point, the distinction between S2, O2 and I2 changes. That is because S2 and O2 are not different parts of I2, but S2 is I2 in relation to I1. And O2 is simply S1 in relation to S2. In other words, S2 is I2 as it “looks downwards”.

Second, there are still the questions raised by each different variation of the combination problem that I discussed earlier (Section 3.2). One of them is the subject summing problem, which is the problem of how subjects could sum into a further subject. In this solution, however, we are not summing the subjects. Maybe one could say that holons are “summed” into other holons—although that also could be problematic—, but no subject is being summed here. Under Coleman’s classification this is an internal problem about how our concept of subject could lead to the constitution of another subject. But Wilber’s solution entails changing the concept of a subject in relation to the subjects of its parts, which is why he bypasses the issue.

One interesting thing to think about is related to what I have already quoted from William James when he talks about summing subjects: “still each remains the same feeling it always was, shut in its own skin, windowless, ignorant of what the other feelings are and mean” (James 1890, 160). This idea of a subject shut in its own skin, windowless, brings up nicely the feature of Wilber’s proposal. If we have a subject S3 who experiences S2 as an object, what is it that S3 sees? Because H2 is composed of H1, S2 is experiencing S1. So S3 experiences S2 experiencing S1. S2 is still shut in its own skin, but now S3 sees ‘through’ it, and that includes S1. So S2 is windowless insofar H2 is seen from the outside. But when H2 becomes part of H3, S2 actually becomes a window through which S3 can see S1, as S1 is an object for S2.

Still, Coleman puts the subject summing as form of explanatory gap, because we don’t have an entailment from microsubjects to macrosubjects. This is closely related to the issue of conceivability: we can conceive of microsubjects without the existence of a further macrosubject. But this works differently under this holarchical solution. Let us sketch a simple argument. Take these premises: (1) the subject of a level is experienced as an object for the subject of a level above; (2) S1 is the subject within a holon H1, as S2 is within H2; (3) holons H1 compose a holon H2. Conclusion: S1 is experienced as an object for S2.

Premises (1) and (2) are part of Wilber's conceptual framework. These, of course, can be questioned. But the issue at hand is to deal with what Coleman called the internal problem of the subject summing, which is a conceptual issue that gives rise to an explanatory gap. Therefore, (1) and (2) can be taken simply as the necessary conceptual framework to solve the combination problem.

What remains as a possible problem is premise (3). Even if we accept Wilber's conceptual framework, there is still the question of how and when holons give rise to new holons, what I called the holon generation problem earlier, in Section 4.1.3. Maybe we can have a bunch of holons H1 come together and never compose a new holon H2, which would lead to a different conceivability problem. This is the same issue I touched upon previously of establishing which are the systems that are conscious, if we do not want to attribute consciousness to every possible system, and this again is tied to the holon generation problem. I will discuss this issue in more detail later, but here it is enough to notice that previously we had two problems: one about subject summing and one about which systems are conscious. Under Wilber's conceptual framework, the first problem collapses into the second, which means that now we have one fewer problem.

Another version of the combination problem regarding subjects is the unity one. Quoting Coleman again: "How could you hope to produce a phenomenally unified, single-perspective, subject by assembling a group of subjects each of which essentially has its own perspective?" (Coleman 2016, 255). This problem is dealt with in a similar way to what we have been discussing so far. The way to unify multiple subjects, each with its own perspective, is to make them objects of experience for a further subject. This further subject is tied to the corresponding holon, and therefore this problem too collapses into the holon generation problem.

Next on our list is what Chalmers (2016) calls the boundary problem and Roelofs (2019) relates to the incompatibility contexts argument. The boundary problem is the problem of how boundaries of wholes and parts relate, since each subject seems to have boundaries around its own experience. The incompatibility contexts argument is a dilemma between saying that either: (1) parts lose their boundaries and have the same experience as the whole; or (2) parts keep their boundaries, which means the experiences that the whole inherits from the parts is not the same as the parts experience them.

Once again, this problem is solved in Wilber's account by the relation between subjects of different levels. In a sense, Wilber chooses option (2) of Roelof's dilemma, because each holon and the corresponding subject keeps its boundaries. But that is not a problem because, as I said before,

when H1 composes H2 which composes H3, we have that S3 experiences S1 through S2. Which means that each subject gets to keep its boundaries at the same time as its experience is seen through by the subject of the level above it. As Wilber puts it, “at every stage ‘a single self is composed of many selves’” (Wilber 2000b, 281).

The last of the combination problems related to subjects is the privacy argument. This is the problem of explaining how a private subjective experience of a microsubject could at the same time be accessed by a macrosystem. As might seem clear at this point, the response also lies in how S3 experiences S2 experiencing S1. The experience of S2 is private in a sense, but at the same time S2 is wholly experienced by S3. That is, S3 experiences S2 as it experiences S1, which means that S3’s experience includes S1.

4.2.4. Qualities are not pixels

Now we get to the versions of the combination problem that are related to phenomenal qualities. The first of them is the palette problem, which is the problem of explaining how we could get a huge array of different macroqualities out of a very limited set of microqualities. This is, as Coleman (2016) says, a bridging problem tied to Russellian versions of panpsychism. I will discuss later to what extent Wilber’s position can be classified as Russellian, but we can assume it is in order to work through this problem. So let us say that there is a limited set of intrinsic properties at a level of holons H1. Any subject S2—that is, at the level of a holon H2—will only be able to see different combinations of S1, which has a limited set of intrinsic natures. But when we get to a level above, S3 is no longer limited to the set of S1, because now what S3 can experience is the whole new set of possible subjects S2. This points to a certain emergent aspect of Wilber’s view, which I will also discuss later, as it relates to the holon generation problem. But in any case, the fact is that by the time you get to holons of very high levels—humans, for example—the multiplicity of qualities to be experienced are no longer a problem.

