

Combining Minds

A Defence of the Possibility of Experiential Combination

by Luke Roelofs

A thesis submitted in conformity with the requirements for the degree
of Doctor of Philosophy, Graduate Department of Philosophy, in the
University of Toronto

© Copyright by Luke Roelofs (2015)

Abstract

Combining Minds: A Defence of the Possibility of Experiential Combination, by Luke Roelofs
Doctor of Philosophy, 2015, Graduate Department of Philosophy, University of Toronto

This thesis explores the possibility of composite consciousness: phenomenally conscious states belonging to a composite being in virtue of the consciousness of, and relations among, its parts. We have no trouble accepting that a composite being has physical properties entirely in virtue of the physical properties of, and relations among, its parts. But a long-standing intuition holds that consciousness is different: my consciousness cannot be understood as a complex of interacting component consciousnesses belonging to parts of me. I ask why: what is it about consciousness that makes us think it so different from matter? And should we accept this apparent difference?

‘Combinationism’ – the thesis that intelligibly constitutive composition is possible in the experiential realm – bears on many debates in the metaphysics of mind. Constitutive panpsychism’s need for combinationism is at the centre of recent criticism of the theory, but physicalists also need an account of how the consciousness, or lack thereof, in two cerebral hemispheres and a whole brain, or a human being and their head, or a social group and its individual members, can be intelligibly related. And further back in history, the supposed simplicity of the soul was held to rule out any form of materialism, in a tradition of argument stretching from Plotinus to Brentano. With an eye to this diversity of debates, I examine the prospects for combinationists with a range of different background views about the nature of consciousness, the ontological status of the subject, the behaviour of the physical part-whole relation, and the notions of constitution and explanation themselves.

The core difficulty for experiential combinationism is that subjects seem to be exclusive and independent in their experiences: no experience of one subject can belong to, or even be deduced *a priori* from the experiences of, another subject. This separateness is thought necessary to do justice to the privacy and subjectivity of experience, and seems to preclude the kind of ontological intimacy

between subjects that combinationism demands. In my third chapters I show that this conflict is soluble; a weakened form of exclusivity still preserves the distinctive privacy of experience, and is compatible with a composite sharing the experiences of its parts. A second major problem concerns the unity, interdependence, or even holism often attributed to each subject's experiential field. In my fourth and fifth chapters I develop a framework for accommodating and explaining this unity while still allowing component experiences to belong to distinct component subjects. Each experience in a unified field has a phenomenal character akin to that of *amodal perception*, indicating the other experiences that it is unified with.

Other challenges arise from particular sorts of wholes and parts. Panpsychists attribute consciousness to our smallest microscopic parts; this raises special problems concerning the grain and structure of consciousness, which I address in my fifth chapter. Other versions of combinationism attribute consciousness to large overlapping parts of a human being, parts capable of formulating self-referential thoughts. This generates problems for self-knowledge, which I address in my seventh chapter. And any combinationist, having claimed that consciousness in a thing's parts accounts for consciousness in the whole, must confront the issues I address in my sixth chapter, concerning those large disunified composite entities, such as galaxies or mereological fusions of people, which have conscious parts but appear to be unconscious.

Acknowledgments:

I owe thanks to many people for helping me in multifarious ways during the production of this thesis. Pride of place in this regard must go to my two supervisors, Jessica Wilson and Bill Seager, who generously provided their time and energy to comment on draft after draft of chapter after chapter. They constantly pushed me to expand the scope and range of my defence of combinationism, even while forcing me to make that defence more rigorous. The other members of my supervisory panel, Diana Raffman and Benj Hellie, also deserve thanks for their outstanding feedback and criticism.

I also owe thanks to all the many people who read and commented on parts of the various drafts this thesis progressed through, or essay-versions of material that came out of went into it: Mark Fortney, Benjamin Wald, Joshua Brand, Aaron Henry, Dominic Alford-Duguid, Jeff Hilderly, Mark Moyer, Dave Chalmers, Philip Goff, Sam Coleman, Hedda Hassel, Reier Helle, Bianca Bosman, Eli Shupe, Ole Koksvik, Pat Lewtas, Greg Horne, Dave Chalmers, Gurpreet Rattan, and Katlyn Beattie. Without their feedback, my work would have been incalculably impoverished.

Conference audiences and commentators also contributed importantly to the development of the ideas in this thesis, among them Greg Horne, Greg Glatz, Steve Pearce, Matthew Ivanovitch, and anonymous referees for *Thought: A Journal of Philosophy* and the *Journal of Consciousness Studies*.

Finally, I want to thank my family - Aurora Roelofs, Jacob Roelofs, Portia Roelofs, Ann Froggatt, and Andrew Roelofs - my partner Katlyn Beattie, and the whole graduate department at the University of Toronto: the former is an excellent family, the second an excellent partner, and the third an excellent graduate department. Without their support I would have gone a little mad.

Contents

Abstract	page ii
Acknowledgements	page iv
Contents	pages v-vi
Chapter 1: What is Experiential Combination?	pages 1-26
<u>Section 1: Defining my Project</u>	<i>p.2</i>
<u>Section 2: Why Do the Mereological Denials Matter?</u>	<i>p.8</i>
<u>Section 3: Versions of Combinationism</u>	<i>p.13</i>
<u>Section 4: Problems for Combinationism</u>	<i>p.21</i>
Chapter 2: What does Combination Require?	pages 27-57
<u>Section 1: <i>A Priori</i> Entailment</u>	<i>p.29</i>
<u>Section 2: Structural Simplicity</u>	<i>p.33</i>
<u>Section 3: Continuity of Natures</u>	<i>p.37</i>
<u>Section 4: Sharing, Inheritance, and Connections</u>	<i>p.42</i>
Chapter 3: Conditional Inheritance, Experience Sharing and the Problem of Subject-Summing	pages 58-118
<u>Section 1: The Subject-Summing Problem</u>	<i>p.59</i>
<u>Section 2: The Easy Case and the Inheritance Proposal</u>	<i>p.64</i>
<u>Section 3: In Defence of Experience-Sharing</u>	<i>p.72</i>
<u>Section 4: Components of Experiential Ownership</u>	<i>p.87</i>
<u>Section 5: Making Basic-Experience Inheritance <i>A Priori</i></u>	<i>p.103</i>
Chapter 4: Explaining the Unity of Consciousness	pages 119-177
<u>Section 1: The Boundary Problem</u>	<i>p.123</i>
<u>Section 2: What is the Unity of Consciousness?</u>	<i>p.133</i>
<u>Section 3: Amodal Perception and Perceptual Adumbration</u>	<i>p.144</i>
<u>Section 4: Adumbration as an Explanation of Conscious Unity, Given Transparency</u>	<i>p.157</i>
<u>Section 5: Adumbration as an Explanation of Phenomenal Unity, Denying Transparency</u>	<i>p.168</i>
Chapter 5: Blurring, Blending, and Mismatch - Microsubjects and the Structure of Experience	pages 178-227
<u>Section 1: The Blurring Problem</u>	<i>p.182</i>
<u>Section 2: The Blending Problem</u>	<i>p.197</i>
<u>Section 3: The Mismatch Problem</u>	<i>p.213</i>
Chapter 6: Group Minds and World Souls	pages 228-255
<u>Section 1: Reasons for Rejecting Mega-Subjects</u>	<i>p.230</i>
<u>Section 2: Could Mega-Subjects have Unified Consciousness?</u>	<i>p.237</i>
<u>Section 3: Human Bodies as Organic Mega-Subjects</u>	<i>p.248</i>
Chapter 7: Knowledge of Self, Other, and Part	pages 256-297
<u>Section 1: Why Don't Our Parts Notice Each Other?</u>	<i>p.259</i>
<u>Section 2: What Does 'I' Mean?</u>	<i>p.271</i>
<u>Section 3: Who am I?</u>	<i>p.284</i>

