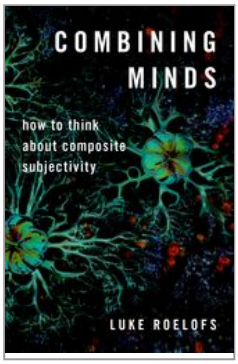

Combining Minds: How to Think about Composite Subjectivity
Luke Roelofs

Print publication date: 2019
Print ISBN-13: 9780190859053
Published to Oxford Scholarship Online: February 2019
DOI: 10.1093/oso/9780190859053.001.0001

Composite Subjectivity and Microsubjects

Luke Roelofs

DOI:10.1093/oso/9780190859053.003.0003

Abstract and Keywords

This chapter is about how to combine the very simple subjects of experience posited by panpsychism, the theory that matter itself is inherently conscious. The combination problem has been most thoroughly discussed in relation to the combination of these microsubjects, and this chapter addresses head-on the two central strands of the combination problem: the subject-summing problem and the problem of the unity, and boundaries, of consciousness. Alternative solutions, including cosmopsychism (the world as a whole is conscious) and panprotopsychism (matter is not conscious, but contains some sort of germ of consciousness) are also discussed. The metaphysics of nature that results from addressing these challenges is both highly counterintuitive and theoretically elegant.

Keywords: panpsychism, cosmopsychism, panprotopsychism, combination problem, unity of consciousness, subject of experience, metaphysics

IN THIS CHAPTER I consider combinationism in relation to what the universe is like at the most basic level, independently of the specifics of the human condition. Going by what modern science seems to reveal, the universe consists in a vast number of absolutely tiny things that can be called “particles,” possessing a few fundamental physical properties, standing in a few fundamental physical relations, and forming aggregates of various sizes which inherit these physical properties and whose structure is determined by these physical relations.

If constitutive Russellian panpsychism (CRP) is true, conscious experience should somehow fit into this world of particles: like their fundamental physical properties, it should be widely distributed and indifferent to the special features of any earthly ape species. More precisely, the fundamental physical things are conscious subjects (“microsubjects”), the fundamental physical properties and relations correspond with experiential properties and relations, and complex consciousness like ours is explained by the combining of these incredibly simple subjects with their incredibly simple experiences.

In this chapter, I sketch and defend a panpsychist theory on which experiential properties work like fundamental physical properties. As well as assuming the **(p.76)** truth of CRP, I will assume that subjects are substrates of experience as opposed to personas, that wholes are aggregates rather than structure-specific, that consciousness is a fundamental property, and that phenomenal unity is a fundamental relation, not reducible to some sufficient density of other sorts of conscious unity. In the coming chapters I will explore different starting points, which generate alternative, complementary theories of mental combination.

The result—panpsychist combinationism—is a combinationist theory that is simple and elegant but deeply counterintuitive in some of its implications. Section 3.1 presents an overview of panpsychist combinationism, identifying how it relates to chapter 2’s arguments against combinationism. The two key claims are experience inheritance (the principle that aggregates inherit the experiential properties of their parts) and the micro-unity hypothesis (the suggestion that one of the fundamental physical relations establishes phenomenal unity). Sections 3.2 and 3.3 then clarify, explore, and defend these claims, and section 3.4 presents a further argument in support of experience inheritance, with the micro-unity hypothesis as a premise.

The most controversial claim made in this chapter is that experience inheritance is true *a priori*, and so my final sections explore the reasons for and against accepting its *a priori* status, and some ways to preserve the spirit of panpsychist combinationism if this status is denied. The next chapter then addresses certain additional problems facing CRP, based on the apparent discrepancy in structure between physical reality and human consciousness.

3.1. A Sketch of Panpsychist Combinationism

I begin my sketch of panpsychist combinationism by supposing that microscopic entities have experiences to combine. More specifically, I suppose that all elementary particles are associated with incredibly simple experiences, whose structure is no more complex than the structure of those particles’ physical properties. Since they are the metaphysical substrates of these experiences, they will count as conscious subjects according to the metaphysical conception of subjects—they are microsubjects.

Admittedly, it is unclear how far contemporary physics supports the fundamentality of particles. Consider the apparent failure of permutations to differ: in the probabilistic mathematics of quantum physics, the outcomes "particle-A in location 1 with X properties and particle-B in location 2 with Y properties" and "particle-B in location 1 with X properties and particle-A in location 2 with Y properties" must be treated as a single outcome (French and Rickles 2003). Or consider the possibility of indeterminate particle-number: in some cases the number of particles in **(p.77)** a system depends, like velocity, on what frame of reference we use to consider it (Domenech and Holik 2007). There are also much older metaphysical questions about the relative fundamentality of particles and space or spacetime. The world might be infinitely divisible, displaying no identifiable "basic parts," or a fundamental whole more basic than any of its parts. For all we know, the fundamental physical entities might be particles, fields, points, spacetime, or something else.

However, these considerations can largely be set aside. They amount to saying that a certain model we are familiar with—of independent, recombinable, countable, more or less solid parts like Lego blocks or the grains of sand in a sandcastle—may not extend all the way down. But that model does extend pretty far down! Cells, molecules, and even atoms are ontologically independent of each other and can be determinately counted, tracked, and recombined, etc. If there is a level where these parts are no longer so "well-behaved," that does not change the fact that human bodies and everything else we encounter are made up of discrete parts on the scales from meters down to nanometers. If subatomic particles are in some sense unreal, my remarks will apply *mutatis mutandis* to whatever the simplest and most basic physical entities are, if any, and moreover will still apply to the comparatively well-behaved atoms, molecules, cells, and so forth.

3.1.1. Microsubjects and Their Experiences

What sort of experiences do microsubjects have? There is probably no way to positively answer that question at the current stage of human knowledge. The best we can do is identify constraints on what kind of experiences they *might* have. On the one hand, the physical simplicity of a given microsubject puts a "ceiling" on how far we can attribute complex experiences to it. Partly this is because of Russellianism: if the basic experiential properties are the real inner nature of the basic physical properties, then the two should line up with one another. Partly it is also about making sense of the extremely simple behavior of the basic physical entities: we should not ascribe to them a conscious life which would lead us to expect more intelligent behavior than they display. On the other hand, we can assign a "floor" by reflecting on the explanatory challenge posed by consciousness. Whatever it is about consciousness that cannot be explained in terms of anything else should be recognized as a fundamental feature of matter. Of course, it is a controversial question what it is about consciousness that is hard to explain. A lot of discussion has focused on sensory qualities, like

experienced redness or pain, and if we thought that these and nothing else produce an “explanatory gap,” we should conclude that microexperiences must involve some sort of sensory quality, but perhaps nothing else. Alternatively, we might think that *meaning*, the capacity to **(p.78)** mentally represent things, to be “about” some object, is just as hard to account for in physical terms as sensory qualities are. In that case we should suppose that the basic experiences also have some rudimentary seed of meaning; they are “about something,” though that something might be as simple and unspecific as “the world” or “this stuff.” If we thought that a sense of the “goodness” and “badness” of certain feelings was likewise primitive, we should include that too at the fundamental level; similarly for will or motivation, or anything else that reflection suggests is not analyzable as some configuration of more basic features.

Some might worry that an expansive conception of the irreducible aspects of consciousness might raise the “floor” so high that no space is left between it and the “ceiling”: human consciousness might be explainable only by a sort of experience that is much more complicated than anything a fundamental particle could plausibly undergo. But I do not think this is true. Here is a speculative example of what it might be like to be a fundamental particle, which includes all the above-mentioned basic features but is compatible with extreme simplicity of physical properties and behavior. Suppose the entire content of this phenomenology is a contrast, between a figure and a background, a “something” and a “something else.” The former aspect of the experience represents the particle’s surroundings insofar as they exert force on it; the latter aspect refers to the rest of the world in general.[1] The particular quality that the “something” is experienced as having will vary depending on the particular way that the particle’s surroundings act on it (degree of force, type of force, etc.), but beyond that these representations are maximally unspecific, mere inchoate gesturings outward. This experience is not strictly visual, or tactile, or emotional, or cognitive: it is poorer and less determinate than any of these. But it has an internal motivating power, rather like pleasure or displeasure. It motivates the electron to move blindly toward or away from the source of the force it is feeling, in virtue of the way that it feels.[2] Moreover, because the experience involves a contrast, it contains an implicit sense of negation—that the “something” is not the “something else.” And philosophers who think consciousness always inherently involves a special relation of “awareness” or “acquaintance” can add that the subject is, in a very crude way, “aware of” this simple experience, though without the cognitive capacity to reflect on or reason about it.

(p.79) Note that this rudimentary consciousness need include no capacity for shifting attention, for focusing first on one aspect and then on another. Microsubjects need not be able to conceptualize the different aspects or actively distinguish them from each other. Given this, I do not think that the phenomenology sketched out here attributes implausibly complex experience to

an electron: there is really no more than a single “bit” of information, this-as-opposed-to-that, motivating one of two responses (“toward” and “away from”), both having a mathematical degree of intensity. And yet it also seems to contain the germs of cognition, volition, affect, and qualitative sensory perception, making it a suitable explanatory base for human consciousness even if those features are all considered primitive.

This is not the only way that it might feel to be a fundamental particle. For instance, maybe the basic motivation is not pleasure or displeasure but a sort of blind love and desire for union with the world that is inarticulately perceived, or a sort of “tension” that is more basic than either pleasure or displeasure (cf. Freud 1961; Schopenhauer 1969; Spinoza 1994). We are not in a good position to decide among these alternatives: positively determining the experiences of the fundamental physical entities is probably beyond current human ability, perhaps requiring a near-completed physics, a near-completed introspective phenomenology, and a near-completed neuroscience (maybe augmented with superhuman powers of reasoning, introspection, and imagination). My goal here is not to decide what microexperience is really like, but to show that there is room for it to be both simple enough to be compatible with our present physics and rich enough that, as long as combination is possible, it provides a sufficient explanatory basis for human experience.

3.1.2. Two Principles for Combining Experiences

Supposing that microsubjects have some sort of microexperience, how do they combine? The first and most basic principle of panpsychist combinationism is that experiential properties are inherited: for any collection of microsubjects, the aggregate they compose is a subject of all their experiences.

> **Experience Inheritance (EI):** Whenever a part of aggregate *x* undergoes an experience (instantiates an experiential property), *x* undergoes that same experience.

So if one particle is undergoing experience A, and another is undergoing experience B, the aggregate of the two is undergoing experience A and also undergoing experience B. This does not involve any duplication of experiences, but rather a **(p.80)** sharing of experiences between part and whole. Thus experiential properties are division-invariant: however we carve up reality, we find the same experiences. In this regard experiential properties work just the same as properties like spatial location, causal power, mass, and charge: wholes have them just in virtue of their parts having them.

The second thesis of panpsychist combinationism concerns the unification of experiences:

> **Micro-Unity Hypothesis (MUH):** The inner nature of one, some, or all of the fundamental physical relations is phenomenal unity; when two

microsubjects are related in the relevant way, their experiences become unified, establishing a composite experience that subsumes them.

The phrase “fundamental physical relations” is meant to leave open what these are: the propagation of ripples in fields of force, the exchange of energy-carrying particles, mere spatial distance, or something else. Thus MUH is only schematic: it does not say how exactly phenomenal unity is distributed, only that its distribution matches that of one or more of the basic physical relations. This is already enough, however, to imply that phenomenal unity runs much more widely than we usually think. This is because all the fundamental physical relations we know of, or seem likely to discover, seem to be very extensive in their scope. For instance, gravitational attraction holds between any two particles with mass, whether they are in the same human brain, in different human brains, or one in a brain and one in the corona of the sun.[3] So MUH makes phenomenal unity nearly pervasive, an implication I explore in section 3.3.

While EI is a general metaphysical principle, to be defended *a priori*, MUH is offered as an empirical postulate: it aims to describe how the fundamental laws of our universe in fact work, not how they necessarily must work. MUH says that when two microsubjects do not stand in whichever of the fundamental physical relations is important, then even if they compose a composite subject, they compose one with disunified consciousness. Adopting the above suggestion that microsubjects are conscious of “this-against-the-background-of-that,” a composite **(p.81)** formed out of causally unrelated microsubjects would experience two “thises” against two backgrounds, each with its respective quality and motivation. But despite being conscious of A and being conscious of B, it would not yet be conscious of “A and B”: A and B do not appear as contrasting with or connected with each other in any way. When they enter into the relevant relation, however, whatever that involves, their experiences become connected, so that they form a single experience which subsumes each and is undergone not by either microsubject but by the composite subject they form.