The second version of the qualities combination problem is the production problem, which is about how two different qualities could in principle combine. This, in contrast, is an internal problem. However, the way it is solved in this context is basically the same as in the previous case. There is no actual combination, because all the different holons at a level H1—and their corresponding subjects S1—retain their own qualities, subjects and boundaries, as discussed previously. Take the case Coleman (2016, 252) discusses of two qualities, red and white, combining into a pink one. Here, if red and white are ‘qualities’ corresponding to two different holons H1, then pink could be a quality that corresponds to some level above it, for example H2. In a sense, pink

could be the quality of the subject S2 while it experiences red S1 and white S1. A pink S2, then, would be experienced by a further subject S3.

Third, we have incommensurability problem. This is the problem of having qualities that are general enough so that they could lead to very different modes of experience such as colors, smells and pain. Here, the exact same solution as before can be applied. Starting with fundamental particles at the bottom, by the time you get to human experiences, the level of depth is very high. So in the case of an auditory or visual system, for example, the layering of subjects—S1 experienced by S2 experienced by S3, etc—can be so deep that, even if we start with the same qualities at the bottom, the overall structure and relation between layers is so dissimilar that we could get to very different qualities of experience.

4.2.5. Structures from the inside and from the outside

Finally, this is the last group of combination problems, related to the structural aspect. Among those, we have first the structural mismatch problem, which is how to explain the difference between macrophysical structures and macrophenomenal structures. If the microexperiential is tied to the microphysical, why is it that the corresponding macro structures do not seem to align? The physical structure of a brain seems very different from the phenomenal structure of what a brain experiences. This is also a particularly pressing problem under Russellian panpsychism, and as I discussed earlier (Section 3.2.3), which is why Coleman (2016) classifies it as a bridging problem.

I will make use of a somewhat silly analogy to illustrate how this can be solved under Wilber's view. Bear with me. Let us say that on a first level, we have four lenses set up in a row, all facing down. On a second level, we have a bigger lens on top of the first two of the first level—covering them completely—and we have another lens on top of the last two of the first level. Then, on a third level we have an even bigger lens on top of the two lens of the second level, that is, covering the whole system. Now, if we look at the lenses from the side, they form a type of structure. However, if we look from the top, the structure seems different. That is because when we look from the side, we are looking at the lenses. But when we look from the top, we are looking through the lenses. Yes, in a sense we can say that we are also looking at the lenses from the top. But what we see is changed by the fact that we are now looking through them all, that is, *through their layering*.

When we look at a holon H3 from the outside, we see its exterior. That exterior has all the structural features that include its H2 parts, and their H1 parts in turn. It's the structure we could see, for example, by analyzing a cell, which is made of molecules, which are made of atoms. But

when we look at the interior of H3, we are looking from the inside. More than that, if we are looking at the interior of a holon, it means that *we are that holon*, in this case we are H3—or any holon above it. And as a subject S3, we are looking *at* S2, but we are also looking *through* S2 and *at* S1. And that is why the structure we see from the inside is not the same as the one we see from the outside.

The other version of the combination problem that is related to structure is the grain problem. This is the problem of explaining why the structure of macroexperience as revealed by introspection seems to be continuous, and not composed of a multitude of discontinuous microexperiences instead. This is the same issue as in the revelation argument (Chalmers 2016, sec. 7.4.6). If we apply Wilber’s view, this can be solved by using the solution to the previous structural problem, in addition to what I have discussed about a subject as that which unifies its objects (Section 4.2.3). A subject S3 experiences a multitude of S2 taken as objects, each in turn experiencing a multitude of S1 as objects.

The first thing to notice is that, even though S3 does experience a multitude of S1 in a sense, it is not in a direct unification of them, but through the intermediate unification given by S2. In the experience of a human being, we could have many intermediate levels of unification. These would ‘smooth out’ the experience, which means we do not directly experience an uncountable number of microexperiences, corresponding to all the relevant fundamental particles. The second thing to notice is what we discussed for the incommensurability problem, that there seems to be some stark difference between modes of perception at high levels of macroexperience. Auditory and visual perceptions are very different, to the point of seeming discrete. But that could be explained by the fact that the unification between both of them happens at a much higher intermediate level.

For example, say that a visual experience is built on a layering of holons up to a level H100. In addition, say that the auditory experience is also built on a layering of holons up to H100, but in a very different structure from the visual one. Because it took 100 levels of layering to get to both the full auditory and visual systems, they can be experiences of very different natures. So even if they are eventually unified under a single experience of a level above, the distinction between them would be very sharp. However, the distinction between the microexperiences at the level of fundamental particles would have been very smoothed out by all the intermediate layers.

4.3. Russellian or not?

As I discussed in Section 2.5.2, Russellian panpsychism is a way to handle the distinction between what matter *is* and what it *does*. Under a Russellian version of panpsychism,

microexperience takes the role of the intrinsic nature of matter. Now, does this apply to Wilber's view? Wilber (2000b) describes the relationship between interiors and exteriors in a way that seems quite different. As I quoted earlier (Section 4.1.3), he says: "interiors are *ultimately* unqualifiable (in my view, every interior is basically an opening or clearing in which correlative exteriors arise [...])" (Wilber 2000b, 277, italics original). This does not seem to be immediately about intrinsic properties, especially if we consider how he highlights the unqualifiable nature of interiors.

But if it is the interiors that allow the exteriors to arise, then those interiors could be understood as playing the role of an intrinsic nature, even if one that is unqualifiable. In forms of Russellian panpsychism, the intrinsic nature can be seen from within—the phenomenal—, while the extrinsic nature can be seen from without—the narrowly physical. And this lines up nicely with Wilber's concepts of interiors and exteriors, which follows the same pattern.

In addition, regarding the relationship among holons in general, and among subjects in particular, Wilber's conceptual framework also makes sense under a Russellian view. Consider the following: a subject S1 is seen as an object by a subject S2. But what is it that S2 is actually seeing? One possible answer could be that what S2 sees is precisely the intrinsic nature of S1. This, again, makes use of the same conceptual change as before, because S2 is not seeing its own intrinsic nature, but that of S1. In this sense, it seems perfectly reasonable to interpret Wilber's view as being Russellian, since it fits with other aspects of his theory.