Chapter 8: Integration and Fusion

pages 298-345

Section 1: Recap

p.298

Section 2: Introducing Mind-Fusion

p.310

Section 3: Fusion from the Perspective of the Parts

p.321

Section 4: Fusion from the Perspective of the Whole

p.335

Glossary of Abbreviations

pages 346-347

Appendix: Why Experiential Combination Matters, and to Whom

pages 348-399

Section 1: From Mental Unity to Metaphysical Simplicity

p.350

Section 2: Homunculi and the Simplicity Diagnosis

p.356

Section 3: Panpsychism and the Combination Problem

p.362

Section 4: Split Brains and Split Persons

p.366

Section 5: Fission and Fusion

p.372

Section 6: Problems of Multiplication

p.378

Section 7: Anti-Nesting Among Materialists

p.386

Section 8: Collective Mentality

p.390

Section 9: Nihilism and Non-Singular Mentality

p.395

Bibliography

pages 400-414

Chapter 1: What is Experiential Combination?

<u>Section 1: Defining my Project</u>	<i>p.2</i>
<i>Subsection 1.1: Explanation and Grounding</i>	<i>p.3</i>
<i>Subsection 1.2: Composition and Combination</i>	<i>p.5</i>
<i>Subsection 1.3: The Three Mereological Denials</i>	<i>p.7</i>
<u>Section 2: Why Do the Mereological Denials Matter?</u>	<i>p.8</i>
<i>Subsection 2.1: The Achilles Argument and Materialism</i>	<i>p.8</i>
<i>Subsection 2.2: Panpsychism and the Combination Problem</i>	<i>p.9</i>
<i>Subsection 2.3: Anti-Nesting and the Simplicity Diagnosis in Physicalist Theorising</i>	<i>p.10</i>
<i>Subsection 2.4: Puzzling Cases and Paradoxes of Overlap</i>	<i>p.11</i>
<i>Subsection 2.5: Collective Minds and Collective Consciousness</i>	<i>p.12</i>
<u>Section 3: Versions of Combinationism</u>	<i>p.13</i>
<i>Subsection 3.1: Different Views of Experience</i>	<i>p.15</i>
<i>Subsection 3.2: Different Views of Composition</i>	<i>p.18</i>
<u>Section 4: Problems for Combinationism</u>	<i>p.21</i>
<i>Subsection 4.1: Explanatory Gaps and Latent Incoherence</i>	<i>p.21</i>
<i>Subsection 4.2: Micro-subjects, mega-subjects, and overlapping sections of subjects</i>	<i>p.23</i>
<i>Subsection 4.3: My Plan</i>	<i>p.25</i>
<i>Summary</i>	<i>p.26</i>

Much has been written about the relation of mind to body. A major issue that has not so far received a dedicated treatment is the apparent structural discrepancy between the divisibility of matter and the unity of the mind. Material things seem to be essentially and importantly composites, each part of which can be considered as a material thing in its own right, externally related to the other parts. The material world is one of galaxies containing nations, nations comprising human individuals, human individuals made up of organs and systems and cells, and cells made up of molecules, atoms, and yet smaller particles - yet it also seems to be a smoothly unified world, where the things at each level are nothing over and above the things at the level below.

Yet many have felt that minds are different: a mind is not composed of cooperating component minds, each seeing the others from the outside. If we try to imagine a conscious galaxy containing conscious nations, conscious nations made up of conscious humans, conscious humans made up of conscious organs and cells and yet smaller particles, we cannot then maintain the world we have imagined is smoothly unified, for the minds at one level will necessarily be ‘something over and above’ those at the level below. Indeed some have even claimed that a mind cannot have minds as parts, or

cannot have parts at all. This apparent discrepancy challenges any attempt to understand how mind and matter fit together.

In this work I will explore the arguments that have been adduced for the supposed indivisibility of the mind, and argue that we should reject them. I do not positively assert that conscious human beings are made up of conscious organs or particles, or that they enter into conscious nations or galaxies, but I try to provide a coherent and intelligible framework within which such claims can become plausible. In this first chapter I explain my terms, my motivations, different versions of my main thesis, and the challenges to it.

Section 1: Defining my Project

What makes the apparent indivisibility of the mind so striking is that in the physical case, we seem to have not only part-whole relations but highly *explanatory* ones. Understanding the physical facts about a thing's parts lets us understand facts about the whole: the micro-facts make the macro-facts intelligible.¹ So when I ask whether minds could be 'divisible', what I am most interested in is whether they could display this sort of intelligibility, whereby facts about their parts fully explain facts about them.

But what does 'explain' mean here? My major point of reference is the supposed explanatory gap between physical properties and consciousness, which contrasts with cases where one set of physical properties intuitively suffices to explain another (see e.g. Levine 1983, Chalmers 1995a, Loar 1990, McGinn 1989). A complete understanding of the microphysical structure and dynamics of a human brain would, in principle, let us 'see' why it must have the macroscopic physical features which

¹ Here and elsewhere, I use the prefixes 'micro-' and 'macro-' simply to indicate size: a micro-subject is just a microscopic subject, and micro-physics is just the physical facts about microscopic things.

it in fact has (Levine: “our knowledge of physics and chemistry makes [such connections] intelligible”, 1983 p.357), yet would not allow us to ‘see’ why it feels a certain way to be that brain (Levine: it “leaves the connection... completely mysterious”, 1983 p.357). The microphysics explains the macrophysics, but not the phenomenology.

In chapter 2 I examine what this talk of ‘explanation’ ‘intelligibility’, ‘understanding’, and so on really comes to. While many analyses are available, none is convincing enough to be adopted across the board, and so I will remain neutral, employing them collectively as a ‘tool-kit’ for sharpening and addressing questions about explanatory part-whole relations. The best I can do by way of fixing ideas is to index my discussion to the explanatory gap between consciousness and physics: could there be the kind of connection between mental wholes and mental parts which defenders of the explanatory gap think is missing between consciousness and physics, but usually accept between physical wholes and physical parts?