That concludes my sketch of panpsychist combinationism. How does it address the five internal problems considered in chapter 2? First, EI conflicts with strong independence, since the (*a priori*) truth of EI implies that when we consider a whole made of conscious parts, we cannot consistently conceive of it as lacking consciousness of its own—because it is just them, considered together. Since it is just them, it has the consciousness they have, just like it has the mass they have, the volume they have, and so on. Moreover, EI contradicts strong privacy but supports weak privacy: it implies that particular experiences can be shared between wholes and parts, but not that they can be shared by discrete subjects. Next, MUH answers the unity argument by treating phenomenal unity as one of the fundamental relations knitting together our universe, holding not just within a single subject’s experience but between the experiences of distinct subjects.

But MUH does not imply—and panpsychist combinationism denies—the key premise of the boundary argument, that unity between two subjects dissolves their distinctness; rather, each undergoes an experience unified with another experience they do not undergo. And the incompatible contexts argument is dodged by not relativizing unity to subjects. Of course, both experience inheritance and the micro-unity hypothesis demand further explication and defense, which I seek to provide in the next three sections.

3.2. Why Are Experiential Properties Inherited?

The most basic objection to combinationism is that consciousness in a thing's parts tells us nothing about consciousness in that thing itself—that part and whole, being different subjects, are too "metaphysically insulated" for any property of the one to be a mere combination of properties of the other. In chapter 2 I called this principle "strong independence," and according to panpsychist combinationism, strong independence is false because the following principle is true:

(p.82)

> **Experience Inheritance (EI):** Whenever a part of aggregate *x* undergoes an experience (instantiates an experiential property), *x* undergoes that same experience.[4]

But why should EI be true? Not all properties are inherited, after all; a heap of beans is not itself a bean, nor even bean-shaped. My primary answer (section 3.4 provides a supplementary answer) is that experiential properties are primitive properties and that primitive properties are inherited.[5]

Thus I appeal not to anything specific about experience but to the general nature of the part-whole relation. Previous writers on the subject-summing problem (Goff 2009a, 2009b, 2017c; Coleman 2014, 33–34, 2017, 256–258) have tended to proceed by scrutinizing consciousness and subjecthood, looking for a distinctively experiential way to "hook" wholes and parts together. I instead claim that they are automatically hooked together, just because of what aggregation *is*. In a slogan, aggregates inherit the primitive properties of their parts because they are simply those parts considered as one thing rather than as many. More precisely, I will defend a principle of the "substantive indiscernibility of parts and aggregate," which entails EI, and which I argue is the best explanation of why many physical properties are inherited.

I should first clarify that when I say that aggregates undergo the same experiences as their parts, I mean "undergo" in the most minimal way. Recall the simplicity of microsubjects and the correspondingly rudimentary way in which they undergo experiences. When a human being experiences something, that generally lets them reflect on it, report it, remember it, focus attention on it, form a concept to name and categorize it, and so on. It is hard to imagine

undergoing an experience but not being able to do any of these things with it—we might even be reluctant to call that “consciousness.” Yet microsubjects cannot do any of these things, so the sense in which they have experiences lacks all of these usual connotations, stripped down to nothing but the “raw feeling” itself. I claim that wholes inherit experiences in this same rudimentary way: a whole undergoes the **(p.83)** same “raw feeling” as its parts but may not be able to attend to, reflect on, or otherwise actively think about its experiences. This kind of very simple and unimpressive consciousness is division-invariant, equally present however we divide up the stuff that instantiates it, even though more cognitively complex sorts of consciousness may not be.

Moreover, the “wholes” which inherit experiences are here thought of simply as aggregates, not as structure-specific wholes. In chapter 5 I will discuss what it takes for a structure-specific whole to inherit experiences; that discussion will turn out to be closely connected to the question of what it takes for something to “undergo” an experience in the richer sense. In this chapter I focus on getting a basic kind of experience for a basic kind of whole, hoping that this more basic task will provide the foundation upon which the more sophisticated task may be accomplished.

3.2.1. The Substantive Indiscernibility of Parts and Whole

Why do aggregates inherit experiential properties from their parts? Begin by considering the “indiscernibility of identicals,” the principle (sometimes known as “Leibniz’s Law”) by which, if two things are identical, every property of one is a property of the other. This principle lets us infer, from the premises that Bob is Robert, and that Bob is eighteen, that Robert is eighteen. This principle is explained just by what identity is: if two things are one and the same, it is contradictory to describe that thing one way under one label, and a conflicting way under another label.

The relation between an aggregate and its many parts is similar to the identity relation in certain salient ways. David Lewis (1991) identifies several similarities, such as the fact that an aggregate and its parts can share the very same locations, and the fact that if you “describe the character of the parts [and] describe their interrelation . . . you have ipso facto described the [aggregate]” (81), which he takes to motivate “the thesis of Composition-as-Identity” (82; cf. Baxter 1988; Cotnoir 2013; Baxter and Cotnoir 2014). This thesis does not quite say that composition and identity are the very same relation, but merely that they are “strikingly analogous” (Lewis 1991, 84). Part of the reason why they are not the very same relation is that the “indiscernibility of parts and aggregate” is clearly false: aggregates often have properties that their parts lack, and vice versa. For example, an aggregate that weighs 10 kilograms can have parts that weigh 1 kilogram; a square aggregate can have triangular parts; an aggregate that contains arsenic can have parts that are arsenic-free. But rather than abandoning the analogy between identity and

composition, we should qualify the indiscernibility of parts and aggregate. **(p. 84)** As Lewis (1991, 87) says, “It does matter how you slice it—not to the character of what’s described . . . but to the form of the description.” To try to capture the way that the “character of what’s described” remains the same, however we slice it, I propose:

> **Substantive Indiscernibility of Parts and Aggregate (SI)**: For every property had by some part of an aggregate, that aggregate has a corresponding property, and for every property had by an aggregate, one or more of its parts have (individually or collectively) a corresponding property.

The idea is that the properties of the aggregate and the parts do not differ “substantively,” in that facts about the whole and the parts involve the same portion of reality. Their differences are just differences in the appropriate way to describe that reality: the properties of one may not be strictly the same properties as those of the other, but they are still “corresponding.” Going along with this, it is natural to think of the fact of an aggregate having the one property as *grounded in* the fact of one or more parts having corresponding properties, and vice versa; that is, there seems to be symmetrical grounding here, just as grounding by identity is symmetrical.

But what are “corresponding properties”? I do not have a precise definition, unfortunately. The idea is that the aggregate having the one property and the part(s) having the other are not really two distinct facts, but the same fact described in two ways. But how do we tell which property corresponds to which? For some properties, analysis of what that property is—what it takes for an object to instantiate it—will by itself reveal the other properties that would correspond to it under various circumstances. In such cases the corresponding property to X is one that either grounds X by analysis or is grounded in X by analysis.[6] But for primitive properties, which do not admit of any analysis, we cannot identify corresponding properties by this method. For such properties, I claim, their only corresponding properties are themselves, and so when they belong to the parts the aggregate must also have them. To put it another way, SI says that for each property of the parts, the whole either has the same property, one grounded in it by analysis, or one that grounds it by analysis. All properties of aggregates will thus be either division-invariant or intelligibly division-relative.

Because experience is irreducible to anything nonexperiential (or so say panpsychists), at least some experiential properties are primitive. Of course **(p. 85)** there might be analyses of some experiences in terms of others (e.g., to feel Schadenfreude is analyzed as feeling pleased as a result of consciously thinking that someone else is unhappy), but the process of analysis must come to an end somewhere. Whichever experiential properties are primitive will be inherited if substantive indiscernibility is true, and if they are inherited, the other

experiential properties that are constructed from them will also be inherited (e.g., if I inherit the pleasure and the thought of another's sadness, properly connected, then I inherit Schadenfreude). Thus if SI is true, EI follows: it is impossible for an aggregate to be made of conscious matter and yet not be conscious, just as much as it is impossible for Robert to be eighteen and Bob twenty-three, if Robert is Bob.

I will try to clarify the idea of "corresponding" properties through examples, and thereby also make my main argument in support of SI. To some extent SI is intuitive by itself: if a change is made to part of something, surely it follows that the whole thing will have been changed in some fashion. But a stronger justification for SI is indirect: it best explains the behavior of physical properties.

3.2.2. Examples of Physical Properties Obeying Substantive Indiscernibility

Begin with spatial location: the following principle is affirmed as obviously true by Van Inwagen (1990, 44), Lewis (1991, 85), Sider (2007b, 52 and throughout), Bennett (2015, 266), and McQueen (2015):

> **Location Inheritance:** An aggregate is located at a given point or region of space whenever one or more of its parts is, simply in virtue of the part being located there.

Of course, there is an obvious sense in which I am located in a different (larger) area than, say, my foot. But there are two senses of "located" (cf. Sider 2007b, 52n4): it can mean "wholly located," and this is not inherited, but this sense can be defined in terms of the other, more noncommittal sense, which is inherited. To be "wholly located" in some region is to be "located," in this noncommittal sense, at all and only the points in that region.[7]

Another plausible example of an inherited property is causal responsibility: what an aggregate does is just all the things which its parts do.

(p.86)

> **Power Inheritance:** An aggregate exercises a given causal power whenever any of its parts do, simply in virtue of the part exercising that power.

To use an example from Merricks (2001, 111), when a baseball shatters a window, each of its atoms causes something (e.g., a slight increase in the energy of one part of the window), and what the baseball causes is just all of these many small effects, which add up to shattering the window.[8]

A good example of correspondence without inheritance is shape. Say a table has a certain shape—roughly, a flattened cuboid with four elongated cuboids attached at its corners. None of the particles that compose it has this shape, yet they clearly display properties that correspond to it: it could not be *shaped* that way unless the parts were *arranged* a certain way. In some discussions of

composition, philosophers have used the phrase “arranged tablewise,” meaning “arranged in one of the ways that corresponds to the kind of shape a table can have” (Van Inwagen 1990, 109ff.; Sider 2007a, 1; Thomasson 2009, 256). Nobody tries to spell out what exactly it means for things to be arranged tablewise, but they do not need to: our grasp of “table-shaped-ness” (a property of the whole) is quite sufficient for us to grasp “arranged-tablewise-ness” (a property of the parts).[9] The very casualness with which terms like “tablewise” can be introduced shows how readily we can translate between the properties of the parts and the properties of the whole. And the difference between shape and location is simply that shape properties can be analyzed, and their analyses will identify distinct ways of describing the same fact—either as many things being arranged or as one thing being shaped. Understanding what it takes for a single thing to be, say, cubic, reveals also what it takes for a set of things to be arranged cubewise.

Of course, being arranged tablewise is not the kind of property that science talks much about. So consider “having a dipole moment”: a water molecule has a dipole moment; i.e., it is relatively negatively charged on one side, and relatively positively charged on the other. This corresponds directly to certain charge-properties and relative-position-properties of its parts. It would be impossible for the atoms to have those properties, and the molecule not have a dipole moment, **(p.87)** or vice versa, and this is something we can see just from analyzing what having a dipole is.[10] Similar things hold for properties like solidity (Jackson 1998, 3–4) and liquidity (Horgan 1993, 379).

A case that has received some debate is the additivity of certain basic physical properties, such as mass and charge. If something is composed of five discrete parts, each weighing 1 kilogram, then it weighs 5 kilograms: its total-mass property corresponds to their component-mass properties. This seems like a paradigmatic case of an aggregate having a property that is a mere combination of the properties of its parts. But some philosophers have denied that the correlation between the values of component masses and total masses is an *a priori* matter: it is rather a contingent fact that we have discovered empirically (e.g., McLaughlin 1997, 38–39; cf. Broad 1925, 63). If it is a contingent fact that mass is additive, it is hard to see it as conforming to any such abstract metaphysical requirement as SI. But in fact I believe that although mass additivity is not *a priori*, something close enough to support SI is.

McQueen (2013, 2015) points out that matters are complicated here because in relativistic physics, mass additivity is actually *false*: an aggregate’s mass depends not just on the masses of its parts but on their velocities. Insofar as we discovered empirically that relativity theory, and not Newtonian mechanics, is correct, we have thereby discovered that mass is not additive, but rather is a function of component masses and velocities. However, physicists do not generally take the composition of masses, following whatever formula, to be itself one of the fundamental laws; it is not something to be included alongside

Newton’s or Einstein’s laws of motion and gravitation. Rather, it is derivable from those laws. The lesson is that while mass additivity is not *a priori*, what is true *a priori* is that if Newtonian mechanics were the whole truth about the microscopic behavior of matter, then mass would be additive. Likewise, it is true *a priori* that if Einsteinian relativity were the whole truth about the microscopic behavior of matter, then mass would be a certain function of component masses and velocities.