However, this also introduces a possible problem. As I mentioned earlier (Section 2.4.1), many contemporary philosophers (e.g. Chalmers 2013; Miller 2018; Roelofs 2019) focus on constitutive Russellian types of panpsychism. That is because these two aspects, 'constitutive' and 'Russellian', go well together. If the micro level constitute the macro level, then we could say that the intrinsic nature of the micro level simply *is* the intrinsic nature of the macro level. But this does not seem to be applicable to Wilber's view.

Let us say that there is a holon H3 with its subject S3. Now suppose, given my interpretation above, that what S3 sees is not its own intrinsic nature, but that of S2. Likewise, S2 sees the intrinsic nature of S1, and not its own. This seems to imply that S1 and S2 have each their own distinct intrinsic nature. But just as H3 is composed of H2, we have that H2 is composed of H1. So how come we get two distinct intrinsic natures in S1 and S2? In order to get two intrinsic natures under the same holon, we would need some form of strong emergence. It is this question of emergence in Wilber's view that I will discuss in the next section, under the holon generation problem.

4.4. *The holon generation problem: constitutive or not?*

Recall that constitutive panpsychism is the view that microexperiences constitute macroexperiences, and it is usually contrasted with emergent panpsychism. Wilber's view brings a conceptual change—subjects of one level become objects for subject of the next level—, which is what makes possible the discussed solution to the combination problem. On one hand, this means that it would not be correct to interpret his view as a constitutive form of panpsychism, since microexperiences are not constituting macroexperiences. On the other hand, the same shift also blocks us from saying that macroexperiences emerge out of microexperiences. This is because we get macro holons out of micro holons, not macroexperiences out of microexperiences, as I will explain.

What is possible, however, is to use these terms to discuss the relation between holons at different levels, in what I have been calling the holon generation problem. A holon H2 is composed of holons H1, but is this a relation of constitution or of emergence? From what I have been hinting at different points, it seems that Wilber's view requires some form of emergence, and that is precisely the position that he takes:

[...] holons emerge. First subatomic particles, then atoms, then molecules, then polymers, then cells, and so on. The emergent holons are in some sense novel; they possess properties and qualities that cannot be strictly and totally deduced from their components; and therefore they, and their descriptions, cannot be reduced without remainder to their component parts. (Wilber 2000a, 54)

There are a couple of things that need to be noticed here. First, under non-Russellian—i.e. dualist—forms of emergent panpsychism, microphenomenal and microphysical properties run in parallel, which then raises the question of how the emergent macrophenomenal relates to the macrophysical.

But in Wilber's position this is not a problem. As I said, we do not have an emergence of macrophenomenal out of microphenomenal; nor do we have an emergence of macrophysical out of microphysical. What we have in Wilber's view is the emergence of macro holons out of micro holons. The phenomenal and the physical are seen as the interior and the exterior of holons. And the relation between interiors and exteriors is the same regardless of the level of the holon, so there is no separate emergence of macrophenomenal and macrophysical. Therefore, I think that Wilber's position could be considered as closer to a Russellian form of emergent panpsychism—and not a dualist one—, as long as we keep in mind the considerations I discussed in the previous section.

One of the possible problems with the move to emergent panpsychism is that it needs to appeal to a stronger type of emergence. This is in contrast to constitutive panpsychism, which only requires a weak type of emergence. The use of strong emergence, however, could erode the argument that lead us to panpsychism in the first place. If we are going to appeal to strong emergence, why not choose emergent dualism instead, and get consciousness out of non-consciousness? Brüntrup (2016, sec. 2.4), discusses this issue, saying that while this might seem to reduce the theoretical elegance of panpsychism, it does not necessarily mean that it defeats its purpose. He says: “The key question, of course, is which notion of strong emergence the non-constitutive [i.e. emergent] panpsychist will employ. Even strong emergence comes in different flavors” (Brüntrup 2016, 65).

The distinction among different flavors of strong emergence could give emergent panpsychism an advantage when compared to emergent dualism. With emergent dualism we get consciousness out of non-consciousness; with emergent panpsychism, we get macroconsciousness out of microconsciousness. So even if the emergent macroconsciousness has something truly above and beyond its microconscious parts, it seems like a shorter leap than what Strawson (2006) considers the brute emergence of getting the conscious out of the non-conscious.

Wilber’s view, in turn, has a similar advantage over emergent dualism, even if we have a strong emergence of holons. Besides, his position also has the advantage of being able to nicely solve all of the combination problems under a single conceptual framework, as I discussed. Appealing to emergence gives us the possibility of new holons, and allows us to solve some of the difficulties we were facing. However, it also creates new problems, as I will discuss next.

4.4.1. Causal efficacy and causal closure

When I discussed the argument for panpsychism (Section 2.4.2), it included the supposition of the causal efficacy of consciousness, that is, that we should consider consciousness as being causally relevant. One of the advantages of constitutive Russellian versions of panpsychism is that it is able to give some explanation of how macroconsciousness could be causally relevant. Since Russellian panpsychism takes microconsciousness to be the intrinsic nature of matter, by definition it is the very reason why matter is causally relevant. What matter *does* in terms of causal efficacy, is due to what matter *is*, i.e., its intrinsic nature.. And for constitutive panpsychism the macroconsciousness is constituted of microconsciousness, so one could argue that having causal

efficacy in microconsciousness would be enough to attribute causal efficacy to macroconsciousness⁸.

But there is another important question here, one about causal closure. I discussed earlier (Section 2.4.1) that in his defense of panpsychism, Chalmers (2013) uses a causal argument against dualism and in favor of physicalism as its antithesis. One of the premises of the causal argument is about the causal closure of physics: everything that is causally relevant for a physical event can be fully explained in physical terms. Of course, Chalmers uses this to get to panpsychism, so the idea that consciousness could be explained in physical terms is rejected. But remember that there is a distinction between the narrowly physical and the broadly physical, and only the narrowly physical can be described in physical terms. So the causal closure of physics is in fact what Chalmers (Chalmers 2013, 14) calls the causal closure of the narrowly physical. But there still remains the question of causal closure of the broadly physical.