Subsection 1.1: Explanation and Grounding

Explanation is an epistemic relation, essentially connected to possible states of understanding by cognising subjects. But it is natural to associate it with metaphysical relations: perhaps one set of facts can be fully explained by another, in the relevant sense of ‘explain’, only if they are also metaphysically ‘grounded in’ the latter. So we might naturally associate the explanation of macro-facts by micro-facts with the grounding of wholes by their parts. The notion of ‘grounding’ at work here is just the rough and intuitive notion of one thing being ‘nothing over and above’ another. To use a popular metaphor, if *a* grounds *b*, then all God would have to do to create both is to create *a*: with *a* in place, *b* already and automatically exists. This general notion is meant as a placeholder for some

potentially more specific relation: either *a* realises *b*, or *b* is reducible to *a*, or *a* and *b* are identical, or some other sort of ontological dependence is involved.²

While it is plausible that explanatory relations generally reflect grounding relations, there are two complications. First, as I discuss more fully in subsection 3.2, wholes might be explained by parts in virtue of grounding their parts, rather than vice versa. That is, understanding the metaphysically more superficial fact might let us ‘see’ that the more fundamental fact must be the way it is, and so a given instance of explanation could run either parallel to or counter to the direction of grounding. Second, even if wholes are grounded in their parts, there might be some stubborn cognitive fact about us which prevents us from understanding the one in light of the other: there might be an explanatory gap even without a real gap, just as it has sometimes been argued that the explanatory gap between consciousness and matter does not betoken any metaphysical difference, but only a limitation of our understanding (e.g. Levine 1983, Loar 1990, Block & Stalnaker 1998, Diaz-Leon 2011). So someone *might* claim that some subject’s consciousness was wholly constituted by that of its parts, even though a full understanding of the latter failed to explain the former. However, such a postulate of ‘brute grounding’ is likely to be unsatisfying: it would seem ‘too easy’, with “the many advantages... of theft over honest toil” (Russell 1919, p.71).

In this work, therefore, I focus on the question of explanation. This does not reflect any lack of interest in metaphysical questions, but a belief that they are most profitably approached through a preliminary conceptual examination.

² Some authors have proposed recognising a distinctive, sui generis and perhaps primitive relation of capital-G ‘Grounding’ as a proprietary concept of metaphysics, to capture the intuitive idea of one fact holding ‘in virtue of’ another (e.g. Fine 2001, Rosen 2010). This notion would unify, replace, or supplement various more specific relations such as realisation, reduction, inherence, constitution, type-identity, and so on. However, I am sceptical of the usefulness of this distinctive notion, largely for reasons laid out in Wilson 2013-b.

Subsection 1.2: Composition and Combination

I use the term ‘combination’ to refer to the sort of explanatory part-whole relation that I am investigating (inspired by widespread discussion of ‘the combination problem’ for panpsychism, a phrase coined by Seager 1995, p.280). This means that ‘combination’ is something stronger than ‘composition’: composition is simply the part-whole relation, such that anything with parts is ‘composed of’ them.³ ‘Combination’ means composition in which facts about the parts fully explain all (intrinsic) facts about the whole; when qualified with an adjective, this explanatory relation is restricted to facts of a certain sort. Thus ‘spatial combination’ means a spatial thing having other spatial things as parts, with the spatial properties and relations of the latter fully explaining the (intrinsic) spatial features of the former; ‘mental combination’ requires the mental properties of a thing’s parts to explain the (intrinsic) mental properties of the whole.

Yet we should not take, say, physical combination to require that every single explanatorily relevant feature of the parts be a strictly and distinctively *physical* feature: that would exclude any appeal to ‘topic-neutral’ properties and relations, like causation, similarity, or temporal succession, which can coherently be ascribed to mental things, physical things, or any other kind of concrete thing. We should make room for any particular sort of combination to incorporate these properties and relations into its explanatory base, so that mental combination, say, requires that the mental properties of a whole be fully explained by the mental *and* topic neutral features of its parts.⁴

³ I take composition to be a synchronic relation, and so views on which a multiplicity of things are ‘fused’ into a single thing at a later time, the latter being diachronically explained by the former but not synchronically composed of them (e.g. Seager 2010, forthcoming, Mørch 2014), are not instances of combination in the particular technical sense I am investigating: they could be called ‘diachronic combination’ in contrast to the ‘synchronic combination’ I discuss.

⁴ What if the topic-neutral, non-mental features were sufficient, even without the mental features? More precisely, what if the experiential properties of the whole could be fully explained based on those topic-neutral properties of, and relations among, the parts, that do not themselves underly or explain any mental features of the parts? Call this ‘trivial mental combination’, as opposed to ‘substantive mental combination’ where the mental features play an essential role. In what follows, reference to ‘combination’ of any particular sort will mean ‘substantive’ combination of that sort.

Though my project bears on mentality in general, I will focus my investigation specifically on conscious experience, which is widely taken to be the hardest aspect of mentality to explain, and to which arguments for the indivisibility of the mind often make special reference. I define ‘experiential properties’ as properties which are individuated by what it is like to have them⁵: on this definition, anything that instantiates experiential properties is a subject, something which it is like something to be.

I take it that we can also talk about ‘experiences’, where the property ‘experiencing X’ is equivalent to ‘having an experience of X’. ‘Experiences’, unlike experiential properties, are particular tokens rather than universal types; talking about them therefore raises questions both about their ontology (what type of thing are they? events? property instances? tropes?) and about their individuation (could an experience have occurred a little later or with a little more intensity than it did? could a subject have two simultaneous and qualitatively identical experiences?). These questions will become important in chapters 2 and 3, where I defend the possibility of experience-sharing: single token experiences with multiple subjects. For now all I am committed to is the stipulative equivalence of ‘experiencing X’ and ‘having an experience of X’.

Experiences themselves do not instantiate experiential properties: subjects do. ‘Experiential combination’ thus involves the composition of subjects by other subjects: it goes beyond the familiar idea that many experiences might together compose a more complex experience. That is, people often suppose that ‘my present total experience’ contains as a part ‘my present visual experience’, which itself contains as parts experiences of colours, shapes, and so on. However, this is not experiential combination, but only the *combination of experiences*. For nobody assumes that there corresponds to

⁵ This is narrower than simply ‘properties which it is like something to have’: if I can desire X both consciously and unconsciously, then there is sometimes something it is like to have the property ‘desires X’, but what it is like is not essential to that property. Cf. Bayne & Chalmers 2003, pp.30-31, on ‘phenomenal states’ vs. ‘phenomenally conscious mental states’.

‘my present visual experience’ some subject which is conscious of only that. Similarly, while it is common to analyse the mind as composed of various faculties and drives, it is not at all common to infer that there is anything it is like to *be* one of those faculties or drives. Experiential combination is far harder to make sense of than the mere combination of experiences.

Subsection 1.3: The Three Mereological Denials

With this sense of ‘combination’ in mind, consider the following three theses:

Experiential Simplicity: A conscious subject cannot have parts.

Anti-Nesting: A conscious subject cannot have parts which are themselves conscious subjects.

Anti-Combination: The experiential properties of a conscious subject cannot be explained by the experiential (and topic-neutral) properties of, and relations among, its parts.

Of these claims, each is weaker than those before it, and implies those following it (given the above stipulation that experiential properties can only be instantiated by subjects. Negating them yields what we may call ‘the three mereological affirmations’, each stronger than those before it:

Experiential Compositeness: A conscious subject may have parts.

Experiential Nesting: A conscious subject may have parts which are themselves conscious subjects.

Experiential Combinationism: The experiential properties of a conscious subject may be fully explained by the experiential (and topic-neutral) properties of, and relations among, its parts.

Note that these six claims concern possibility, not actuality. Even if combinationism were granted, it would be a further step to conclude that any actual subjects of experience are in fact composite, nested, or explained by their parts.