But how exactly does mass additivity, or an alternative formula, follow *a priori* from the laws of microphysical behavior? McQueen (2013, 2015) offers a detailed philosophical examination of several derivations presented in physics textbooks (e.g., Kibble and Berkshire 2004, 12; Lindsay 1961, 19–21); in the space available I can only briefly summarize the general pattern of these derivations. They appeal **(p.88)** to the laws connecting force and acceleration, with “mass” defined as the objective feature of things which accounts for their resistance to acceleration given a force on them. An aggregate’s acceleration is then defined, in terms of the acceleration of its parts, and the force acting on it is likewise defined, as equal and opposite to the force it exerts, which in turn is defined by the forces its parts exert. With figures for the aggregate’s acceleration under a given force, we can compute its mass. The key physical premises are the relationship between mass, force, and acceleration, and the law that to each force exerted there is an equal and opposite force. The key metaphysical premises are location inheritance and power inheritance, which are implicitly required to specify, respectively, the acceleration of (McQueen 2013, 53–54) and force on (54–55) the aggregate. This fits neatly into my account of corresponding properties, because the derivation rests on an analysis of our most revealing concept of mass—namely, that it is the objective feature of things that mediates between force and acceleration, according to a certain set of empirically discovered equations. Reflection on that analysis is what reveals the exact mathematical structure of correspondence between mass in the aggregate and mass in the parts.

Even if many physical properties obey substantive indiscernibility, are there others that do not? Can we find either properties of wholes that correspond to nothing about their parts, or properties of parts that correspond to nothing about the wholes they form? Both sorts of example have been discussed by metaphysicians, but I believe neither really threatens SI.

First, what about wholes which *lack* properties corresponding to some of those of their parts? There do seem to be cases that could readily be analyzed this way, like the whirlpool or biological cell mentioned in chapter 1, section 1.3. But these are clearly examples of what in chapter 1 I called structure-specific wholes, and so do not threaten SI as a principle applying to aggregates.

For properties of wholes corresponding to nothing about the parts, we might consider emergent properties, appearing only in certain wholes and not predictable in advance even from the most complete knowledge of the individual parts. There are currently no generally accepted examples of such emergent properties, but for a period in the nineteenth and early twentieth century many philosophers thought, apparently quite reasonably, that both chemical compounds and biological organisms had causal powers that could not be predicted from the most complete knowledge of their parts (e.g., Mill 1882; Broad 1925). While the progress of science seems to have refuted these ideas, I follow McLaughlin (1992) in thinking that the emergentist's views were not *a priori* false: nature could have turned out to work the way they thought it did.

(p.89) Does SI conflict with the apparently reasonable, though seemingly false, empirical hypothesis of emergence? Not necessarily. First, SI is about aggregates, and it might be that emergence is possible only for structure-specific wholes, or even true units. Second, although emergence is usually discussed in part-whole terms, the scientific implications of traditional scientific emergentism can be equally well cashed out in terms of general and specific conditions. On emergentist scientific theories, matter has some powers that manifest only under very specific circumstances, while on nonemergentist theories, matter displays all its powers under a wide range of circumstances. This contrast can be expressed entirely in terms of properties belonging to the smallest particles of matter. For example, Broad (1925, 65) suggests that the properties of silver chloride are "emergent" relative to those of silver and of chlorine. This amounts to saying that atoms of silver and chlorine (the parts) have certain powers that manifest only in the very specific case where they are bonded together—hence those powers "cannot . . . be deduced from the most complete knowledge of [silver and chlorine] *in isolation or in other wholes*" (63, emphasis added).[11] This does not threaten SI: when the special powers are manifest, they are manifested both by whole and by part, and when they are merely latent, we can ascribe them both to the parts and to the impotent wholes they then form. If silver chloride has the "emergent" powers, we should both ascribe these powers, when they are manifested, to the silver atoms and the chlorine atoms, and also ascribe these powers, when they are latent, to any aggregate containing silver or chlorine not bonded to each other.

3.2.3. Why Do Physical Properties Obey Substantive Indiscernibility?

So physical properties seem to obey substantive indiscernibility. Why? The explanation I prefer is that substantive indiscernibility is true across the board, about physical and mental properties alike. Sider (2007b) draws a similar conclusion: the best explanation for things like the inheritance of location is something about the composition relation itself, namely that it is akin to the identity relation. This seems to me the simplest and most satisfying account, and it implies that SI holds for experiential properties, thereby supporting EI.

One alternative explanation would be that SI holds for physical properties because of something about matter specifically: not all things show this intimacy of whole and part, but only the physical stuff that happens to make up our universe. This explanation still supports EI, given the sort of universe we appear to inhabit: **(p.90)** a universe composed entirely of matter, some or all of which is conscious. For in such a universe, all the experiential properties that occur will belong to matter, and so obey SI.[12]

An explanation that was genuinely problematic for combinationism would be if SI held for various physical properties for disparate reasons, with no general unifying theme or principle behind it. If location was inherited for one reason, causal effects for another, charge and mass for some other unrelated reason, and so on, then there would be no basis for drawing any conclusion about experiential properties. However, it is implausible on the face of it that so many different properties should show such similar behavior for completely different reasons: it seems more reasonable to prefer a more systematic explanation for their behavior given above.

It would also be a problem for combinationism if physical properties obeyed SI specifically because they are physical *properties*, rather than because they are properties of a certain stuff. In particular, Russellian panpsychists hold that physical properties are, ultimately, "structural" properties, abstract descriptions of the ways that some nonstructural properties are related, while experiential properties are the underlying nonstructural properties that do that relating. A critic might claim (and from conversation I believe this is Goff's position) that structural properties are inherited, because the roles which define them can be played equally well by a thing's own properties and by the properties of its parts. In particular, a structural property's defining roles might be played equally well by the property of *experiencing X* and by the property of *having a part that experiences X*. Then a whole could have the structural properties implemented by the experiences of its parts, without itself experiencing anything. I do not have a decisive objection to this account, though it seems to me still to leave something unexplained: Why is the part-whole relation such that wholes can have structural properties in virtue of the nonstructural properties of their parts? Other relations (being larger than, being fond of, being genetically related to, being earlier than) do not work like this: I cannot instantiate structural properties in virtue of the properties of my relatives, or things larger than me, or things I am fond of. The most satisfying explanation for this special feature of the part-whole relation is that wholes in some sense *are* their parts, and not a genuinely distinct thing related to them. But **(p.91)** that is not a knock-down argument for accepting SI, and critics might think that the apparent conceivability of a nonconscious whole made of conscious parts (like the microexperiential zombie) counts against it; in section 3.5 I will consider in more

detail what panpsychist combinationism can say about the microexperiential zombie and the threat it seems to pose to EI.

3.3. How Do Microexperiences Become Phenomenally Unified?

Whereas the previous section's case for experience inheritance was pure metaphysics (parts and wholes necessarily share properties), the relations among experiences depend on the contingent laws of our universe. In chapter 2 I distinguished "primitivism" from "reductionism" about phenomenal unity: Is phenomenal unity just a matter of being sufficiently richly connected in various other ways (causally, representationally, functionally, etc.), or is it a distinct and more basic relation? If reductionism is true, then the challenge of explaining phenomenal unity is nothing beyond the task of explaining those other relations, which are better tackled in chapter 5 as part of functionalist combinationism, not here. But let us suppose that primitivism is true. In that case I propose that phenomenal unity is the inner nature of one or more of the fundamental physical relations:

> **Micro-Unity Hypothesis (MUH):** The inner nature of one, some, or all of the fundamental physical relations is phenomenal unity; when two microsubjects are related in the relevant way, their experiences become unified, establishing a composite experience that subsumes them.

This postulate could be attacked in a number of ways, so in this section I will defend it against the following three objections:

1. Why think that phenomenal unity corresponds to one of the fundamental physical relations (and not to some relation that holds only within animal brains)?
2. Doesn't MUH, by extending phenomenal unity beyond the bounds of human brains, dissolve individual subjects into an all-encompassing cosmic consciousness?
3. Doesn't MUH, by extending phenomenal unity beyond the bounds of human brains, contradict the starting assumptions that gave us a grip on the very idea of phenomenal unity?

(p.92) #### 3.3.1. Why Unity Is Probably Everywhere

The most obvious objection to MUH is simply a question: Why on earth think that phenomenal unity is coextensive with one of the fundamental physical interactions?

Since we ourselves enjoy unified experiences, we know that many microexperiences do in fact stand in the relevant unifying relation: the laws of nature dictate that microexperiences, at least sometimes, form composite experiences. MUH goes beyond this, to claim that the laws of nature associate phenomenal unity with one of the fundamental physical relations (or with several of them). Perhaps, for instance, the situation we describe mathematically in

terms of two particles imposing electromagnetic forces on one another, or the situation of their imposing any forces on each other, or even the situation of their being at some spatial distance, is simply the situation of two microsubjects' experiences becoming phenomenally unified. Of course we cannot at present have much idea of *which* relation; like the question of what experiences microsubjects have, this question could likely be answered only at some future time when we have advanced considerably in physics, in phenomenology, and in neuroscience.

But why shouldn't phenomenal unity be associated with a nonfundamental physical relation, or even with no physical relation at all? After all, that would allow us to vindicate the natural assumption that phenomenal unity holds within, but not between, human brains, by defining some nonfundamental relation such that it holds only between particles in a human brain. We could then say that only when particles are related in *that* way—only when they interact in the right kinds of patterns, with the right degree of strength, in the right kinds of materials—do their experiences come to mutually co-present. Of course we would need to do a lot more neuroscience to know exactly how to define this relation, but there are problems with this approach that can be recognized in advance of such work. I believe that any theory on which unity is associated with a nonfundamental physical relation will lose out, in a comparison of theoretical elegance and simplicity, to a theory which associated it with a fundamental physical relation.

To see this, suppose we have identified some relation R which holds between particles in a human brain (and perhaps some other sorts of brain), but not between particles in different brains, and we hypothesize that it is *this* relation which correlates with phenomenal unity. Now consider a hypothetical sequence of steps, as gradual as we like, between a set of particles in a brain, related by R, and a set of particles spread across two brains, not related by R, such that at each point in the sequence the particles are interacting *slightly* more strongly than before in whatever ways are relevant to R. We want to say that at one end of this sequence there **(p.93)** is no phenomenal unity, while at the other there is, but from the perspective of fundamental physics there is no point where any fundamental break occurs: it is all just particles and forces.[13]

Things might not have turned out this way—we might have found that certain fundamental forces kick into action only within the boundaries of human heads, so that no such gradual sequence could be constructed. Indeed, we might have found that consciousness depends on immaterial souls, interacting with but separate from the material brain. In that case there would be a clear, sharp, nonarbitrary boundary between the kind of interactions that take place between two of my experiences (within a soul, with no physical steps at all), and between mine and yours (a soul-state causes a physical state, which causes another physical state, which causes a state in a different soul). In such a world, the

intuitive distinction noted in chapter 2 between the “direct” relations that my experiences seem to bear to each other and the “indirect” relations they seem to bear to your experiences would be real. But scientific investigation has not revealed a world like that; it has revealed a world where the interactions within one mind and the interactions between minds differ only in degree—in strength, sensitivity, reliability, speed, and so on. In such a world, for any proposed R, we can construct a sequence as above, where each one of the many tiny steps is, from a physical perspective, as trivial and unimportant as any other, and none of them seems like it could make the difference between unity and disunity.

Do we really want to say that phenomenal unity is something that “kicks in” only when a certain *precise* degree of physical relatedness is present? Can we believe that, to put it crudely, there is no unity when two distant particles exert a force of 24-trillionths of a Newton on each other, nor when they get a little closer and exert a force of 25-trillionths of a Newton, or 26-, but as they grow nearer and the force between them increases—perhaps when they reach 1,301-trillionths of a Newton, perhaps 2,753-, who knows—there is suddenly unity?