The causal closure of the broadly physical is the idea that everything that is causally relevant for a broadly physical event is also broadly physical, regardless of whether it can be described by physics or not. But within the broadly physical, we can make a further distinction: the causal closure of the broadly microphysical. Since the broadly microphysical would include both the microphysical—as in narrowly microphysical—as well as the microphenomenal, I will simply call this *the causal closure of the micro*. Goff describes a similar idea: “Micro-level causal closure is the thesis that everything that happens has a sufficient micro-level cause” (Goff 2007, 11). I will, however, use the concept of ‘*causal closure of the micro*’ in a slightly different way, but that in the end serves the same purpose.

Notice that Goff says that the idea is that micro-level causes are sufficient for *everything that happens*. This means that it is sufficient for everything that happens at both the micro-level and the macro-level. But I want to define the causal closure of the micro more strictly: the causal closure of the micro is the idea that everything that is causally relevant *for the micro-level* is also at the micro-level.

If we take the macro-level to be nothing but the micro-level—as in constitutive panpsychism—, then this is not a problem and both definitions are equivalent. But under Wilber’s view, we are considering holons at the macro-level as something that newly emerges and is not completely

⁸ On one hand, this view could be accused of being too deflationary about the causal efficacy of macroconsciousness, since it would be nothing above the causal efficacy of microconsciousness itself. On the other hand, it lines up with the intuition of how the causal efficacy of a macrophysical entity such as a hammer could be understood as nothing above the causal efficacy of its parts.

reducible to the holons at the micro-level. And at the same time, we want the holons at the macro-level to have causal efficacy.

But these two ideas seem to be in conflict: (1) causal closure of the micro-level; (2) non-reducible causal relevance of the macro-level. If (1) is correct then there seems to be no role for any causal relevance as described by (2). How could the macro-level affect the micro-level, if there is causal closure at the micro-level? Maybe we could say that the causal relevance of the macro-level is limited to the macro-level itself, and therefore is independent of the micro-level. But this would simply lead us back to some form of emergent dualism, which we are trying to avoid. If we want micro and macro levels to be both causally relevant, while remaining ‘on the same side’ of the ontology, then we need a way to make sense of their causal relation. This question connects in general to the holon generation problem, but also specifically to the problem of establishing which systems are conscious and which are not. I will discuss this question in the next section.

4.4.2. Individual and collective: which systems are conscious?

At the beginning of this chapter I mentioned how Wilber has the idea of four quadrants, which are the result of distinctions made in two axes: interior-exterior and individual-collective. So far we have been focusing on holons as individuals, and dealing with the combination problem in those terms. But the individual is just one side of the axis. At this point, it is important to consider how the individual and the collective relate, because it has direct implications in the questions of holon generation, of causal efficacy and of which systems are conscious. However, to discuss the relation between individual and collective in depth could by itself fill a whole dissertation. I will, therefore, focus only on the aspects that are relevant for our purposes.

Wilber talks about individual holons and social holons, but not as a sharp distinction:

This distinction between an individual holon and its social holon (environment in the broadest sense) is not as easy to draw as it may first appear, however, because it’s almost impossible to define what we mean by an individual in the first place. [...] On the other hand, we do recognize that enduring holons possess a specific form or pattern, and this pattern is to some degree autonomous, or resistant to environmental obliteration. And this is usually what we mean by calling a holon an “individual”—we mean an *enduring compound individual*, compounded of its junior holons and adding its own defining form or wholeness (Wilber 2000a, 72 italics original)

If we take atoms as individual holons, then a group of atoms can be seen as a social holon at that level. Of course, ‘social’ here does not mean ‘human social’, but just as a collective of individuals. However, when two atoms come together and form a molecule, for example, that

molecule can now be understood as an individual holon, because it has a pattern that endures. But if both the individual and the social are holons, and we have been talking about panpsychism in terms of holons, would that mean that social holons are also conscious? Wilber says they are not:

I agree entirely with Leibniz/Whitehead/Hartshorne/Griffin that only the entities known as compound individuals (i.e., holons) possess a characteristic interior. Holons are different from mere heaps or aggregates, in that the former possess actual wholeness (identifiable pattern, agency, regime, etc.). [...] Heaps, on the other hand, are holons that are accidentally thrown together (e.g., a pile of sand). Holons have agency and interiors (every whole is a part, and thus every holon has an interior and an exterior), whereas heaps do not. A social holon stands between the two: it is more than a heap, in that its individuals are united by patterns of relational exchange, but it is less than an individual holon in terms of the tightness of its regime: social holons do not possess a locus of self-awareness [...]. (Wilber 2000b, 279)

In this quote Wilber is not always explicit, but it seems clear enough that he is talking about holons as individual holons, in contrast to social holons and in further contrast to heaps. So it is only individual holons that have a characteristic interior, i.e., are conscious. According to him, social holons are not conscious—do not possess a characteristic interior—, even though they are made up by conscious individuals. But now we can ask: what is it that makes one group of holons a further conscious entity, while other groups are not? In other words, how is it that a new individual—which is conscious—emerges out of a collection of holons that was previously a heap or a social holon?

This takes us back to Skrbina's criticism: "Wilber [...] draws the burden of explaining just how and when an interior appears in, say, a molecule of salt when one does not exist in the Na and Cl pair just before bonding, or, for that matter, how a new interior of a brain is created from the union of independent interiors of the neurons" (Skrbina 2005, 221).

By now we know that for Wilber it is not interiors that emerge out of interiors, but instead holons that emerge out of holons. Even though we are considering Wilber's view as somewhat Russellian—with interiors functioning as intrinsic nature—, he is explicit when saying that there are no interiors without exteriors, that they co-emerge. So the question is not about how we get new interiors, but about how we get new individual holons, which are conscious and therefore have characteristic interiors.