Since Experiential Simplicity and Anti-Nesting both entail Anti-Combination, any arguments for the former will be arguments for the latter. Thus I can condense the question that I investigate in

what follows to: what reason is there to accept Anti-Combination? I will argue that these reasons are not conclusive, though they have some force given certain supplementary premises; thus the dissertation amounts to an extended though qualified defence of Experiential Combinationism (which I will generally call simply ‘combinationism’).

Section 2: Why Do the Mereological Denials Matter?

The import of the mereological denials is widely felt but rarely articulated. My appendix lays out how a series of philosophical debates are impacted by explicit or implicit adherence to one or more of these three theses. Here I briefly summarise this impact. Readers interested in a fuller discussion of combinationism’s relevance to these debates should consult the appendix.

Subsection 2.1: The Achilles Argument and Materialism

First, many opponents of materialism, especially in the Early Modern period, have argued that conscious subjects *must* be simple, and therefore cannot be material. For instance, in the *Critique of Pure Reason*, Kant labels this argument ‘the Achilles of all dialectical inferences of the pure doctrine of the soul’ (A351): apparently invincible but in fact fatally flawed, though he himself rejects it only as part of his overall rejection of speculative metaphysics. Other versions of this ‘Achilles argument’ appear in authors as far back as Plotinus (1956, pp.255-258, 342-356), as well as in Proclus (1963, p.163), Avicenna (1952, pp.47ff), Descartes (1985, Volume 2, p.59), Butler (1860), Mendelssohn (2002), Clarke (1978, Volume 3, p.759), Bayle (1991, pp.128-134), Lotze (1894, p.158) and Brentano (1987, pp. 290-301).

Despite the differences among presentations of the Achilles argument, certain themes recur, particularly the idea that “representations that are divided among different beings... never constitute a whole thought” (Kant 1781, A353). This claim is often supported by an analogy to groups of people, as in this quotation from Brentano (Cf. Plotinus 1956, p.346, de Courcillon, 1684):

If, when we see and hear, the seeing were the property of one thing and the hearing the property of another, then how could there be a comparison between colours and sounds? (It would be just as impossible as it is for two people, one of whom sees the colour and the other of whom hears the sound.) (Brentano 1987, p.293)

This suggests that a major driver of the Achilles argument is Anti-Combination: a composite cannot be conscious because no set of facts about its parts could add up to a unified consciousness for the whole. (This is reinforced by the role of a background commitment to physical combinationism, e.g. Clarke 1978, VIII p.759, Kant 1781, A352, Mendelssohn 2002, s.II, Bayle 1991, p.129.)

Subsection 2.2: Panpsychism and the Combination Problem

More recently, experiential combination has been intensively discussed by defenders and critics of panpsychism, the view that all the fundamental material constituents of the universe instantiate fundamental experiential properties, making them each a conscious subject. Given this commitment to ‘micro-subjects’, the panpsychist has an obvious interest in experiential combination, to ‘build up’ familiar human consciousness from this fundamental level.

Not all panpsychists are combinationists, but only those labelled ‘constitutive panpsychists’: they contrast with ‘emergentist panpsychists’, who make macrosubjects emergent relative to

microsubjects.⁶ The apparent impossibility of experiential combination is sometimes used as an argument by emergentist panpsychists against constitutive ones, and as an argument against panpsychists generally, by critics who assume that the only or best forms of panpsychism must be constitutive. One common reason for this assumption is the perception that non-constitutive panpsychism offers no advantage over non-panpsychist emergentism. Consequently many critiques of panpsychism involve arguing for Anti-Combination, and the arguments that arise in this process have been a major source of material for the following chapters.

Subsection 2.3: Anti-Nesting and the Simplicity Diagnosis in Physicalist Theorising

Debates over the relative merits of functionalism, behaviourism, and mind-brain identity theory have also been impacted by the mereological denials. For instance, Putnam's influential formulation of functionalism stipulates, in his suggested schematic definition of 'pain', that "No organism capable of feeling pain possesses a decomposition into parts which separately [satisfy this definition]" (1965/2003, p.215), this stipulation being intended "to rule out such 'organisms'... as swarms of bees as single pain-feelers" (p.216). Tononi also incorporates Anti-Nesting into his 'Information Integration Theory of Consciousness' (2012, pp.59-68, Tononi & Koch 2014, p.6): consciousness is integrated information, but if a system is contained within a supersystem with equal or greater informational integration, or contains a subsystem with greater, it cannot be conscious.

These principles are criticised by (Block 1978/93) and Schwitzgebel (2014), who present thought experiments involving outwardly intelligent complexes formed of a large number of minute aliens in close communication: they find it implausible that the presence of these minute aliens should

⁶ There are different definitions of 'constitutive' and 'emergentist' available; my usage of the labels may thus diverge from some other authors, since I define them by reference to the notion of explanation (admittedly itself somewhat flexible).

defeat our natural inclination to ascribe consciousness to the whole based on its behaviour. Yet Block himself makes much use of thought-experiments in which we are supposed to find consciousness in certain beings composed of other conscious beings implausible: the ‘Homunculus-Head’ and the ‘Nation of China’. Block regards consciousness in such entities as “an absurdity” (p.79), and therefore rejects forms of functionalism that imply it. While Block maintains that the mereological aspect of these cases is irrelevant to their force, others disagree: David Barnett has argued (2008) that the best explanation for these intuitive judgements is that “our naïve conception of a conscious being demands that conscious beings be simple” (p.309).

Barnett applies this diagnosis also to John Searle’s ‘Chinese Room’ (1980/2003), and to his own thought-experiment, where two human beings are shrunk to the size of someone’s cerebral hemispheres, trained to exactly imitate the latter, and inserted into someone’s head as a replacement (2008 p.312-315). In general, he claims, “whether the pair we consider is a pair of people, a pair of dogs, or a pair of inanimate objects... we have the intuition that the pair itself cannot be conscious.” (Barnett 2008, p.315)

Subsection 2.4: Puzzling Cases and Paradoxes of Overlap

Real cases, as well as thought-experiments, force us to evaluate the mereological denials. Consider two complementary pathological cases: the split-brain syndrome and dissociative identity disorder. Both cases seem somewhat intermediate between what we would normally count as one subject and what we would normally count as many subjects, but accepting Anti-Combination forces us to regard those as the only available (and mutually exclusive) options. Moreover, the reflections prompted by these cases – in particular, realising that each cerebral hemisphere can support consciousness even without the other – can reveal a similar problem in everyday cases (cf. Nagel 1971, pp.409), where

Anti-Combination again constrains us by insisting that if my hemispheres were conscious, that would necessarily have nothing to do with my consciousness, rather than intelligibly constituting it.

These questions do not stop at hemispheres: a human head, or a whole-human-being-minus-one-atom also seem intrinsically capable of being conscious. Merricks (2001) and Unger (1980, 2004, 2006) have both argued that under standard materialist assumptions, any situation we would count as containing “one person” actually contains a multitude of overlapping persons, which they consider sufficiently absurd to motivate revising those assumptions to avoid such multiplication. Both authors explicitly accept that a similar multiplication of inanimate objects is not similarly absurd (Merricks 2001, p.49, 95, 106, Unger 2006, pp.378-379).