It is deeply unappealing to think that a change of 1-trillionth of a Newton should make such a big difference. For one thing, it seems completely arbitrary that it should be 2,753 rather than 2,752 or 2,754. For another thing, it seems to make phenomenal unity basically epiphenomenal: there is virtually no difference in the causal behavior of two things that exert 2,753- rather than 2,752-trillions of a Newton on each other, so it becomes hard to understand how phenomenal unity could play any significant role in what causes what.

(p.94) Of course this is just one more version of the ancient “sorites paradox”: no tiny step can suddenly produce unity (just as no single-hair-removal can make someone bald), but a transition from disunity to unity must occur at some point (just as progressive removal of hair must eventually make someone bald). Analyses of the sorites paradox abound, but broadly speaking, most solutions involve the idea that the word being applied can be applied in a variety of fractionally different ways. For instance, the vagueness of “bald” lies in the fact that any of a range of maximum numbers of hairs is an equally good candidate for the meaning of “bald.”[14] When someone has little enough hair to be bald by one standard, but too much by another standard, it is semantically indeterminate whether they are bald.

A solution like this works well for concepts like “bald,” whose meaning can be specified in more basic terms (number of hairs). The different acceptable ranges of application can be understood as different precise descriptions in these more basic terms. But this requires an analysis of the vague concept in terms of degrees on an underlying spectrum. Being indeterminately bald just means having a number of hairs we are unsure how to classify, but what is it to be indeterminately unified? After all, when experiences are unified they are parts of

a composite experience, so if it is semantically indeterminate whether two experiences are unified, it will be semantically indeterminate whether there is a subsuming experience or not. If consciousness were analyzable into some complex of physical facts, and phenomenal unity into a complex of physical relations, then we could make sense of indeterminacy as a matter of most but not quite all of those complex physical conditions being present. But if consciousness is a fundamental property, as panpsychist combinationism supposes, then there is no in-between possibility which our label “conscious experience” indeterminately applies to. Either a state is a conscious experience or it is not; either there is something it is like to be in that state (even if something incredibly simple, dim, and faint), or there is nothing it is like. But then the boundaries of unity could not be vague.

So I find myself driven to a surprising conclusion: since the only nonarbitrary physical boundaries lie between not interacting at all and interacting a bit, even a tiny bit, there is probably phenomenal unity whenever two experiences are physically related at all, regardless of how strongly.

(p.95) 3.3.2. How to Address the Boundary Problem

MUH implies that the phenomenal fields in the world are of vast size, limited only by the reach of the fundamental physical relations. Each of these fields will extend far beyond a human being, on the scale of light-seconds or light-years. And perhaps (especially if phenomenal unity is transitive, so that wherever the fields overlap they form a single larger field) subsuming all of these fields is the single huge field of the universe. In either case, we are dealing with fields on a scale far larger than the human individual.

Indeed, the transitivity of phenomenal unity may provide an alternative route to the same worrying conclusion. Dainton (2011, 255), for instance, argues that given transitivity, and given that “every particle composing the planet Earth is linked (directly or indirectly) to every other by a chain of physical connections (or interactions)—a by no means implausible assumption—then the entire planet will consist of a single fully unified consciousness.” Here the point is that even if the interactions which establish unity are quite local and small-scale, the existence of long connecting chains will still generate large, suprahuman, phenomenal fields. Consequently, avoiding this kind of result is hard for the combinationist panpsychist: they must *both* deny the transitivity of phenomenal unity *and* associate it with a sufficiently specific and rare sort of interaction. Rather than contorting the view to make it fit commonsense intuition, I prefer to embrace the counterintuitive consequences of allowing fundamental experiential relations to behave like fundamental physical relations.

But isn’t this result disastrous? The idea that all the experiences in the universe are phenomenally unified, subsumed by the vast phenomenal field of the whole cosmos, is precisely what was worrying proponents of the boundary argument.

Even if they sympathize with the motivations for MUH, they might feel that its implications are so patently false as to make it a nonstarter. After all, MUH seems to imply that all of my experiences are unified with all of yours, and those of everyone else on earth, as though we all formed some kind of vast superconsciousness. But (these critics might say) of course we don't. Rosenberg (2004, 88) suggests that "this view . . . banishes middle-level individuals from existence."

But does the presence of this larger unified mind, containing us as parts, really conflict with our existing and having the kind of consciousness that we do? No. Wholes do not in general "banish their parts from existence," even if the same relations that hold within the whole also hold between it and other things. A table leg is not banished from existence by the table, even if the same molecular bonding relations that connect its parts to one another also connect them to the rest of the table.

(p.96) Of course, if we thought that phenomenal unity between two experiences confers both on the subject of either, then we would have a problem, for then my experiences being unified with yours would imply that I was having your experiences, and you were having mine, and everyone was having all the experiences; there would then be no subjects who, like us, undergo only some and not all of the world's experiences. But we should not think that unity has this effect—the idea that it does is what in chapter 2 I called "boundedness":

> **Premise D3 (Boundedness):** For any experience e_1 belonging to a subject s, if another experience e_2 is unified with e_1, then e_2 must also be had by s.

Panpsychist combinationism denies boundedness for phenomenal unity and claims that my experiences are unified with yours even though I am not undergoing yours. Whenever two human brains exert (say) a minuscule gravitational attraction on each other, this suffices to establish a composite experience that subsumes the experiences of each but leaves each still only undergoing their own experiences, not those of the other.

Why might someone think boundedness is true? One reason might be that they are impressed by the thesis that I called interdependence—the way that when two experiences become unified it changes the character of each:

> **Premise E3 (Interdependence):** The phenomenal character of an experience often depends partly on its phenomenal context, i.e., on the set of other experiences it is unified with.

If interdependence is true, then we might reason as follows: if my experiences were unified with yours, they would to some extent reflect the character of yours. When you started to feel sad or happy, that would impact the way I felt—we would have a kind of telepathy. Avicenna (1952, 5.3.7–8) seems to express

this thought when he rejects the idea of a single continuous soul in all bodies because then “it would have knowledge or ignorance in all [the bodies] and it would not be hidden for Zayd what is in the soul of Amr” (cf. James 1909, 201–203; Goff 2012).

That would be a reason to think that MUH is incompatible with the actual degree of mutual ignorance that we humans experience. The further step to thinking that boundedness is true across the board might go as follows: if each of two unified experiences makes its subject aware of the other, then the subject is introspectively aware of both, and this seems tantamount to having them both. After all, the subject is being made aware, consciously, of what that experience is like. Isn’t that just what having an experience is? I am not sure exactly how plausible such **(p.97)** an inference is; I am to some extent trying to reconstruct what is motivating those who affirm a principle I deny. (In Roelofs [2016] I examine in more detail what reasons might be given for affirming boundedness for phenomenal unity.) Here I will simply say why MUH, with its implication of unity between my experiences and yours, is neither necessarily false (due to boundedness) nor empirically false (due to our obvious ignorance of each other’s experiences).

Let us grant that unified experiences affect each other’s character; this does not stop us from saying that they may do so more or less strongly. Maybe when I see someone’s dancing at the same time as I hear the music, the two experiences interact deeply, allowing me to see their movements as synchronized with and expressing the music. But I am also having plenty of other experiences at the same time (the feel of my clothes on my arms, an idle thought, a vague nagging feeling of having forgotten something), with far less impact on either how I see the other person or how I hear the music. Indeed, some unified experiences (e.g., my current feeling of my socks and the visual image of a lampshade beside me) are phenomenally unified but seem to make virtually no interesting difference to one another. Perhaps a strong supporter of interdependence will insist that there is still some sort of mutual influence (see, e.g., Dainton 2010; Chudnoff 2013), but it must be so mild as to be easy to miss. Plausibly, the strength of this mutual influence is linked to the strength of the causal interactions between the experiences; if this is the case, then the mutual influence established by phenomenal unity between my experiences and yours will be fainter even than that between two mutually irrelevant experiences of mine, like those of my socks and the lampshade.

The objector might press the point as follows: Doesn’t MUH still imply some sort of phenomenal influence of my experiences on yours, however faint? I reply: Sure! But such an influence will (1) be so faint that we will not be able to use it to learn anything specific about each other’s experiences, (2) will blur together with the influence of everything else in the universe, and (3) will have been present constantly throughout our lives, so that we have no idea what it would

be like to live without it. Given this, what evidence do we have that this influence is not actually being exerted? How do we know that in a world of fundamentally separate minds, like the world envisaged by Descartes, we would not have lived all our lives with a slightly different “background feel” to our experiences, though without ever being able to identify it because we never had anything to contrast it with?

The objector might push their objection a slightly different way: Surely MUH still implies the possibility of some kind of strange supernatural telepathy? After all, if we simply made it so that our brains exchanged more information, then the mutual phenomenal influence they have in virtue of being phenomenally unified would become stronger and more distinctive, allowing each of us to know, just **(p.98)** from the character of our own experiences, what the other was experiencing. I reply: Sure! That is called communication, and we do it regularly. When I talk to my friend, or gesture at them, or smile at them, I am manipulating the physical relations between us (whose inner nature, we are supposing, involves phenomenal unity) so that their experiences will give them knowledge of mine. MUH simply says that what I am doing to my friend is not fundamentally different from what my left hemisphere does to my right hemisphere when it transmits nerve signals down the corpus callosum: both involve physical transfers of information; both involve phenomenal unity. A more detailed examination of this sort of process will have to wait until chapters 5 and 6, but for now I will simply say that according to MUH, it is a mistake to speak, as the objection does, of “supernatural” telepathy: telepathy is the most natural thing in the world.

3.3.3. Why Consciousness Seems Bounded

A final objection would allege that MUH is stretching the very idea of phenomenal unity past the breaking point. Didn’t we introduce phenomenal unity as a relation that characteristically holds within, but not between, the consciousness of different humans? And so if we start to think that my experiences actually are phenomenally unified with yours after all, don’t we lose our grip on what phenomenal unity was meant to be?

It sometimes happens that we initially form a concept to refer to something based on its evident presence in some cases, and apparent absence in others, only to later find that it is actually present in both cases. For example, the concept of “gravity” was originally introduced for the downward tendency of stones, water, and other bodies that fall downward, while air and fire were ascribed a corresponding property of “levity,” a natural tendency upward. As it turns out, air and fire have gravity (weight) as much as earth and water, and the upward force on them is a result of gravity: denser substances around them jostle them out of the way. So gravity turned out to be present even in the cases it was initially contrasted with. This did not undermine the original meaning of the concept, because gravity was still *evident* in earth and water in a way that it

was not in air and fire, so that its evident presence in heavy substances could be used to fix reference to it.

Something similar applies to phenomenal unity, if MUH is true. Each human being enjoys overall unified experiences, and moreover this unity is evident to their introspective self-awareness; even if pairs of human beings enjoy overall unified experiences, neither the pair nor either individual has introspective awareness of that fact. This pattern of introspective impressions was our original datum, and we need not violate it by extending phenomenal unity more widely: we simply **(p.99)** replace one explanation of the lack of an introspective impression of unity in the pair (viz. that there is no unity) with a different explanation (viz. that there is unity, but no capacity to introspect it).

The evidentness of phenomenal unity in reflective beings here plays the same role as the evidentness of gravity/weight in objects heavier than air. Each of us can report a rough summary of all the many experiences we are having right now—I can "cast my inner eye about" and tell you that I am seeing certain things, feeling and hearing others, with my mind on a certain topic, and so on—and that I am experiencing all these things together. You can tell me similar things: each of us thus reports the content of a single complex experience which subsumes many others. But this is not enabled simply because those complex experiences exist and subsume their parts: it is enabled because of the human brain's capacity to do this kind of introspective survey, to synthesize an overall impression of a conscious field.[15] According to MUH, there is another, even more complex experience, which subsumes both of ours (at least as long as we are close enough for our brains to be exerting forces on one another). This experience has a subject—it is being undergone by a pair of people, the composite that you and I are parts of. But this subject is not integrated enough to survey its field of consciousness and report a summary. Equivalently, the composite experience which mine and yours form together is not in a position to produce an introspective summary; it is not poised for that kind of access. To that extent, it could be misleading to describe it as a conscious state (in a sense nobody is "conscious of it," i.e., nobody is "access-conscious" of it), even though it is phenomenally conscious in the sense that there is something it is like to undergo this state.