Wilber talks about the emergence of higher holons as being tied to the integration of lower holons in increasing complexity. He says: "Differentiation produces partness, or a new 'manyness'; integration produces wholeness, or a new 'oneness.' And since holons are whole/parts, they are

formed by the joint action of differentiation and integration” (Wilber 2000a, 75). Brüntrup (2016) takes a similar path⁹:

If the underlying levels of nature reach a certain threshold of complexity of configuration, then an emergent individual is likely to appear. The causal properties of the emergent entity will go beyond the summation of the causal powers at the underlying microlevel. There will be a genuinely new entity with new causal powers. (Brüntrup 2016, 66)

Here is the point that we are able to connect the emergence of new individuals with the problem of causal efficacy. A new individual is distinct not only because it is conscious, but also because it gives rise to new causal powers, in what is called *downward causation*. Having downward causation, here, means that the new individual has the ability to causally influence its own parts, a causal influence that is distinct from what the parts themselves already had.

In order to deal with the possibility of downward causation without breaking any causal relation at the micro level—e.g. the laws of physics—, Wilber’s solution is: “The lower sets the possibilities of the higher; the higher sets the probabilities of the lower” (Wilber 2000a, 61). He emphasizes that the causality at higher levels can not break any law governing the lower levels:

even though a higher level “goes beyond” a lower level, it does not violate the laws or the patterns of the lower level. It cannot be reduced to the lower level; it cannot be determined by the lower level; but neither can it ignore the lower level. [...] This is what is meant by saying that a lower sets the possibilities, or the large framework, within which the higher will have to operate, but to which it is not confined. (Wilber 2000a, 62)

Brüntrup makes a similar point:

If the basic rules are indeterministic such top-down influence might even happen without ‘breaking’ the most fundamental rules, only the probability distribution will be slightly affected. Such a system is conceptually coherent. If our world were like this, then it would contain strongly emergent entities with downward causal powers. (Brüntrup 2016, 66)

For that idea to work, we need to have a fit between the philosophical framework and the empirical restrictions that are given by science in general, and by physics in particular. But there seems to be room for such views, and one of the common ways of doing so is to appeal to an

⁹ Funnily enough, both Brüntrup and Wilber use the same quote from Whitehead when discussing these issues: “The many become one and are increased by one” (Whitehead 1978, 21). This might be a common influence that partly explain why they get to similar solutions to some of these issues, even though none of the two mention the other, as far as I am aware. But this adds to what I think is the unfortunate fact that Wilber’s ideas are not being explored in this context.

indeterministic aspect of quantum physics. Brüntrup points to the phenomenon of quantum entanglement as such a possibility:

Physics does not know of any viable procedure for reducing the entangled state to a summation of classical states and hence reducing quantum mechanics to classical physics. But if that is granted, then the properties of entangled atoms might well be causally efficacious for the future dynamics of the world. (Brüntrup 2016, 67)

In cases like this, for example, the entangled system of two particles can not violate the fundamental laws of physics guiding the particle themselves. Each particle has a limited set of possible states it can be in. However, the probability distributions of the particles depend on the entangled system as a whole. Again, as Wilber said: “The lower sets the possibilities of the higher; the higher sets the probabilities of the lower” (Wilber 2000a, 61).

In summary, what we have so far is that, when the complexity increases and you have enough integration among holons, there comes a point when a new individual holon emerges. That new individual holon has downward causal powers, which affect the probabilities of the lower level, without breaking any of the laws governing them. However, this is not a full answer to Skrbina’s criticism of how Wilber does not explain when it is that a new interior appears. There are two important considerations to be made.

On one hand, concepts such as ‘complexity’ and ‘integration’ can be vague, and therefore are able to hide a more significant gap in the explanation. Therefore, there is plenty of room for refinement in the philosophical discussion about what it means to say that a heap is different from a social holon, which is different from an individual holon. Wilber himself recognizes that this is a significant omission in his work (Wilber 2002).

On the other hand, the fact that we are dealing with strong emergence limits to what extent we can get a full account of the emergent holons, at least in terms of pure philosophical theory. There is a strong sense in which we can not derive the higher from the knowledge of the lower. This means that, although we must try to make the philosophical theory more precise, there is still a big part of the problem that is related to empirical research. What I have discussed about Wilber’s view gives us a conceptual framework, which in turn can be connected to empirical theories about the specifics of emergence. On this, Goff says:

if emergent entities have distinctive causal powers, then there will be an empirically discernible distinction between those systems that have and those systems that lack such emergent causal powers. The behavior of the latter but not the former will be predictable on the basis of the behavior of the system’s parts. (Goff 2016, 296)

One possible source of a connection between these aspects—integration, downward causation and empirical research—, is with the *Integrated Information Theory* (IIT), which is usually associated with the work of Giulio Tononi (Hoel, Albantakis, and Tononi 2013; Oizumi, Albantakis, and Tononi 2014; Tononi and Koch 2015; Tononi et al. 2016). Brüntrup explicitly mentions IIT as one of the ways to make downward causation to work under panpsychism (Brüntrup 2016, 67), and Chalmers even classifies IIT itself as a form of emergent panpsychism (Chalmers 2016, 193). In the next section I will explore a possible relation between IIT and Wilber’s view.

4.5. *Integrated Information Theory, externalism and vanishing subjects*

IIT is a theory of consciousness that has gained some attention in the last year due not only to its attempt at a precise philosophical formulation, but also because of its associated empirical research. Some of its authors explain:

IIT addresses the hard problem in a new way. It does not start from the brain and ask how it could give rise to experience; instead, it starts from the essential phenomenal properties of experience, or axioms, and infers postulates about the characteristics that are required of its physical substrate. Moreover, IIT presents a mathematical framework for evaluating the quality and quantity of consciousness. (Tononi et al. 2016, 450)

I will not go into the details of the theory, but there are some elements that are interesting to highlight in the context of our current discussion. The first element is what they mention as the mathematical framework for evaluating consciousness. The specifics of that framework has changed—the theory in its third version (Oizumi, Albantakis, and Tononi 2014)—, but the core idea is that there is a quantity associated with the level of integration of information, which is denoted by the Greek letter Φ ¹⁰.