What makes the difference is physical combinationism: any given material object can be viewed as many million overlapping material object, but each of these overlapping objects are metaphysically and explanatorily nothing in addition to the others. Similarly, if we ask how many *brains* the split-brain patient has, any uncertainty as to whether to say ‘one’ or ‘two’ seems merely semantic, because we can always shift away from a language of discrete brains and speak in terms of neural parts and their relations. We can say that all the normal parts of a brain are present, but they are no longer interacting as they were before. But when we come to conscious subjects, Anti-Combination makes cases involving them far more problematic, while combinationism would provide the conceptual flexibility to treat them analogously to their purely physical analogues.

Subsection 2.5: Collective Minds and Collective Consciousness

A different class of real-life examples involve humans composing a group agent, like an organised social group. Schwitzgebel (2014) argues that materialist premises should lead us to regard the United

States, and similarly complex collectives, as conscious. However, while many philosophers have argued for genuine collective mentality – most commonly intentions (Searle 1990, Velleman 1997, Bratman 1999), but also beliefs, desires, and all the cognitive and conative apparatus of agency (Pettit & List 2011) – genuine conscious experience in social groups is widely rejected. Even those who go furthest in defending genuinely collective mental states stop short of collective consciousness: Gilbert 2002 and Huebner 2011 both argue in support of collective emotions, but do so by trying to break the link between emotion and consciousness, arguing that genuine emotions may be devoid of phenomenology. Thus they seek to prevent “the implausibility of collective consciousness... impugn[ing] the possibility of collective emotions.”(Huebner 2011, p.102)

I believe that Anti-Combination explains this discrepancy in attitudes towards collective intentionality and collective consciousness. Authors wish to avoid irreducible, strongly-emergent group minds which stand above individual minds; such beings are regarded as mysterious, unparsimonious, and even ethically threatening (Searle 1990, p.404, Petit & List 2011, p.9). This allows for collective intentionality, since that intentionality can be explained through and grounded in that of individual members. But if consciousness in a whole cannot be explained by or grounded in consciousness in its parts, then this reductive attitude rules out collective consciousness. Combinationism dissolves this conflict: social groups can be nothing over and above their members, while still being just as literally conscious as their members. That is not to say that combinationism entails collective consciousness is ubiquitous or even actual; it merely shows that the possibility is not absurd.

Section 3: Versions of Combinationism

Combinationism is compatible with a wide range of views about experience and composition, but different background views can change both its importance and the power of the objections to it. For

ease of reference later on, and to indicate the range of views available, I will survey a few ‘types of combinationist’, explaining their differences and their different relationships to combinationism. All of these views fall within a certain broad grouping, defined both by ‘weak metaphysical naturalism’, and by a ‘level-connecting’ approach to composition. Views in this family give reason to expect, and to hope for, the possibility of experiential combination, for they imply both physical combination and a deep unity between the experiential and physical realms.

Weak metaphysical naturalism is simply the view that consciousness is a natural phenomenon in a broadly material universe. More precisely, it holds that the stuff of which paradigm material objects are made, and which natural science investigates, is the only type of stuff in the concrete universe. If there are concrete things made of some other kind of stuff, naturalism is false; if different paradigm material objects, like spoons, brains, and toenails, are made of fundamentally different substances, then naturalism collapses. According to naturalism thus construed, what differentiates conscious humans from inanimate objects is not the kind of stuff they are made of but simply how that stuff is arranged: this covers both physicalism and many forms of property dualism or neutral monism.⁷

A ‘level-connecting’ view of composition is any view on which the complete set of fact at one mereological level can ground all the facts about all the other mereological levels. On such views, distinguishing among mereological levels never means distinguishing two *independent* sets of facts, though they may disagree on what the privileged or fundamental level is. This conflicts with ‘autonomous-level views’, on which facts at different mereological levels, or facts that span

⁷ Weak metaphysical naturalism contrasts with strong metaphysical naturalism, which holds that the only fundamental properties are those studied by the natural sciences, and methodological naturalism, which holds that the only appropriate methods for seeking knowledge are those of the natural sciences.

mereological levels (such as ‘emergence laws’) may be equally fundamental, ontologically independent, and not explicable through each other even in principle.⁸

Subsection 3.1: Different Views of Experience

My first two divisions are that between physicalist ‘primitivist’ combinationists, who differ on whether all experiential facts admit of complete explanation in non-experiential terms,⁹ and that between panpsychist and non-panpsychist combinationists, who differ on whether experience is present throughout the fundamental parts of physical reality. These two divisions usually correlate, since panpsychists are much more likely to be primitivists, and because non-panpsychist primitivists are typically committed to strong emergentism and so are not combinationists. But exceptions are possible, such as physicalist-panpsychists, who hold that experience is reducible to an omnipresent but ultimately physical or functional property (Tononi’s ‘Integrated Information Theory’ comes close to this).

Panpsychists and their critics have been especially interested in the issue of experiential combination recently, because it is central to their claim to offer a better explanation of macroexperiential facts than physicalism can. The non-panpsychist combinationist has a different sort of need for combinationism: rather than making component subjects part of the ultimate explanatory base, they simply need to reconcile them with composite subjects so as to make their co-existence plausible. For instance, the puzzles discussed in subsection 2.4 and 2.5, about parts of humans and

⁸ Does a level-connecting view of composition rule out ‘strong emergence’? Yes, if that requires equally fundamental laws at both ‘emergent’ and ‘base’ levels, or crossing levels. But if strong emergence simply requires fundamental features of wholes that are not grounded in facts about their parts, it is compatible with monistic level-connecting views, on which all facts are ultimately explained by and grounded in facts about the universe as a whole.

⁹ The most obvious way to be a primitivist about consciousness is to be a ‘property dualist’, but the view that experience is irreducible to physical or functional facts is compatible with views like idealism and Russellian monism, on which physical properties are reducible to experiential ones.

groups of humans, confront the non-panpsychist as much as they do the panpsychist, even if the former accepts physicalism and hopes to ultimately explain both the parts and the whole by fundamental physical facts.

A second division concerns the possibility of ‘phenomenal overflow’: conscious experiences of a subject who cannot access them. The everyday term ‘consciousness’ can mean both pure phenomenality (‘there being something it’s like’) and cognitive access by a range of ‘content consuming systems’ (Block 1995), but it is disputed whether these two meanings can come apart. (I take ‘experiential’ to be synonymous with ‘phenomenally conscious’.¹⁰) On the ‘no overflow’ position (e.g. Rosenthal 2007), my states can be phenomenally conscious only if I can report them, commit them to memory, use them to guide future plans, etc. This is often expressed with the slogan that “conscious states are simply mental states we are conscious of being in” (Rosenthal 1986, p.329). On this view, we can have fairly direct empirical evidence for whether a given state is phenomenally conscious: its manifestation in report, memory, planning, etc.