The situation is rather as though we are standing under two different streetlights, which reveal circular areas of an unbroken, continuous street surface. Introspection is the light, and phenomenal unity is the continuity of the surface that it reveals. It is compatible with what we are seeing that the continuity of the surface stops where light stops, so that we stand upon separate pedestals with no link between them. But it is also compatible with what we are seeing that the surfaces extend beyond where the light stops, and in fact are just one big surface. I think our default assumption, if we ever consider the question, is that we in the first situation—standing on "separate pedestals," unified minds

with no unity out beyond them. It's an easy, intuitive, idea, and it might have been true: we might have inhabited a world where every mind was a true unit, with a precise and **(p.100)** objective boundary between itself and the rest of the world. Plenty of philosophers have defended such a view, and it is arguably implicit in the very widespread view that for each human being, there is a distinct entity, a "soul," which can and hopefully will survive the dissolution of the body. I think the seeming bizarreness of MUH partly reflects that such a view is baked into how most of us think about the world, whether or not we would on reflection accept the metaphysics it implies.

But that world of sharply bounded minds does not seem to be the world we inhabit. Insofar as scientific investigation of the world has revealed no fundamental physical breaks in the fields of matter and energy whose perturbations make up human brains and bodies and everything else, scientific investigation gives us reason to prefer the rival hypothesis, that there is just "one big surface," one big phenomenal field. Our two lights illuminate different sections of the same surface; our two brains enable introspective awareness of different subfields of the same phenomenal field.

What enables the formation of a second-order thought about the totality of my current experiences is that each of these experiences broadcasts enough information to the rest of my mind. This broadcasting of information changes and enriches each experience, so that each reflects and carries information about the others. This interweaving of experiences into more complex wholes is examined in detail in chapter 5; the key thing to note here is that, unlike phenomenal unity itself, this sort of access-unity does come in degrees. One part of my brain can send out more or less information, or better or worse information, or information that is harder or easier to detect, to other parts. And between the within-brain information transfers enabled by synapses and nerve fibers, and the minuscule gravitational attraction that any two brains exert on each other, there is a smooth continuum of intermediate degrees of communication.

Consequently the form of access-unity involved in the capacity for introspective survey and summation, unlike phenomenal unity itself, can be semantically vague. Recall that it can be neither determinately true nor determinately false that a man is bald, because our concept "bald" does not specify precisely how much hair is "enough" hair not to be bald is. Likewise a set of experiences can be neither determinately unified nor determinately disunified, because being unified requires "enough" information to be shared between experiences and synthesized by various mental systems, and our intuitive conception of unity does not specify how much information is "enough." This lets us say, as we were inclined to, that experiences are unified within the confines of organized structures like the human brain, disunified between disparate objects like two

people lying on a beach, and in in-between cases, like perhaps split-brain patients, semantically indeterminate between unity and disunity.

(p.101) 3.4. The Subsumption Argument from Micro-Unity to Inheritance

In the previous two sections I defended experience inheritance as an *a priori* necessity and the micro-unity hypothesis as an *a posteriori* conjecture. There is, however, also an argument in support of experience inheritance (or something close to it) which takes MUH as a premise, and thus establishes only that EI is true in the actual world if MUH is. That argument can be summarized quite readily: If every set of parts is connected by the fundamental physical relations, then MUH implies that every set of parts has unified experiences; this means that their experiences are subsumed by a composite experience; but since experience implies a subject, this means there must be a composite subject to undergo the composite experience. Call this “the subsumption argument.”

3.4.1. The Subsumption Argument Laid Out

Formally the argument goes like this:

Premise 1: The deep nature of one of the fundamental physical relations is phenomenal unity (the micro-unity hypothesis).

Premise 2: All of the fundamental physical relations hold either universally or nearly universally.

Premise 3: When a set of experiences is phenomenally unified, there is a composite experience which subsumes them.

Premise 4: For any experience, there must be a subject.

Premise 5: To undergo a composite experience involves undergoing all the experiences it subsumes.

6 *(from 1 and 2)*: For all, or almost all, aggregates in our universe, the experiences belonging to its parts are all phenomenally unified.

7 *(from 3 and 6)*: For all, or almost all, aggregates in our universe, there is a composite experience that subsumes all the experiences of its parts.

8 *(from 4 and 7)*: For all, or almost all, aggregates in our universe, there is a composite subject that is undergoing a composite experience which subsumes all the experiences of its parts.

9 *(from 5 and 8)*: For all, or almost all, aggregates in our universe, there is a composite subject that is undergoing all the experiences of its parts.

Let me first say a bit about the premises of this argument, and then comment on the conclusion, which is not quite EI but comes very close.

The first premise is, of course, simply MUH, defended in the previous section. The second premise is an observation about how nature appears to be organized, **(p.102)** again discussed in the previous section. “Nearly universally” here means over large enough scales to connect all sets of things whose consciousness might be of interest to humans—not just “the parts of my brain” but also “my brain and your brain,” or “all the brains on earth,” or “all the toenails and cactuses on earth.” And if phenomenal unity is transitive, “nearly universally” will likely imply “universally,” because even things too distant to stand in some fundamental physical relation are probably connected indirectly by a chain of such relations.

Premises 3 and 5 are just articulating the plausible Bayne-Chalmers analysis of phenomenal unity: for two experiences to be unified is for there to be something it’s like to undergo both together. The key new claim is premise 4, that experiences require subjects. We might call this the “ownership principle”:

> **Ownership Principle:** For any experience, there must be a subject who experiences it.

I think the ownership principle is very plausible, as long as it is kept in mind that we are operating with a distinctly unimpressive sort of subjecthood: if panpsychism is true, then it must be possible for subjects to be as unintelligent as an electron is. If we took “being a subject” to imply intelligence, self-awareness, and so on, then the ownership principle might be much more doubtful, but readers who demand such things of anything called a “subject” will probably find functionalist combinationism, outlined in chapter 5, more congenial. According to panpsychist combinationism, a subject is simply any entity that it is like something to be, however simple, and experiences are instances of properties defined by what it is like to instantiate them. Given these definitions, the ownership principle is more or less tautologous.

Suppose we accept the premises: What about the conclusion, item 9? This stops short of saying that all aggregates undergo the experiences of their parts: instead it says that “all or almost all” aggregates are associated with a subject which undergoes the experiences of their parts. The first difference—all versus “all or almost all”—is one that, by definition, matters little for any human purpose. Even if mental combination occurs only in “almost all” aggregates, that is more than enough to account for the formation of human-scale subjects and to generate the same counterintuitive profusion of subjects that EI does.

What about the second difference? The subsumption argument establishes only that for each aggregate, there is a subject sharing all the experiences of its parts—not that this subject *is* the aggregate. But this difference too is not of very great significance, because in this chapter I am assuming the metaphysical conception of subjects. If this conception is correct, then an experience’s subject is **(p.103)** the entity which metaphysically underlies an experience; the

aggregate in question has a very good claim to being that entity, and so plausibly this difference from EI vanishes. On the psychological conception of subjects, we might instead say that the subject is the composite experience itself, or that it is a distinct being constituted by that experience but nothing over and above it (cf. the discussion of "thin" subjects explained by unity in Mendelovici 2018). I do not think the difference between such a view and panpsychist combinationism, as I have defended it so far, is a huge one.[16]

Interestingly, the subsumption argument could have been used as an argument *against* combinationism, if we accepted the principle of strong privacy. The argument shows that one thing combinationists might want (unifying relations among distinct subjects) implies experience sharing. The composite subject undergoes the subsuming experience, and thus also undergoes the experiences subsumed: each part's experience is thus shared by the composite. I take this to support my decision in chapter 2 to deny strong privacy (cf. Roelofs 2016, 6–11).

3.4.2. Is This Phenomenal Bonding?

It is worthwhile here to compare panpsychist combinationism, particularly as it comes out of MUH and the subsumption argument, with the "phenomenal bonding" account offered by Goff (2017a, b). On the latter account, we posit a "phenomenal bonding relation," about which we cannot know anything except that when subjects stand in this relation, they form a composite subject with unified consciousness. Goff (2017a, 183–186) moreover postulates that this unknown relation is most likely the deep nature of the relation of spatial distance, and that consequently it will bond microsubjects into composite subjects more or less indiscriminately.

(p.104) There are some important similarities between my theory and Goff's. We both posit that one of the fundamental physical relations has a phenomenal deep nature, and that this relation accounts for the combination of microsubjects into composite subjects with unified experiences. We both are inclined to think that such mental combination probably occurs universally or nearly universally, on the grounds that any restriction would have to be arbitrary. I have no objection to calling the relation I posit among microsubjects a "phenomenal bonding" relation. But there is an important difference between my theory and Goff's (even setting aside my *a priori* defense of EI): the bonding relation I posit is an introspectively familiar one, not a mysterious one we have no conception of. The relation I posit is simply phenomenal unity among experiences, the same basic relation we are each acquainted with among our own experiences.[17] In this way my theory avoids what Goff calls the "mysterianism" of his theory, the need to postulate something unknown so as to make combination work. While such a claim is (by its nature) hard to decisively refute, it is also somewhat unsatisfying, not only because it posits something whose nature is permanently opaque to us, but also because this posit makes it harder to see what explanatory role is played by the parts themselves being subjects (cf. Coleman 2017, 256–258).

After all, we could just as easily posit an unknown “relation X,” which forms subjects out of parts which are *not* themselves subjects. Without understanding the nature of either mystery relation, and how it connects with the subjecthood of the component subjects, there is little obvious advantage in postulating phenomenal bonding rather than X.

This difference between my theory and Goff’s flows directly from our different attitudes toward experience sharing. Because I think that wholes undergo the same experiences as their parts, I can say that the relation which bonds their parts together into a subject is simply a relation among those experiences. And we are already acquainted with relations among experiences, in particular with phenomenal unity. Goff, by contrast, denies experience sharing, holding that the experiences of the parts “constitute” but do not “characterize” the experiences of the whole. Consequently, the relation which bonds the parts together into a whole is not any relation which we wholes might be acquainted with among our parts. Indeed, Goff argues for precisely this reason that we will never be able to form any positive conception of the phenomenal aspect of this relation.

(p.105) Overall, I am not sure if it is embarrassing or reassuring to have two completely different justifications for the same principle (EI), one entirely *a priori* and one deriving it from a posited contingent fact about our universe. On the one hand it may seem a bit like having two completely independent accounts of how one came into legitimate legal possession of certain goods: each story casts suspicion on the other. But on the other hand, it does reinforce my feeling that even if I have taken a wrong turn somewhere, the basic ideas of panpsychist combinationism—unity and subjecthood pretty much everywhere, experiences shared between parts and wholes—are along the right lines, at least if constitutive Russellian panpsychism is true.

3.5. Why Do Microexperiential Zombies *Seem* Conceivable?

In section 3.2 I made a case for the *a priori* truth of experience inheritance, the claim that experiential properties are division-invariant, present in a given portion of reality regardless of whether we see it as one whole or many parts. This implies the impossibility—indeed the inconceivability—of a nonconscious whole made of conscious parts, like the “microexperiential zombie” envisaged by many critics of constitutive Russellian panpsychism. And yet such a scenario has seemed conceivable to many philosophers; indeed, most people probably think it is not only conceivable, and possible, but also *actual.* (The solar system has me as a part, and I am conscious, but it is not.) Indeed, it sometimes seems conceivable to me, when I think about it the right way. So even if we thought that the argument given in section 3.2 is sound, we still need some account of what mistake explains the apparent conceivability of nonconscious wholes with conscious parts, and why smart people make that mistake.

What has gone wrong? I think there are two distinct conceptual confusions at work, one concerning wholes and one concerning consciousness, both of which are hard to avoid. In short: it is easy to overinflate what wholes are, so as to give a spurious appearance of conceivability to scenarios where their properties diverge from those of their parts. It is also easy to conflate phenomenal consciousness with access consciousness, and thus to mistake a scenario where a whole with conscious parts lacks access consciousness for a scenario where it lacks phenomenal consciousness. When considering something like the macroexperiential zombie, these two confusions work together in a mutually reinforcing way. In this section I try to motivate this view, that the conceivability of nonconscious wholes with conscious parts is only apparent. But I will also, in the next section, try to show that much or all of panpsychist combinationism can survive even if this claim for *a priori* status is abandoned.