Under the right circumstances, the level of integration of information in a system is associated with the level of consciousness of the system. Those right circumstances are the second element that is important for us. For IIT, in order to know whether a system is conscious or not, it is not enough to look at the system itself; we also need to look at all overlapping systems. Only the system with maximal Φ among all the ones that overlap has a corresponding consciousness associated. This means, for example, that if we have a system A within a system B within a system C, then system B is conscious only if neither A nor C has a higher level of integrated information—nor any other partially or fully overlapping system.

¹⁰ There are actually finer distinctions among different quantities in the theory, but for our purposes it is enough to put all of them under the same name.

The problem with this view is that it creates a counter-intuitive result: consciousness becomes an extrinsic property. Whether system B is conscious is determined not just by that which is within B—e.g. system A—, but also by what lies beyond its limits—e.g. system C. This is a problem that was briefly pointed out by Chalmers (2016, 212, note 8), and that has been further discussed by Mørch (2019). Mørch explains how the problem is mostly a result of one of the postulates of IIT, the *exclusion postulate*. This, in turn, is associated with a corresponding *exclusion axiom*:

Consciousness is exclusive: each experience excludes all others – at any given time there is only one experience having its full content, rather than a superposition of multiple partial experiences; each experience has definite borders – certain things can be experienced and others cannot [...]
(Oizumi, Albantakis, and Tononi 2014, 3)

This notion of exclusion means that, in the example I gave above, the consciousness of systems A, B and C exclude each other. Only one of them can be conscious. But now let's say that initially A had maximal Φ , and was therefore conscious. Then the integration of B increases until it becomes larger than A's. Since A is within B, the moment that B becomes conscious, A is no longer conscious. A was initially a subject, but now it no longer is, even though A has not changed.

The exclusion axiom has some intuitive appeal, since it seems strange to consider that two overlapping systems could be conscious of some shared experience at the same time. But notice that this is precisely what I have been discussing under Wilber's solution to the combination problem. On one hand, we do not have two subjects having directly the same experience, as I discussed for the boundary problem, for example.

On the other hand, when an individual holon H1—a conscious system—becomes part of another individual holon H2, there is no need for H1 to stop being conscious in order for H2 to become conscious. H1 retains its subjectivity in S1; it is just that S1 is now an object for S2. This is a direct result of Wilber's conceptual understanding of consciousness as a relation among holons. When a higher holon emerges, also does its consciousness. But the lower holons do not lose their own consciousness; they just enter into a new relation with the higher holon.

It seems to me that IIT can give us some specificity and empirical support to the idea of integration as leading to the emergence of individual holons, which then have downward causation. But Wilber's conceptual framework can provide a better understanding of how holons of lower and higher levels can relate to each other in terms of consciousness, subjects and experience. With that, it can help us avoid problems such as the externalism of IIT, for example. I believe there are interesting and potentially fruitful paths to be pursued along those lines of inquiry.

4.6. *What lies ahead*

I want to close this chapter by discussing some possibilities of how to move forward based on these ideas. The first possibility is that Wilber's view could help us solve what Chalmers (2018) recently called *the meta-problem of consciousness*. Chalmers explains:

The meta-problem proper, however, is the problem of explaining problem intuitions: intuitions that reflect our sense that there is some sort of special problem involving consciousness, and especially some sort of gap between physical processes and consciousness. For example, 'I can't see how consciousness could be physical' is a problem report, and the disposition to judge and report this is a problem intuition. (Chalmers 2018, 12)

In other words, the meta-problem is about understanding why we think that there is a problem of consciousness in the first place. I believe that Wilber's ideas—e.g., that subjects of consciousness are not summed, but layered—can provide some insights around this question. One way of approaching the meta-problem could be by investigating how it is that we can take both first and third person perspectives. This can be helpful because the hard problem of consciousness is tied to the tension between these perspectives. And it seems that any theory of consciousness should at least help explain them, since they are perspectives that are taken by conscious beings such as ourselves.

Wilber connects much of his theory to developmental psychology, where there is a lot of theoretical and empirical work regarding the development of perspectives, in both children and adults. So let us consider the possibility, for example, that when we develop as complex holons we gain more and more depth, and with each level of depth there are new capacities. If taking a first and a third person perspective are capacities that we gain at some level of complexity, then it is conceivable that, with more complexity, there comes a point when we are able to see how those perspective are at odds with each other. It is important to notice that I am not talking about having interiors or exteriors, since those are present at any level. I am talking about the cognitive capacity of taking those perspectives.

Chalmers (2018) talks about distinct categories of intuitions regarding the meta-problem of consciousness. From those, he says that at the core of the problem are *explanatory intuitions*, and closely related are *metaphysical intuitions*. The former is about the difficulty of connecting explanations of physical processes with those about consciousness. The latter is about the somewhat dualistic intuition that consciousness is non-physical, even fundamental.

One possible way of clarifying such intuitions is by better understanding what it means to take a first and third person perspective. This is not part of Wilber's view, but I believe that it is compatible. First, let us define perspectives in the following convoluted way: a first person perspective is the perspective of which the objects of knowledge are *necessarily* part of the person taking the perspective; a third person perspective is the perspective of which the objects of knowledge are *not necessarily* part of the person taking the perspective. For example, an experience such as a thought is an object of knowledge—we know *about* the thought, *about* the experience—, one that is revealed in a first person perspective. As an experience and a thought, it is necessarily part of the person taking the perspective. In contrast, an atom is an object of knowledge—we know *about* the atom—, one that is revealed in a third person perspective. The atom is not necessarily part of the person taking the perspective, since it could be known regardless of being part of that person.

This is a simple definition of first and third person perspectives, but it is one that seems to fit with our intuitions. But more importantly, it is one that immediately creates an exclusive disjunction between the two sets of objects of knowledge. Either it will be an object of knowledge of a first person perspective—necessarily a part of the person—, or it will be an object of knowledge of a third person perspective—not necessarily part of the person.