On the ‘pro overflow’ position (defended by Block 1995, 2005) there can be phenomenal states whose subject is unaware of them – what it’s like to be me could go beyond what I can report, reflect on, etc. Certain empirical results – such as change blindness and inattention blindness – have been argued to support this possibility (Block 2005, 2011). Once it is accepted that phenomenal consciousness can outstrip access-consciousness, it becomes an open possibility that we might find reason to postulate vast reaches of inaccessible phenomenal consciousness – perhaps most of one’s current consciousness lies outside what can be accessed.¹¹ This possibility affects what needs to be

¹⁰ I will assume that thoughts and judgments, as well as sensations and emotions, can be phenomenally conscious, as is held by what Bayne 2010 calls ‘liberals’ about consciousness (e.g. Horgan and Tienson 2002, Siewert forthcoming), as against ‘conservatives’ (e.g. Tye 2003, Prinz 2012). Nothing I say will hinge on this position.

¹¹ For panpsychists this possibility is particularly live (as recognised by Block 2007, pp.535-536) since they must accept that phenomenal consciousness, in its simplest forms, does not require any sophisticated behavioural accompaniment.

explained: on a no-overflow position, combinationists must explain not only the *presence* of certain experiences in the whole, but also the *absence* of whatever is not accessible to that whole. On a pro-overflow position, some of the parts' experiences might be present but inaccessible in the consciousness of the whole.

My third division is between what I call the 'experience-first' and 'subject-first' views. On the former, experiences are ontologically prior to the entities we say 'have' them: a subject is nothing over and above a certain collection or stream of experiences, suitably related in some fashion. On the latter, the thing that has experiences is some sort of enduring substance, which is more basic than experiences and could exist without them. This substance might be an immaterial soul, a material thing with non-physical experiential properties, or something purely physical. The classic experience-first view is Hume's account of the mind as a 'bundle of perceptions' (1888, p.252), though other versions of the view may not *identify* the subject with any collection of experiences (Parfit 1984, 1999, Dennett 1990, cf. Strawson 2009). Experience-first and subject-first combinationists relate in different ways to the combination of experiences, as distinct from experiential combination, for the experience-first view takes 'part' and 'whole' to have a specifically experiential meaning when applied to subjects: for one subject to contain another as a part is for the set of experiences which constitutes one to contain as a subset those which constitute the other.

The subject-first view seems implicit in our habit of ascribing experience to tangible human beings, and regarding people as existing even when wholly unconscious. An adherent of the experience-first view might analyse this as shorthand for ascribing the experiences to a subject who then 'has' but is distinct from a certain material body. Nevertheless, most experience-first combinationists will at least accept that there *are* enduring substances which metaphysically underlie experiences (e.g. human brains): call these the 'bases' of experience, so that the subject-first and

experience-first views disagree over whether an experience's basis is also its subject.¹² The experience-first combinationist may end up saying things about part-whole relations among bases that are similar to what the subject-first combinationist says about part-whole relations among subjects. Alternatively, they might reject even bases (e.g. process-based ontologies, where events are more fundamental than objects), putting them in a similar position to the physicalist combinationist: their fundamental explanatory base does not involve any micro-subjects, but they still need an account of how one subject (and the experiences that constitute it) can relate explanatorily to another (and the experiences that constitute it).

Subsection 3.2: Different Views of Composition

Combinationists might take a number of views on the part-whole relation, within the broad class of level-connecting views. The most natural view is probably an unspecific sort of 'priority pluralism' (as defined by Schaffer 2010; cf. Nagel 2004, p.138, Russell 1985 p.36, McTaggart 1988, p.271), holding that composites are entirely ontologically grounded in their parts, though nevertheless distinct from them. This often goes with the idea that composition is widespread but not necessarily unrestricted. This sort of vague pluralism contrasts with three more specific doctrines about the part-whole relation, some of which raise specific issues for combinationism.

The 'composition-as-identity' combinationist holds that the parts of something are, collectively, numerically identical to that thing. 'Parts' and 'whole' are different forms of description for the same reality, so that the properties of wholes are trivially determined by those of parts, and vice versa (see

¹² There is also room for views intermediate between pure subject-first and experience-first views, such as Shoemaker's Neo-Lockeanism (1997, 2003a), which takes experiences and subjects as co-constitutive and dependent on a material basis, or the idea that subjects and experiences are two ontologically independent things, which become related in some way (see Unger 1990, p.177-184).

e.g. Baxter 1988, Lewis 1991 pp.80-87, Cotnoir 2013, Baxter & Cotnoir forthcoming). On this view, as on pluralism, the whole is grounded in ('nothing over and above') its parts, but the parts are not ontologically prior to the whole, for nothing is ontologically prior to itself.¹³ Composition-as-identity likely also implies 'universalism' about composition, since everything is self-identical: it is unclear how a set of things could fail to compose anything (though see Cameron 2011). Given universalism about composition, there is increased scope for experiential combination, since more overlapping conscious entities will exist.

The monist combinationist holds that facts about parts are grounded in more fundamental facts about wholes, and everything is ultimately grounded in the properties of the maximal whole of which everything is a part (Schaffer 2010, 2012). On an extreme version of this view, there is only one thing (Horgan & Potrc 2012): what we think of as a thing's parts are in fact logical constructions out of its properties.¹⁴ The conjunction of monism with panpsychism, holding that the fundamental reality is a conscious cosmos, has recently been labelled 'cosmopsychism' (coined by Gaudry, 2008).

It may seem strange to hold that micro-facts both explain, and are grounded in, macro-facts, but these two claims need not be at odds. Understanding something may shed light not only on what it grounds, but on the more fundamental reality that grounds it.¹⁵ Thus a pluralist's account of what micro-facts generate certain macro-facts can be taken over by a monist as an account of what

¹³ There may be ways to make out a claim of ontological priority between a thing and itself relative to different descriptions: counterpart theorists can take a single thing to have different modal profiles under different descriptions, and might take one description to thereby qualify as more fundamental (this is considered in Wilson 2013-b, with a nod to Lewis 1971 and Gibbard 1975). Jenkins (2011) suggests that discussions of metaphysical dependence induce intensional contexts where Leibniz's Law will fail. Nevertheless, identity claims imply grounding without *automatically* implying any asymmetry. Thus one can endorse composition as identity while denying pluralism.

¹⁴ There are also hybrid options: one might think that the cosmos has ontological priority over fundamental particles, which in turn have ontological priority over all the other sub-cosmic entities they compose.

¹⁵ To give a non-mereological example, suppose that a statue is grounded in the lump of clay that it is formed out of, and that its having a certain shape is grounded in the clay's having that shape. Given that clay has that shape, it is intelligible ('makes sense') that the statue does too, and we can deduce that the statue must have that shape. But we can also deduce that the clay must have that shape, from knowing that the statue has that shape, and it is intelligible that, *given* the statue has that shape, the clay must too.

micro-facts ‘reflect’ or ‘express’ those same macro-facts. Monists, as much as pluralists, need to vindicate their claims about ontological priority by explaining how their purported fundamental basis gives rise to the observed middle-sized world. A system of intelligible cross-level connections in the experiential realm is useful to both.

Finally, the nihilist combinationist denies the literal existence of composite things: there are only trillions of particles, and what we think of as ‘wholes’ are just logical constructions out of these simple objects (Sider 2013; the model for many nihilistic accounts of ordinary language is Van Inwagen 1990, pp. 98-114). Combinationism might seem incompatible with nihilism: if there are no wholes, what is there to explain? But the type of relations which combinationism claims between parts and wholes are precisely those relations which the nihilist can easily do away with: if all the putative facts about experiential wholes are fully explicable through facts about their parts, then nothing is lost if the former are eliminated in favour of the latter. If the parts, so to speak, do all the work, then cutting out the whole is more defensible. Thus the nihilist combinationist endorses something that preserves the spirit of combinationism:

Experiential Combinationism_N: The experiential properties of some jointly-conscious things may be explained by the experiential properties of, and relations among, those things.