(p.106) 3.5.1. Confusion about Wholes, and the Apparent Conceivability of Compositional Nihilism

Let me first describe the confusion I think is possible regarding parts and wholes. The best way to start is by observing that some philosophers have thought it conceivable that physical combination could fail, because I think it is more intuitively obvious in this case that some sort of mistake must have been made. Back in section 3.2 I mentioned Merricks's discussion of what a baseball smashing a window causes (namely, just what its parts cause). But in fact, strictly Merricks denies that the baseball causes whatever its parts cause, because he denies the baseball exists. He thinks that if there were such physical composites, they would have the same causal powers as their parts, but he employs this as a premise in an argument that believing in the existence of wholes like baseballs is unnecessary. He reasons as follows: Baseballs, and other physical wholes, might or might not exist in addition to their microscopic parts. Whether they do exist or not makes no difference to what happens (since anything they would do, their parts are doing already), and since we should not posit things which make no difference to what happens, we should eliminate physical wholes from our theories (Merricks 2001). We should conclude that, strictly, there are no tables, no planets, no houses, no rocks, but just atoms "arranged tablewise," "arranged planetwise," etc.

Merricks, and other "nihilists" or "eliminativists" about ordinary physical wholes (e.g., Van Inwagen 1990; Unger 2006), clearly take it to be conceivable that the atoms composing a baseball exist, but no whole composed of them exists. (Indeed, they take it to be actually true.) But their arguments, it seems to me, are in fact an excellent reason for thinking that this situation is *inconceivable*. For the conclusion of their arguments—that tables, planets, houses, rocks, etc. do not exist—is highly implausible, and yet follows fairly reasonably from that claim of conceivability. If it is genuinely conceivable that there could exist atoms arranged tablewise but no tables, that would suggest it was a real possibility. But then we have to ask: What evidence can we find to tell us whether that

possibility is actual? And the way the choice has been set up, we will of course be able to find no evidence: because wholes do the same things that their parts are already doing, they are not the kind of thing whose existence would produce detectable evidence. I conclude, though, that the choice is badly set up: the nonexistence of tables despite the existence of atoms arranged tablewise is *not* an option, and to think of it as an option is to misunderstand the relation between parts and wholes. At least, it seems far more likely to me that these philosophers have misunderstood this relation in some way, than that there are no tables.

(p.107) To summarize my argument here: it seems obviously reasonable to believe in tables, but there is no empirical evidence in support of their existence. I conclude that the reasonableness of believing in them is not evidence-based, because their nonexistence is not a genuine option: when there are atoms arranged tablewise, *of course* there is a table, since it is just those atoms considered together.[18] If it were consistently conceivable that a table not exist, despite there being atoms arranged tablewise, that would support thinking that such a thing was possible, a genuine option. Since it is not a genuine option, we should suppose that it is not consistently conceivable.

Why is it not conceivable? Where is the contradiction? The contradiction lies in violating substantive indiscernibility, which is a special case of the contradiction involved in violating the indiscernibility of identicals. It is contradictory to say that Bob exists, and Robert is Bob, and yet Robert does not exist. Likewise, I maintain, it is contradictory to say that atoms arranged tablewise exist, and that a "table" is a whole composed of atoms arranged tablewise, and yet no table exists. But then the question recurs as: Why does it seem conceivable—to Merricks, to Van Inwagen, perhaps to you the reader, even to me when I get myself into the right mindset—that there could be atoms arranged tablewise, but no table?

I think it is because of an inflated idea of what wholes are supposed to be. I think atoms arranged tablewise logically entail the existence of tables because I think tables are (with a few caveats) just those atoms taken together: they are aggregates. But (as discussed in chapter 1, section 1.3.4) some philosophers have felt that there is a major difference between a "mere" aggregate and a "true unit," and that only the latter really deserves the title of "whole" (since aggregates, after all, are just their parts taken together!). If, when we talk about "wholes," we are really looking for true units, then aggregates will not really count. Aggregates, after all, are meant to be just their parts taken together, and so by the standards of a seeker for true units, calling them "wholes" is like a sleight of hand, a sneaky attempt to "launder" many into one. By the standards of a seeker for true units, views like panpsychist combinationism are perhaps better seen as nihilist about wholes: on such views there are no wholes, only parts (or, even better, only uncountable "stuff" that is neither singular nor

plural). After all, such views take human beings to be basically equal in their metaphysical status to tables, or any other sort of physical aggregate, and so if tables don't exist, neither do we. But to **(p.108)** my mind it is more reasonable to say that both we and tables exist, though neither of us is, ultimately, anything over and above a bunch of particles interestingly arranged.[19]

The conceivability of particles arranged tablewise without tables makes sense, if "tables" must be true units or nothing, since true units by definition have features that are not derived from their parts, and so will not be conceptually entailed by their parts. But I think it is a mistake to think this applies also to "tables" understood as aggregates; at least, it seems that going down this road even for aggregates has seemingly absurd conclusions, such as the nonexistence of tables. An inflated idea of what wholes should be makes it seem coherent to vary their existence and properties independently of those of their parts.

I think the same inflation is at work in the apparent conceivability of nonconscious wholes with conscious parts, like the microexperiential zombie. To conceive of such a scenario, we first assign some properties to the parts (including consciousness), and then introduce a further entity, the whole, to which we assign a different and independent set of properties. We construct, so to speak, separate files for the many parts and for the one whole, and what we write in one file need not also be written in the other. But if the whole involved is just an aggregate, then it is a mistake to construct two separate files. There is just one file, read by different programs; one reality, differently described. To construct a second file is implicitly to treat the whole as something emergent, a true unit of the sort that Merricks, Leibniz, and others are looking for.

But what separates the microexperiential zombie from things like tables is that a second confusion, over the nature of consciousness, is working to reinforce this first confusion, by connecting it to a real and important possibility which is easy to misdescribe, namely that of a non-access-conscious whole with access-conscious parts.

(p.109) 3.5.2. Division-Invariant Phenomenal Consciousness and Division-Relative Access-Consciousness

Here is something that is certainly true: just from some part of me undergoing an experience, it does not follow that I as a whole have cognitive access to that experience, or that it is informationally integrated with my other experiences, or that it will make a difference to my behavior. And in everyday talk this is precisely what would lead us to call it "unconscious": even though it is going on in me, I have no idea about it, and it makes no difference to what I do. In an important sense it is not conscious: it is not access-conscious.

According to many philosophers, that is all there is to consciousness: if some event in me is not cognitively accessible, if it does not play the right functional role in how I work and how I behave, then it is not a conscious experience (in

chapter 5 I call this “pure functionalism”). If these philosophers are right, then experience inheritance is false: since what plays the right functional role in some part of me may not play that role for me as a whole, consciousness is not division-invariant. On the plus side, though, a functional analysis of consciousness would allow for consciousness to be *intelligibly* division-relative, and thus would still be compatible with a form of combinationism. But that form will not be panpsychist combinationism: it will be either functionalist combinationism or psychological combinationism, laid out in chapters 5–8.

Panpsychist combinationism is a form of constitutive Russellian panpsychism, which rejects any functional analysis of consciousness. No complicated organizational structure is, by itself, sufficient for consciousness; while it may make a difference to the structure of consciousness, it does not account for raw experientiality, for phenomenal consciousness. And, according to panpsychism, phenomenal consciousness can exist in very very simple forms, without cognitive access, behavioral complexity, or any of the others things we usually associate with it. But this distinction between phenomenal consciousness and access-consciousness is in some ways a hard distinction to keep clear—because whenever we try to focus on the phenomenality of some experience we are having, we must do so by cognitively accessing it!

I think the conflation between these two senses of “consciousness” is partly responsible for the apparent conceivability of nonconscious wholes made of conscious parts. We intuitively recognize that a whole with conscious parts may not be *access*-conscious of their experiences, and indeed may not be access-conscious of *any* experiences. It may be entirely without anything we would want to call cognition. Indeed, most wholes containing conscious parts are like this: for any of a million different random assemblages of things we might consider, which **(p.110)** happen to include me, almost none will be consistently organized so that their overall functioning is sensitive to my individual experiences. To inherit access-consciousness from me, the whole needs to be organized in the right way, to give my experiences the same functional role in the whole as they have in me.

But that does not mean that these wholes lack phenomenal consciousness of my experiences. And if substantive indiscernibility, as I have interpreted it, is right, then they cannot lack that, because they are just me and some other things taken together. And phenomenal consciousness (we are assuming in this chapter) is not susceptible to any kind of analysis that would reveal that it means different things relative to different divisions (in the way that access-consciousness is susceptible), so there is no difference between the aggregate having that experience, considered as one, and its having that experience, considered as many things including me.

So what explains the apparent conceivability of nonconscious wholes with conscious parts is the interaction of the phenomenal/access confusion with the aggregate/true unit confusion discussed in the previous section. When I ask myself whether wholes containing me might lack consciousness, I first find myself struck by the very salient possibility of wholes containing me which lack access-consciousness of my experiences. And, second, because it is hard to keep phenomenal consciousness and access-conscious apart in imagination, this pushes me toward imagining that the wholes in question lack phenomenal consciousness of my experiences as well. And then, third, because of the ease of thinking of wholes as true units, floating free of their parts, I find myself easily able to keep simultaneously in mind the thought that the whole entirely lacks phenomenal consciousness while its parts do not, without feeling that the two are contradictory. But in fact both the second and third steps in this mental construction are mistaken.

3.5.3. Intuitive Cases of Mental Combination

I think it will help to illustrate and motivate these claims if I discuss a particular type of mental combination which is much more familiar than the sort panpsychists posit. It is an instance of what in chapter 1 I called “trivial combination,” where a composite subject’s consciousness is fully accounted for by the consciousness of just one part (and that part’s relations to the others). Consider, for instance, your own brain. Suppose that we were willing to say that the brain itself is conscious—there is something it is like to be a brain. Suppose, in fact, that the brain’s consciousness is just the sort of consciousness we typically ascribe to people; after all, it is widely accepted that all the conscious states we ascribe to people are in some sense based in the brain. Now consider the whole human being—the thing which has the brain, and all other organs—as parts. Could that **(p.111)** composite thing lack consciousness, given that it contains this conscious brain as a part? (A precisely analogous question can be asked of substance dualists: If the soul is a part of the whole human being, and the soul is conscious, could it be that the whole human being fails to be conscious?) It seems to me that the most plausible answer is no: the whole human being cannot fail to be conscious, because it automatically “picks up” or “inherits” the consciousness of its brain.

This is just how the fundamental physical properties work: the whole human being cannot fail to occupy the space its brain occupies (as well as that occupied by its other parts), to have the mass its brain has (as well as that of its other parts). To conceive of the whole human being as having parts with a certain mass, or location, but as itself lacking those features, is incoherent: it is to both affirm and deny certain properties of a single entity—because the whole human being just is its brain and other parts, considered together.[20]

Let us consider a few ways that someone might reject this account of the brain-person relationship. They might simply deny that whole human beings, of the sort I have described (which have organs as literal parts), are ever conscious. One way to do this is to say that such things are just human *bodies*, things with biological persistence conditions: conscious beings have psychological or phenomenal persistence conditions. But this is just to reject the metaphysical conception of subjects, which I am supposing here. (They would then, presumably, deny that brains are the right kind of thing to be conscious either.) Fortunately, denying strong independence is considerably *easier* on the psychological conception of subjects, as I discuss in chapter 7.

Another response would be to say that whole human beings are conscious "only in a derivative sense": brains are properly conscious, and whole people merely by courtesy (cf. Dainton 1998, 681–682; Bailey 2015). But surely the most natural thing to say, on first reflection, is that I myself am a whole human being, and that I myself am literally conscious, not "conscious" by some sort of polite or convenient fiction. So this move, though possible, is itself counterintuitive.

Might an objector insist that, even if human beings inherit the consciousness of their brains, it is at least still conceivable that they not do so? But this opens them up to some very unnerving consequences. If we cannot establish this inheritance *a priori*, then surely we would need to find some empirical evidence that human beings really are conscious (and not just unconscious vehicles for conscious brains). And even if we assured ourselves that they are, we would then need **(p.112)** to explain that: Why are human beings conscious? Could some be conscious and others not? It seems to me that these are not serious questions we need to empirically investigate (any more than we need to empirically investigate whether, in addition to particles arranged tablewise, there are tables); whatever the right answer is, it must be one that we settle on *a priori*.