If this definition holds, then we can see why we could expect the explanatory intuition of the meta-problem of consciousness. And if we take a realistic stance towards those objects, then we also have the metaphysical intuition, since we can not see how the two types of objects—i.e. consciousness and physical processes—could exist under a single category.

Now, going back to Wilber, if such capacities of taking first and third person perspectives are capacities that a holon gets at a certain level of complexity, this means that they are available for the subject at that level. Now let's say that the complexity increases even more, and that subject is then seen as an object for a further subject. Then it could be the case that this further subject sees the incompatibility between the perspectives. And as the higher subject sees this incompatibility, the problem of consciousness arises, which is what the meta-problem is asking about.

Besides the meta-problem of consciousness, there is another path that I believe can be further explored. When Goff (2016, sec. 11.7) discusses the question of what conditions make it so that objects combine into a further object, he says that there are two options. The first, *unrestricted composition*, is the idea that any set of objects, no matter how unrelated, compose a further object. In this case, there is even an object that is composed of: my computer, the moon and Wilber's left

toe. The second option, *restricted composition*, is the idea that some sets of objects form a further object, while others do not. Goff says:

It is likely, then, that the emergentist will support some form of restricted phenomenal composition, looking to the empirical facts for the boundary between systems that are mere aggregates of micro-level subjects and systems that are conscious in their own right. (Goff 2016, 296)

There is indeed a sense in which an empirical restricted composition is relevant, as I pointed out. However, I think that there is the possibility of a hybrid position. I have also discussed how, for Wilber, there is a distinction between a heap, a social holon and an individual holon, but that those are not sharp distinctions. In addition, I have briefly mentioned (Section 4.1.1) that for Wilber holons go not just all the way down, but also all the way up. Although Wilber does not seem to speak explicitly in these terms, it seem reasonable to understand that since holons are whole/parts that go all the way up, then in some sense there is an unrestricted composition of holons. Every holon it not only a whole, but also a part of another holon. But that holon, in turn, also is a part of another holon, and so on and so forth.

This does not mean that there would be an unrestricted composition for individual holons—i.e., holons possessing a characteristic interior and downward causal powers. The emergence of individual holons would still need to be empirically discovered. Instead what we would have is that holons as such could fall under an unrestricted composition, while individual holons would fall under a restricted composition.

These considerations are obviously very tentative and speculative. But if the recent rise of interest in panpsychism has showed us something, it is that initially counter-intuitive ideas can be very fruitful and worth exploring.

5. Conclusion

Panpsychism is not a single unified theory, but a group of theories. Still, we can define a coherent general thesis of panpsychism as the view that fundamental entities has some form of experience. Panpsychism is can be connected to broadly physicalism—everything made of the same ‘stuff’ as tables and chairs—, but rejects narrow physicalism—physics covers all the truths about reality. If to the rejection of narrow physicalism we add the rejection of dualism and of brute emergence, then get to panpsychism as viable position.

The current biggest challenge to panpsychism is the combination problem. This is the problem of explaining how we could get higher order consciousness out of consciousness at fundamental entities, even if we assume that panpsychism is true. We can distinguish three aspects of the combination problem. First, the combination of subjects: it seems difficult to conceptualize how we could put together a group of microsubjects and get as a result a macrosubject. Second, the combination of qualities: the qualitative aspect of experience at the human level does is very broad and complex, and does not seem to be just a combination of very simple qualities. Third, the combination of structures: human consciousness seems to be associated with human brains, but even if fundamental entities are conscious, the resulting combination into the structure of brains do not match the structure of experience.

Ken Wilber has proposed a solution to the combination problem that has received little to no attention, but one that I think has the potential to dealing with all the versions of the problem. In his view, the world is made of holons, which are wholes/parts. His solution to the combination problem involves rethinking the way the subjects of lower and higher level holons relate to each other. For him, the subject of one level becomes the object for the subject of the next level. So indeed subjects do not combine, but instead they are layered. Wilber's view require a form o strong emergence, and this allows higher level holons to gain downward causal powers, which in turn makes consciousness causally relevant. Despite the fact that strong emergence can lead to problems of causal closure, I believe there are ways around those problems, and that Wilber's view should be considered as an important step in solving the combination problem, and ultimately the hard problem of consciousness.

6. Bibliography

- Brüntrup, Godehard. 2016. "Emergent Panpsychism." In *Panpsychism: Contemporary Perspectives*, edited by Godehard Brüntrup and Ludwig Jaskolla, 48–71. New York: Oxford University Press.
- Brüntrup, Godehard, and Ludwig Jaskolla, eds. 2016. *Panpsychism: Contemporary Perspectives*. New York: Oxford University Press.
- Chalmers, David. forthcoming. "Idealism and the Mind-Body Problem." In *The Routledge Handbook of Panpsychism*, edited by William Seager. Routledge.
- . 1995. "Facing up to the Problem of Consciousness." *Journal of Consciousness Studies* 2 (3): 200–219.
- . 1997. "Moving Forward on the Problem of Consciousness." *Journal of Consciousness Studies* 4 (1): 3–46.