The above views conflict in various ways – some deny the reality of wholes, some deny the reality of parts, and some identify parts with wholes. Nevertheless they are all ‘level-connecting’ views, agreeing that the universe is not a fundamentally multi-level place, with independent facts obtaining at different mereological levels. Pluralism and nihilism take everything to be built up from the ‘bottom’ level, holism takes everything to trickle down from the ‘top’ level, and composition-as-identity takes all levels to be different ways of describing the same reality.

Section 4: Problems for Combinationism

Why should we accept Anti-Combination? While some authors reject experiential combination out of hand with little explanation, others present challenging arguments against it. The least articulated objection is simple incomprehension of how it could work, as in Nagel's claim that:

We cannot at present understand how a mental event could be composed of myriad proto-mental events on the model of our understanding of how a muscle movement is composed of myriad physico-chemical events... We lack the concept of a mental part-whole relation. (1986, p.50)

But to develop such a concept we must survey and analyse the problems it faces. In this section I do so, starting with problems that confront any version, and moving on to those which deal with specific cases.

Subsection 4.1: Explanatory Gaps and Latent Incoherence

There are two types of argument against experiential combination *per se*: that experiential parts cannot explain experiential wholes because of how *little* they imply, and that trying to connect experiential parts with experiential wholes in an explanatory way will imply *too much*, yielding incoherence.

The former arguments, which I will call 'explanatory gap' arguments, are divided by Chalmers into three sub-problems: explaining the structure of macroexperience, explaining the qualities found in macroexperience, and explaining how macrosubjects can be constituted by distinct microsubjects (forthcoming-a, pp.4-5). This type of argument is often presented through the use of examples. First we imagine a group of conscious human beings, standing in a circle or holding a conversation or related in some other fashion: we observe that nothing need follow about any composite consciousness. But if

nothing follows, then how do these individual minds explain anything? And how could component consciousnesses go any way towards explaining our, supposedly composite, consciousness? This style of argument appears in James's seminal discussion of the combination problem for panpsychism (1890, p.160), but also appears repeatedly in earlier authors arguing against materialism (cf. subsection 2.1) which they construed as implying combinationism. A more elaborate way of pushing this argument, employed by Goff (2009a) would imitate conceivability arguments against physicalism, by considering a 'microexperiential zombie', a physical and functional replica of a human being, every part of which is conscious, but which lacks consciousness as a whole. A third way of presenting an explanatory gap argument is to imagine a change in our parts, with them gaining or losing consciousness, and observe that this seems compatible with our consciousness being largely unchanged. Block argues along these lines against Anti-Nesting (1978/92, pp.77-78), and Sebastian (2013), drawing on Chalmers (1995b) has formulated such a challenge to panpsychism.

Explanatory gap arguments against experiential combination seem much more compelling than any parallel arguments against physical combination: this applies especially to the explanation of macrosubjects (Goff 2009a, 2009b, 2010) I believe one major reason for this is the sense that consciousness is 'private': each subject's experiences are 'cut off' from those of others in a way that their physical properties are not, breaking any explanatory connection between them or at least making it irredeemably mysterious.

The other class of arguments, which I will call 'latent incoherence' arguments, take two forms. The first claims that because of the holistic character of each subject's 'perspective', no such perspective can contain others. Any given element they share would have to have two phenomenal characters: one *qua* element of the whole's perspective (reflecting its unity with the rest of that perspective) the other *qua* element of the part's perspective (reflecting its unity with the rest of that

perspective). Call this the ‘incompatible characters problem’. Arguments of this sort (prominent in recent work by Coleman, 2013, and Basile, 2010), could be expressed dramatically by saying that having parts would ‘tear apart’ the unity of any supposedly composite perspective.

A different worry could be put by saying that combining into a whole would ‘dissolve’ any supposed component minds. This worry is expressed particularly by Rosenberg (1998, 2004), and suggested in certain remarks by Dainton (2011, pp. 257-259); I follow Rosenberg in calling it ‘the boundary problem’. The incompatible characters problem and the boundary problem work together to make experiential combination appear incoherent: either the parts retain their distinctness and rip the whole asunder, or the whole maintains its unity and swallows up the parts.

A combinationist needs to address both types of argument, though in different ways. With latent incoherence arguments, they need to identify a move in the argument which is unsound or invalid, so as to avoid the purported consequences – or else explain why a consequence considered absurd is actually not. But explanatory gap arguments demand something more constructive; an exhibition of how and why certain sorts of composite mentality are explained fully by certain configurations of component mentality. That is, the explanation which is being asked for must to some extent actually be given, and so the burden of proof lies with the combinationist.

Subsection 4.2: Micro-subjects, Mega-subjects, and Overlapping Sections of Subjects

As well as the above problems that apply to any sort of combinationism, there are also specific issues that arise from what we might call ‘exotic subjects’. First, consider ‘micro-subjects’, conscious beings far smaller and simpler than we can easily imagine, at the level of cells and molecules and even fundamental particles. Panpsychist combinationists are already committed to the existence of such

beings, and monist combinationists might end up committed to them: if the consciousness of the universe is expressed in the consciousness of its human-sized parts, why wouldn't the same principles imply that the latter expresses itself as consciousness in smaller parts?

Setting aside the plausibility of micro-subjects *existing at all*, does it really make sense to think, as panpsychist combinationists do, that *our* complex consciousness is based on these trillions of micro-consciousnesses? It has been argued (Maxwell 1978, Lockwood 1993, Coleman 2012, Chalmers forthcoming-a) that such an aggregation of many tiny parts would have to produce a form of experience very unlike ours – one that was *qualitatively homogeneous, spatially structured, and enormously fine-grained*, while ours is *qualitatively diverse, non-spatially structured, and relatively coarse-grained*.

Next consider 'mega-subjects', conscious beings larger and more spatially disparate than humans, such as nations, mereological sums of people, or the cosmos itself. Cosmopsychist views, especially in the idealist tradition of Hegel and Schopenhauer, are already committed to the existence of such beings, and other combinationists may find themselves committed to them by principles which derive experiences in a whole from experiences in its parts. If we did accept the existence of mega-subjects, a dilemma arises: either their consciousness is unified, or it is not. If it *is*, then we can ask why our individual experiences as human beings seem so separate, so cut off from each other; if it *is not*, we can ask whether it really makes sense to posit such massively disunified consciousness.

Third, large overlapping sections of subjects pose distinctive epistemological problems. Beings such as my head, my brain, and my left half are, if conscious, plausibly capable of self-awareness, and of distinguishing themselves from others. But their overlap seems to make this problematic – for they must share the self-referential thoughts in which they would think specifically of themselves. And there

is a deep and challenging question about how *their* difficulties of self-identification might imperil the epistemic capacities of the whole they compose.