Finally, an objector might say that the privacy of experiences breaks the analogy between experiential and physical properties. The reason I inherit the physical properties of my brain is that doing so requires no multiplication of property-instances, but merely their sharing between part and whole. Yet strong privacy rules out an analogous sharing of experiences. But we have already rejected strong privacy, while proposing weak privacy in its place. Indeed, the case of a human being and their brain intuitively supports denying strong privacy: if both are conscious, surely they do not have two streams of experience proceeding independently, but rather share a single stream of experience.[21]

Sutton (2014) analyzes cases like this as involving two beings that "share a supervenience base for" the relevant properties (622): if the part of something whose intrinsic features are sufficient for a given property to be instantiated is shared by another whole, then both wholes will instantiate that property in a

"non-summative" way (622). "Non-summative" means that though two beings instantiate the property, the total amount of that property is not thereby increased: there is just one instance of the property, but it belongs to multiple beings. Sutton offers this analysis, which explicitly allows for "two non-identical beings that think the same thoughts" (622) as not only plausible on its merits, but also as the best way to resolve the "too-many-minds" problem (mentioned in chapter 1). This problem faces any view that distinguishes a person from the living organism or material body that constitutes them at a given time, or which distinguishes the latter two objects: since all appear to be capable of thinking (they "share the equipment"), these theories seem to imply "too many minds." But the problem also faces any view which recognizes the existence of both humans and their heads. **(p.113)** According to Sutton, there are not too many minds because person, animal, and body all share the same mind. I take no stand on the proper analysis of persons, animals, and bodies, but I agree with Sutton that many beings can share the very same thoughts and experiences, if they overlap.

Finally, of course, an objector might point out that if the brain were isolated from the other organs—unable to control them, unable to receive sensory inputs from them, just sitting there like a tumor—we might not think that the whole human being shares the consciousness of the brain. And this is completely right: if the brain was differently connected to the rest of the body, the whole human being would not be access-conscious of the brain's experiences, just as the solar system is not access-conscious of my experiences. It may well be that access-consciousness and phenomenal consciousness are intimately connected, so that nothing can be phenomenally conscious if it is not access-conscious of at least some things, but in this chapter I am assuming the opposite: that phenomenal consciousness can and does exist without access-consciousness. In light of that, the fact that the whole human being is not always access-conscious of the brain's experiences does not entail that it is not phenomenally conscious of them.

3.6. What If Experience Inheritance Isn't *A Priori*?

I do not expect that the previous sections' arguments will convince everyone who worries that violations of EI are conceivable, and that EI is thus not *a priori* true. It is worthwhile, therefore, to say a bit about what follows if EI's *a priori* status is denied.

I think that the theory I have sketched (EI and MUH), or something like it, is what best follows through on the ambition of constitutive Russellian panpsychism, of placing consciousness at the base level of reality and treating it like other fundamental properties. And it does so most fully and most satisfyingly if EI is an *a priori* truth, grounded in the substantive indiscernibility of parts and wholes. But even if we give up on EI being *a priori*, we could retain the core and spirit of panpsychist combinationism, in at least two major ways: through a form

of pan*proto*psychism on which something analogous to EI is *a priori* true, or through a version of panpsychism in which EI is an *a posteriori* truth about our universe.

Both of these "fallback" options retain the idea that something about consciousness is fundamental. Someone obviously might reject EI's *a priori* status because they think it is simply a mistake to treat any aspect of consciousness as fundamental. In particular, if physicalism is true, and the only fundamental facts are those stable by physics, then all facts about consciousness are, ultimately, facts **(p.114)** about some very complicated sort of structure in the brain: all that it means for consciousness to exist is for nonconscious matter to be arranged the right way. In that case, though, panpsychist combinationism was wrong from the get-go, so its failure is not particularly surprising. Combinationism may still be true, but it will have little to say about the fundamental architecture of reality. Readers convinced of this kind of view should simply skip this chapter and the next one, and proceed immediately to chapter 5.

What would be more worrying would be for EI to not be *a priori*, even though something about consciousness is fundamental; this is the more pressing sort of objection, which I have encountered from several philosophers sympathetic to panpsychist combinationism. It is for them that I offer the following fallback options, tailored to two different sorts of critics. One sort thinks not only that EI is not *a priori* true, but that it is definitely false, because it implies conscious subjectivity in things which are *clearly* not conscious subjects. The other sort thinks EI might be true of our world, but is not a necessary truth.

3.6.1. Trading Experience Inheritance for Proto-Experience Inheritance

Consider first the critic who thinks EI is definitely false. They might say something like the following:

> *If EI were true, then piles of sand, clouds of dust, and even such widely scattered aggregates as "the sum of all the left toes in the Milky Way" are conscious subjects: it is like something to be them. But this is absurd; we know* a priori *that to be a conscious subject requires more complexity, or more integration, or more of something, than is possessed by random aggregates*.

Back in section 3.2, I noted that EI made sense only as a principle governing "bare" experiential properties, ones which do not presuppose self-awareness, unity, intelligence, and so on. In effect, this first sort of critic is saying that such properties are not really any sort of experiential property: once you subtract those things, what is left is not really conscious experience, and instantiating such "stripped-down" properties is not enough to be a conscious subject. This sort of critic may well be skeptical of panpsychism in the first place, since

anything as simple as an electron would likely not qualify as a subject by these standards. If they were nevertheless persuaded by the arguments usually taken to support panpsychism (e.g., that consciousness is not fully explained merely by physical structure), they might naturally gravitate toward "panprotopsychism," the idea that all the fundamental **(p.115)** physical entities are endowed with something "proto-experiential" but are not themselves conscious in the strict sense.

The fallback position I would suggest for critics inclined to think EI is definitely false is panprotopsychist combinationism. It is like panpsychist combinationism, but is panprotopsychist instead of being panpsychist, replacing EI with an analogous principle for proto-experiential qualities:[22]

> **Proto-Experience Inheritance:** Whenever a part of *x* instantiates a proto-experiential property (e.g., unexperienced qualities), *x* also shares that instance of the proto-experiential property.

The notion of "proto-experience" is admittedly rather obscure. Sometimes it is taken to be something completely unknown, but such that if we *did* understand it, we would be able to deduce from its nature that consciousness must arise from it under certain conditions (see Chalmers 1996, 2015). Less mysterian versions might say that the proto-experiential is something which we are acquainted with in consciousness, even though at the base level of reality is exists without consciousness. The most developed form of this is Coleman's (2013, 2015, 2017) constitutive Russellian panqualityism. Coleman analyzes consciousness as involving two key components: qualities and awareness. Qualities characterize the distinctive "what it is like" of each different experience, but without awareness there is no consciousness because there is nothing the experience is like *for* anyone. Coleman thus splits the task of explaining consciousness into two subtasks: explaining why there are qualities, and explaining why there are subjects who are aware of them. The latter subtask he attempts to address structurally: a subject exists whenever there is a complex system set up in the right way to represent its own states (an analysis borrowed from physicalists like Rosenthal [2005]). It is only qualities, he thinks, that are fundamental and inherent in matter.

Panpsychist combinationism already has much in common with Coleman's view. We agree that what is simplest and most basic in consciousness is inherent in matter, and that humanlike subjects arise when these raw materials are embedded in the right kind of sophisticated information-processing structure. We agree that the raw materials inherent in matter generate the experiences we have in a wholly constitutive way—complex but not in principle mysterious. Our disagreement **(p.116)** is simply about whether these raw materials—"what is simplest and most basic in consciousness"—are something that deserves to be

called “consciousness,” or whether that label should belong only to the sophisticated humanlike form.

Whatever protoexperiential properties are, proto-experience inheritance is meant to be derived, like EI, from the substantive indiscernibility of parts and wholes, together with the idea that proto-experiential properties are fundamental. It says that these properties are pervasive in nature, and moreover that for a whole to instantiate them is nothing over and above its parts instantiating them. But consciousness itself requires that these raw materials be organized into some sort of richer structure—most likely some minimum degree of causal interdependence, representational unity, access-unity, and global coherence, perhaps accompanied by such capacities as attention, introspection, memory, and so on. In chapter 5 I lay out functionalist combinationism, which examines precisely this process of constructing and then combining complex, intelligent conscious systems: those who find EI absurd may find that theory more amenable.

3.6.2. Treating Experience Inheritance as an *A Posteriori* Truth

Above we considered critics who say not only that EI is not *a priori* true, but that it is definitely false—and who thereby commit themselves to some sort of substantive *a priori* preconditions for being a conscious subject. A very different style of critic might say the following:

> *We cannot decide anything about EI* a priori: *it might conceivably be true, it might conceivably be false. There is nothing intrinsically absurd about the idea that, in the universe we inhabit, every composite thing, including widely scattered aggregates, is a conscious subject. Consciousness is a fundamental property of nature, and could in principle exist in any sort of distribution, with or without intelligence or unity. Nothing about either the nature of consciousness or the nature of composition can tell us whether composites of conscious parts are themselves conscious*.

To these critics I would suggest that, even if EI is not true *a priori* and necessarily, it is probably true contingently. That is, I would accept the possibility of microexperiential zombie worlds, where functional duplicates of us exist, made of the same conscious parts, yet entirely lacking in consciousness themselves. But this world is not such a world: in this world, wholes inherit consciousness from their parts.

This invites two questions: If EI is not true *a priori*, what justifies us in believing it to be true? And if it is not true necessarily, what explains its being true in the **(p.117)** actual universe we inhabit? I have one answer to the first question, and three alternative answers to the second. The answer to the first question is, in short, that EI is the simplest explanation of the fact that we are conscious. We know that we human beings are conscious, and everything we learn empirically

about the way the world works seems to show that we are neither fundamental (we are just grand assemblages or configurations of matter) nor fundamentally different from our surroundings (the way that matter is organized in a human being differs only by degree from the way it is organized in a water droplet, or a table, or a frog). The simplest hypothesis consistent with all of this (if we accept that physicalist reductions of consciousness fail) is that matter itself is conscious, and that assemblages of matter are conscious in virtue of this. More precisely, EI has the following virtues:

1. It is very simple to state: like the fundamental equations of physics, it could be "written on a T-shirt" (Lederman 1993, 21–22).
2. It is uniform: it applies to human beings no more or less than any other organism, and to organisms no more or less than to any other sort of system.
3. It is compatible with the continuity of nature: it does not try to draw any sharp boundaries onto the messy smear of reality, between systems that strike us as "integrated enough to be conscious" and those that do not.
4. It closely parallels the behavior of the fundamental physical properties, rather than diverging from them.
5. It is ontologically nonadditive: it does not require that putting together conscious parts generate anything "over and above" those parts, anything in addition to them; it simply says that their aggregate (i.e., them considered as one) is related to their experiences the same way they are.
6. And in virtue of being nonadditive, it avoids causal exclusion: since wholes have the very same experiences as their parts, when those experiences make a causal difference we can say that both parts and whole are simultaneously causally responsible.

I cannot say in advance that no better principle than EI could be devised, that would outdo it in these respects; if such a principle can be shown, that will only strengthen combinationism. And of course EI does not all by itself explain why humans experience things in the particular way that they do; that requires an account of how the multitude of simple experiences they inherit from their parts can add together into a recognizably human sort of experience, which the remainder of this book explores. But considering the subject-summing problem by itself, **(p.118)** apart from those other problems, I think EI is the most theoretically elegant principle by which to get conscious wholes out of conscious parts.

But what explains EI's being actually true in this world, if it is not true necessarily? Consider three suggestions. First, in section 3.4 I presented an argument that the widespread phenomenal unity posited by MUH implies composite subjects for every set of conscious parts. Perhaps it is the fact that phenomenal unity pervades the cosmos that explains EI. Or (second) perhaps

Goff is right to be “mysterian,” and what explains the existence of composite subjects is something about matter, or the relations among pieces of matter, which we do not have any positive conception of. On either of these two hypotheses, composite subjects are explained by a fundamental but contingent fact about our universe. Just as our universe happens to be a spatial universe, or a universe suffused with electromagnetic force, or a universe containing neutrinos, though there is no logical necessity to its being any of these things, so also it happens to be a universe tied together by some form of phenomenal bonding.

The third possible explanation of why EI is contingently true would be that it follows from the contingent truth of “cosmopsychism,” a version of panpsychism on which the universe as a whole is the most fundamental thing. As I noted back in section 3.1, it is still not clear what the fundamental physical entities are, and in particular whether they are very very small and numerous, or very very large and unique. (This question has some affinity with philosophical debates stretching from the ancient Greeks to the present day [cf. Schaffer 2010] about whether the universe is fundamentally many, forming one whole, or fundamentally one, refracting into many parts.) Different versions of panpsychism have been distinguished based on this question: “micropsychism” (which takes the fundamental conscious entities to be very very small and numerous) and “cosmopsychism” (which takes the fundamental conscious entity to be the universe as a whole).