- . 2013. "Panpsychism and Panprotopsychism." *The Amherst Lecture in Philosophy*, no. 8: 1–35.
- . 2016. "The Combination Problem for Panpsychism." In *Panpsychism: Contemporary Perspectives*, edited by Godehard Brüntrup and Ludwig Jaskolla, 179–214. New York: Oxford University Press.
- . 2018. "The Meta-Problem of Consciousness." *Journal of Consciousness Studies* 25 (9–10): 6–61.
- Coleman, Sam. 2006. "Being Realistic: Why Physicalism May Entail Panexperientialism." *Journal of Consciousness Studies* 13 (10–11): 40–52.
- . 2012. "Mental Chemistry: Combination for Panpsychists." *Dialectica* 66 (1): 137–66.
- . 2014. "The Real Combination Problem: Panpsychism, Micro-Subjects, and Emergence." *Erkenntnis* 79 (1): 19–44.
- . 2016. "Panpsychism and Neutral Monism: How to Make Up One's Mind." In *Panpsychism: Contemporary Perspectives*, edited by Godehard Brüntrup and Ludwig Jaskolla, 249–82. New York: Oxford University Press.
- Goff, Philip. 2006. "Experiences Don't Sum." *Journal of Consciousness Studies* 13 (10–11): 53–61.
- . 2007. "Panpsychism." In *The Blackwell Companion to Consciousness*, edited by Max Velmans and Susan Schneider. New York: Wiley-Blackwell.
- . 2009. "Why Panpsychism Doesn't Help Us Explain Consciousness." *Dialectica* 63 (3): 289–311.
- . 2016. "The Phenomenal Bonding Solution to the Combination Problem." In *Panpsychism: Contemporary Perspectives*, edited by G. Bruntrup L. Jaskolla, 283–302. New York: Oxford University Press.
- Hoel, Erik P., Larissa Albantakis, and Giulio Tononi. 2013. "Quantifying Causal Emergence Shows That Macro Can Beat Micro." *Proceedings of the National Academy of Sciences* 110 (49): 19790–95.
- Jackson, Frank. 1982. "Epiphenomenal Qualia." *The Philosophical Quarterly* 32 (127): 127–136.
- James, William. 1890. *The Principles of Psychology*. Read Books Ltd.
- Koestler, Arthur. 1967. *The Ghost in the Machine*. Macmillan.
- Levine, Joseph. 1983. "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* 64 (4): 354–361.
- Lockwood, Michael, and Howard Robinson. 1993. "The Grain Problem."
- Miller, G. E. 2018. "The Combination Problem for Panpsychism: A Constitutive Russellian Solution." Phd, University of Liverpool.
- Mørch, Hedda Hassel. 2014. "Panpsychism and Causation: A New Argument and a Solution to the Combination Problem." PhD Thesis, University of Oslo.
- . 2019. "Is Consciousness Intrinsic?: A Problem for the Integrated Information Theory." *Journal of Consciousness Studies* 26 (1–2): 133–162.

- Nagel, Thomas. 1974. "What Is It like to Be a Bat?" *The Philosophical Review* 83 (4): 435–450.
- Oizumi, Masafumi, Larissa Albantakis, and Giulio Tononi. 2014. "From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0." *PLOS Computational Biology* 10 (5): e1003588.
- Roelofs, Luke. 2019. *Combining Minds: How to Think about Composite Subjectivity*. Oxford University Press.
- Rosenberg, Gregg H. 1998. "The Boundary Problem for Phenomenal Individuals." In *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*, edited by Stuart R. Hameroff, Alfred W. Kaszniak, and Alwyn Scott. MIT Press.
- Seager, William. 1995. "Consciousness, Information and Panpsychism." *Journal of Consciousness Studies* 2 (3): 272–288.
- Seager, William E. 2019. *The Routledge Handbook of Panpsychism*. Routledge.
- Skrbina, David. 2005. *Panpsychism in the West*. Cambridge, MA: MIT Press.
- Strawson, Galen. 2006. "Realistic Monism: Why Physicalism Entails Panpsychism." *Journal of Consciousness Studies* 13 (10–11): 3–31.
- Tononi, Giulio, Melanie Boly, Marcello Massimini, and Christof Koch. 2016. "Integrated Information Theory: From Consciousness to Its Physical Substrate." *Nature Reviews Neuroscience* 17 (7): 450–61.
- Tononi, Giulio, and Christof Koch. 2015. "Consciousness: Here, There and Everywhere?" *Philosophical Transactions of the Royal Society B: Biological Sciences* 370 (1668): 20140167.
- Whitehead, Alfred North. 1978. *Process and Reality: An Essay in Cosmology*. Edited by David Ray Griffin and Donald W. Sherburne. New York: Free Press.
- Wilber, Ken. 2000a. *Sex, Ecology, Spirituality: The Spirit of Evolution, Second Edition*. Shambhala Publications.
- . 2000b. *Integral Psychology: Consciousness, Spirit, Psychology, Therapy*. Shambhala Publications.
- . 2001. *The Eye of Spirit: An Integral Vision for a World Gone Slightly Mad*. Shambhala Publications.
- . 2002. On Critics, Integral Institute, My Recent Writing, and Other Matters of Little Consequence: A Shambhala Interview with Ken Wilber. Shambhala Publications Website. Version from The Internet Archive at <https://web.archive.org/web/20020618222815/http://wilber.shambhala.com/html/interviews/interview1220.cfm/>. Accessed on April 29, 2020.

Abstract

The hard problem of consciousness is the problem of explaining the existence of subjective experience in a world that is purely physical. As the usual theories fail to provide an adequate explanation, many philosophers started looking for alternatives. One of such alternatives is panpsychism, which has been increasingly gaining attention in the past few decades. Panpsychism is the view that fundamental entities of the world—fundamental particles, for example—have some form of very simple consciousness. And, in the theory, it is that spark that allows for the existence of more complex consciousness, such as in humans.

However, panpsychism faces a major difficulty with what is called ‘the combination problem’. This is the problem that, even if we accept that fundamental entities have some form of simple consciousness, it is not clear how one could go from a plurality of micro consciousness to a complex unified macro consciousness. If we think that an atom is conscious, for example, this seems to imply that it is a subject, a conscious subject. But how do we get to a human subject, with a complex human brain? It seems difficult to understand how putting together many conscious subjects—e.g. atoms—could lead to a further conscious subject—e.g. a human.

In this thesis I discuss panpsychism and some of the arguments that support it. In addition, I explain the combination problem and its different aspects, as well as how it relates to panpsychism and the hard problem of consciousness. Finally, I discuss a solution to the combination problem that I believe has not been given proper attention in contemporary philosophy, which is one provided by the philosopher Ken Wilber. His solution involves changing the conceptual framework of how we understand the relation between macro and micro subjects. I also apply his theory to the variations of the combination problem, discussing how we could approach these issues under a single coherent view.

Thesis word count: 21740.