I said in section 3.1 that this question actually matters little for evaluating panpsychist combinationism. This is for two reasons: firstly, empirically it seems as though all the physical facts about, e.g., a human body, can be traced to facts about particles and their interactions—even if those particles are in turn explained by facts about the universe as a whole. That is, micro-level entities still have explanatory priority over macro-level entities, so it makes sense to focus on how that micro-to-macro explanation works. Second, I think that in a certain sense the whole question of micropsychism versus cosmopsychism is mis-posed. That question is often framed as whether wholes are more fundamental than their parts, or parts more fundamental than their wholes. But it seems to me that opposing parts and wholes like this, as though they are in competition for priority, misses the key point, namely that neither is more fundamental than the other because they are **(p.119)** not really distinct. The whole is neither prior to, nor posterior to, its parts: it just is them. Whether we take them together or divide it into parts, it is the same reality; the difference is in our ways of conceiving and describing it. So in that sense panpsychist combinationism, deriving from the substantive indiscernibility of parts and whole, is neither micropsychist nor comospsychist.

But perhaps I am wrong. Perhaps it really does matter whether we start with a conscious universe and try to extract human-size subjects from it, or start with conscious particles and try to compose human-size subjects from them. In particular, there seems to be a feeling animating many contemporary cosmopsychists that the former is somehow easier than the latter, because it allows the human-size subjects to be genuinely real and yet be nothing in addition to the whole. By contrast, cosmopsychists suggest, if a human-size subject were formed out of conscious parts, it would either be strongly emergent, or it would not be genuinely real, but just a sort of convenient fiction (e.g., Goff 2017a, 209ff.; 2015; Nagasawa and Wager 2017; cf. Miller 2017; Albahari 2018; Shani and Keppler Forthcoming).

If it is true that middle-size subjects can be adequately explained by cosmopsychism but not by micropsychism, then panpsychist combinationism should be replaced by “cosmopsychist combinationism.” As with the other fallback options considered so far, this is not as great a change as it might seem. Even though I have spoken in “micropsychist” style so far, building up larger wholes out of smaller parts, I have been led to recognize principles that entail a universal subject: the cosmos itself inherits the experiences of all its inhabitants, and there is extensive unity among these experiences. In this sense the psychic cosmos is already implied by panpsychist combinationism; the only change to turn this into cosmopsychist combinationism concerns the metaphysical priority of this cosmic subject over other things. This is compatible with EI and MUH: EI simply says that when parts have experiences, so do wholes, while MUH says that when experiences stand in certain relations, their sum is an experience. Although it is natural to assume that the former clauses (parts having experiences, experiences standing in certain relations) express the more fundamental facts, this is not required; the more fundamental facts might instead be expressed in the latter clauses (wholes having experiences, composite experiences existing), which identify the grounds for the former.

3.7. Conclusions

The universe described by panpsychist combinationism, and its two key theses, experience inheritance and the micro-unity hypothesis, is in many ways a strange one. The starting point—that all fundamental physical entities are **(p. 120)** conscious—was strange enough. Because of EI, all aggregates of these fundamental entities will also be conscious, sharing the experiences of their parts: this is a ‘universalist’ form of panpsychism (cf. Buchanan and Roelofs 2018). And because of MUH, all aggregates whose parts physically interact will have composite experiences which unify those many simple experiences. Indeed, overlapping aggregates with interacting parts will share these composite experiences with each other. In short, consciousness works just like other basic physical properties. Every fundamental physical entity has some degree of, say, energy or location, and aggregates of these entities share the energy and location of their parts. Overlapping aggregates share some of their accumulated

properties with each other. If we wish to measure the total energy of a system, it does not matter whether we count it as one whole, as many tiny parts, or as two large parts: it is the same reality, with the same energy, however we carve it. Similarly, it has the same consciousness however we carve it. And just as the basic physical relations run indiscriminately throughout the universe, not respecting the boundaries of organisms or anything else of interest to humans, so does the basic phenomenal relation. The world that physics seems to reveal is one in which it is hard to recognize ourselves: a great endless stretch of mass/energy, with each of us just a complicated ripple in this sea, and no privileged way to divide it up into macroscopic parts. According to panpsychist combinationism, this world is also a world of consciousness, and its consciousness is structured in just the same way: an endless phenomenal field, with each of us just a complicated ripple in this sea, and no privileged way to divide it up into macroscopic subjects.

Notes:

(1) There is thus here a primitive form of the causal theory of reference: experiences refer to whatever causes them, just as we might think human perceptions and perception-based thoughts refer to whatever is causally responsible for them. Note that the particle's experience may refer without having truth-evaluable content—there may be something it is "about" without there being anything it "says" that could be true or false, rather like the primitive "ur-intentionality" posited by some enactivists about cognition (e.g., Hutto and Myin 2017).

(2) This sketch is partly inspired by Mørch's (2014) idea of the basic experiences as indissolubly combining an intrinsic qualitative nature and a causal tendency.

(3) In relativistic physics no interaction is instantaneous, so for any time period *t*, there will be a sphere around an object beyond whose perimeter it cannot have any causal effect within time *t*. (This zone of interaction is sometimes called a "light-cone," since it is conical in four dimensions.) But clearly phenomenal unity does arise, so if all interactions take some amount of time, then either phenomenal unity must take some amount of time to establish, or else the physical relation which is phenomenal unity is just spatial distance, not any kind of interaction.

(4) It might be worried that this gives wholes experiential properties "only in a derivative sense," not in the proper and primary sense in which we ourselves have them. But this is an equivocation on "derivative": while it is true that wholes have their properties in virtue of other things doing so, they still literally have those properties just as their parts do: a ten-ton weight that inherits its mass from its parts still literally weighs ten ton.

(5) By a “primitive property” I mean a property, the most revealing concept of which is primitive. A primitive concept is one that cannot be defined in terms of any other concepts, except those that are reciprocally definitionally dependent on it. I say “the most revealing concept of which” because one property might be represented by multiple concepts, including some which aren’t at all revealing of its nature (e.g., “the property I just thought of” is not a very revealing concept; it tells us nothing about the property it represents).

(6) Or both—it may be that neither the whole’s property nor the parts’ property is more fundamental than the other.

(7) This is a general pattern with many terms: compare the two readings of “fills this cup,” either as “fills this cup exactly” (which tells us the thing’s volume) or as “fills this cup and possibly more” (which tells us only a lower bound on its volume), or of “ate some of the cake” either as “ate a small portion of the cake and no more” or as “ate at least a small portion of the cake.”

(8) It is necessary to formulate causal powers here in ways that make no reference to their bearer as such. For instance, the power “to attract negatively charged particles” is implicitly the power to attract them to *oneself*, and so will mean different things when ascribed to a proton and to a building that proton is a part of. So we should instead specify this power in terms that do not make any reference to its bearer (e.g., “to subject negatively charged particles to a force of magnitude x and direction y”).

(9) Being arranged tablewise is a collective property: the parts (plural) are thus arranged, but no individual part could be said to be thus arranged by itself. But just as it is clear how the whole being table-shaped corresponds to the parts together being arranged tablewise, it is also clear how the latter fact requires that individual parts have certain (locational and relational) properties.

(10) At least, that is impossible as long as we hold everything else fixed: the molecule might lack a dipole moment simply because it has some other atomic parts whose charges and positions allow them to cancel out those of the others, or the molecule might have a dipole moment while the normal parts were not charged, if some other parts were. The point is that we cannot change things at either level without some corresponding change at the other level.

(11) Shoemaker (2002) calls these “micro-latent” causal powers, as opposed to “micro-manifest” ones.

(12) EI would then be nomologically necessary but not metaphysically necessary, though it would still be metaphysically necessitated by more basic facts about our universe, namely that it is exclusively composed of a kind of stuff for which SI holds.

(13) An argument very like this one is made at greater length in Goff (2013). This argument closely resembles the “continuity argument” for panpsychism itself, outlined in chapter 1, as well as the “vagueness argument” for unrestricted composition given by Lewis (1986, 212) and Sider (1997), and certain arguments concerning human subjects made by Parfit (1984, 231–243) and Unger (1979, 1990, 191–206).

(14) On “epistemicist” approaches (Williamson 1994), some particular one of these meanings is in fact the true meaning of “bald,” but we are unable to know which; other analyses differ in the role of contexts (e.g., Graff 2000), in whether the multiplicity of acceptable ranges of application is simply a failure of specification or a positive specification built into the meaning of the words (Fine 1975; Raffman 1994, 2013), and on other points. But the plurality of acceptable ranges of application is common ground.

(15) In the case of animals which lack the concept of “an experience,” and so cannot reflect on their total experience as such, there at least remains a capacity to survey all the different things they are conscious of—all the smells, sights, sounds, and so on—so as to access an overarching sense of their manifest environment as a whole.

(16) There is the following wrinkle: for all the subsumption argument shows, the subject undergoing the composite experience associated with one aggregate need not be distinct from the subject undergoing that associated with another aggregate. There might thus be fewer composite subjects than aggregates. But the principle of avoiding arbitrariness tells against this possibility: what precise, nonarbitrary principle could determine which aggregates are subjects and which aren’t *while ensuring* that every nonconscious aggregate is contained within at least one conscious aggregate? The only nonarbitrary view I can see here is the very extreme one that there is, in addition to the microsubjects, only *one* further subject, associated with the whole universe and undergoing all the composite experiences associated with any aggregate at all. Such a view implies that there is no such thing as me or you, unless both of us are identified with this one cosmic subject, and thus with each other. But even if this—incredibly radical—view is correct, there must be some sort of “figure of speech” by which we can make sensible claims about “Luke Roelofs” and other people, and their differing properties. After all, even if ultimately there is just the one cosmic mind, we want to capture the difference between “true” statements like “Luke Roelofs believes in combinationism” and “false” statements like “Everyone believes in combinationism.” Given that there must be such figures of speech available, all my talk of aggregates should be reinterpreted as involving such figures of speech.

(17) Chalmers (2017, 200–201) briefly suggests the possibility of developing a view like Goff's in this direction, with phenomenal unity as the bonding relation, but considers the boundary argument (which I discussed in sections 3.3.2 and 3.3.3) especially pressing against such a view. Cf. Miller 2018.

(18) This is a slight oversimplification: plausibly tables are structure-specific wholes, not aggregates. But, as I argue in chapter 5, structure-specific wholes derive their existence from aggregates instantiating the right structural properties.

(19) There is no need to dig in our heels over how to use the word "whole." I can happily accept the result that panpsychist combinationism says there are no wholes. Since there are no true units, there are no wholes by the standards of those who want their wholes to be true units. What we thought were wholes, including our own selves, are not single things at all: they are just many, many particles arranged in various ways. And what we each thought of as "my" stream of consciousness is in fact a stream of consciousness undergone collectively by these particles. The principle of experience inheritance, which I presented as connecting experiences in a whole and in its parts, should instead be understood as connecting experiences undergone by many things collectively with experiences undergone by many things individually: it says that whatever experiences one member of a group undergoes individually, the group members undergo collectively. Whether we refer to a group of particles experiencing things together as a composite subject is, in the end, a matter of semantics.

(20) As with the table, this is oversimplifying a bit: the human being is not *exactly* just the parts taken together, because it would cease to exist were they to be widely separated. It is a structure-specific whole constituted by the aggregate of those parts—it is them, taken together, arranged in a particular way. See chapter 5 for a fuller discussion.

(21) Another case that works to undermine strong privacy is that of conjoined twins fused at the skull. Such twins can have nerve tissue connecting their brains, and there is no reason in principle that there could not be shared brain parts, connected with and fully integrated into both brains. Would it not then seem reasonable for there to be a single experience, arising from this shared brain area, belonging simultaneously to both twins. An actual case does exist in which a "bridge" of nerve tissue connects the thalami of two twins, and anecdotal evidence indicates that this allows some sharing of perceptual information. Relatively little study has been done on this case because the twins, Krista and Tatiana Hogan, are still so young (Dominus 2011; Langland-Hassan 2015; cf. Montero 2017, 220). Since this case may involve more of a "bridge" between two brains than a fully shared brain structure, it is not clear whether we should think of the Hogan twins as literally sharing particular experiences;

nevertheless, an extrapolated case where that would seem the right thing to say, on both anatomical and functional grounds, is not hard to imagine.

(22) Panprotopsychist combinationism also, of course, needs to replace MUH with an analogous principle holding that one of the fundamental physical relations somehow functions to “unify” proto-experiential properties, where the relevant sort of “unity” is simply whatever is needed to allow distinct proto-experiential properties to generate a unified consciousness. (Coleman [2017, 261–262] at one point suggests quantum entanglement as the best physical relation to play this role.)