Subsection 4.3: My Plan

These problems will be successively addressed in the coming chapters. Chapter 2 lays the ground for this discussion by examining in more detail the notion of ‘explanation’ at play, and investigating physical combination as a standard. In Chapter 3 I then directly confront the explanatory gap problem, through an examination of experience-sharing, experiential ownership, and the idea of property inheritance. I will argue that combinationists have good prospects if they endorse the principles I call ‘Weak Sharing’ and ‘Conditional Experience Inheritance’. I also discuss here how different versions of combinationism may become committed to micro-subjects or mega-subjects, in their attempts to close their explanatory gap.

In Chapter 4 I address latent incoherence arguments, and expand on the explanatory proposal of the previous chapter, through an exploration of the unity of consciousness. I will argue that combinationists have good prospects if they analyse the unity of consciousness as involving a relation of ‘adumbration’ holding among experiences, whereby each gives its subject a partial, indirect, ‘sketchy’ awareness of the others.

In the next three chapters I examine the issues raised by exotic subjects: chapter 5 considers the structural issues raised by supposing our consciousness to be constituted by that of micro-subjects, chapter 6 considers the possibility and status of mega-subjects, and chapter 7 considers the epistemic problems raised by large overlapping sections of subjects. Finally, in chapter 8 I draw together the ideas

of the preceding chapters through a detailed consideration of a particularly challenging thought-experiment, that of two persons fusing into one.

Summary:

In this chapter I introduced a cluster of theses regarding the mereological structure of the mind –whether conscious subjects can have parts, whether those parts may be themselves conscious, and so on. But I identified one particular thesis as crucial, and as the target of my investigation in what follows:

Anti-Combination: The (intrinsic) experiential properties of a subject of experience cannot be explained by the experiential (and topic-neutral properties) of, and relations among, its parts.

This is a claim about conscious experience, rather than mentality in general, and about explanatory composition ('combination'), rather than composition in general. I called the negation of this thesis Experiential Combinationism; someone who endorses Experiential Combinationism can be called a 'combinationist'. This dissertation amounts to an examination of the dispute between combinationists and their opponents.

Combinationists may be physicalists or panpsychists, accept or reject phenomenal overflow, make subjects or their experiences ontologically prior, and hold a number of different views on the part-whole relation, ranging from nihilism to monism. But they all face a variety of serious challenges, which I divided into explanatory gap challenges (addressed in chapter 3), latent incoherence challenges (addressed in chapter 4), and challenges relating to various exotic subjects (addressed in chapters 5-7). While each challenge has some force, I think that combinationism emerges from the discussion in the next seven chapters as a coherent and defensible view on the metaphysics of consciousness.

Chapter 2: What does Combination Require?

<u>Section 1: <i>A Priori</i> Entailment</u>	p.29
Subsection 1.1: <i>A Priori</i> Entailment and Empirical Knowledge	p.30
Subsection 1.2: <i>The Case of Mass Additivity</i>	p.32
<u>Section 2: Structural Simplicity</u>	p.33
Subsection 2.1: <i>Simplicity in the Set of Principles and in Individual Principles</i>	p.35
<u>Section 3: Continuity of Natures</u>	p.37
Subsection 3.1: <i>The Problem of the Bonding Relations</i>	p.38
<u>Section 4: Sharing, Inheritance, and Connections</u>	p.42
Subsection 4.1: <i>Physical Cases of Token-Sharing and Conditional Inheritance</i>	p.43
Subsection 4.2: <i>Unconditional Inheritance - Location Properties</i>	p.45
Subsection 4.3: <i>Unconditional Inheritance - Causal Powers</i>	p.47
Subsection 4.4: <i>Relations among these Principles</i>	p.51
Subsection 4.5: <i>The Schema Applied to Experiential Combination</i>	p.53
Summary	p.55

In the last chapter I defined 3 theses: that conscious subjects cannot be divisible into parts, that they cannot be divisible into parts which are themselves conscious subjects, and that their consciousness cannot be explained by the experiential (and topic-neutral) properties of and relations among their parts - i.e. ‘experiential combination’ is impossible. In the rest of this work I make a case for rejecting these, and in particular the third, weakest, thesis, which I called Anti-Combination. But evaluating claims about combination requires some understanding of the notion of ‘explanation’ at play (and related notions like ‘understanding’, ‘intelligibility’, and ‘mysteriousness’), and so this chapter is devoted to distinguishing and contrasting available accounts of explanation. I discuss *a priori* entailment, simplicity of laws, continuity of natures, and identity of tokens: none is clearly the correct model to adopt, and so rather than deciding among them, I will keep all four in play, investigating which, if any, experiential combination can satisfy.

The significance of Anti-Combination derives particularly from its seeming to reveal a contrast between the mental and physical realms. The physical properties of composite objects are often thought to be fully explicable by the physical (and topic-neutral) properties of and relations among their parts.

So the best way to investigate the kind of explanation needed for experiential combination is to see what kind of explanation seems to be present in physical combination.

Not all sorts of explanation are relevant here. First, since combination is a synchronic relation, explanation of events in terms of distinct, temporally prior, events is not directly relevant. Second, I wish to abstract from considerations of practical feasibility: combinationists about some domain claim only that macrofacts in that domain are explicable through microfacts ‘in principle’, i.e. for some arbitrarily smart subject under arbitrarily good conditions. Third, we must differentiate the sort of explanation combinationism requires from certain kinds of ‘explanation’ that have usually been regarded as falsifying combination – in particular, explanation by reference to ‘emergence laws’, since (strong) emergentism is often understood as the *denial* that microfacts fully account for macrofacts. We might put this by saying that combinationism involves ‘reductive’ rather than ‘emergentist’ explanations. But both of these terms are ambiguous. It is common to distinguish ‘strong’ from ‘weak’ emergence (e.g. Bedau 1997, Chalmers 2006, Wilson 2010, Seager forthcoming), and ‘reductive’ can also be read in a stronger sense, in which it entails identity and rules out multiple realizability, and in a weaker sense, in which it entails just some kind of intelligible grounding (e.g. Chalmers 2003, Fn.4). ‘Non-reductive physicalism’ is weakly but not strongly reductive in my sense, and while weak emergence can be weakly reductive, strong emergence cannot. My use of the terms ‘reductive’ and ‘emergentist’ will typically mean ‘weakly reductive’ and ‘strong emergentist’, respectively.

A convenient framework for considering competing accounts of explanation is to consider conditionals with purely microscopic antecedents and macroscopic consequents, such that we can explain the holding of the consequent in terms of the holding of the antecedent. Principles of roughly this sort are called ‘principles of composition’ by Broad (1925), and ‘application conditionals’ by Chalmers & Jackson (2001, p.325 ff), Diaz-Leon (2011, p.102 ff), and McQueen (2013, p.97 ff). I will

use Broad's term 'principle of composition', or 'composition principle', defining it as a principle of the form:

If a set of parts $p_1, p_2, p_3 \dots$ possess properties $F_1, F_2, F_3 \dots$ and stand in relations $R_1, R_2, R_3 \dots$ then they compose a composite c which possesses property G .¹

So if combination occurs for some class of properties (e.g. experiential properties), that means that there are composition principles in which G , some of the F s, and possibly some of the R s, are properties and relations of that sort (e.g. experiential properties and relations), and in which consequently c and some of the p s are entities of the sort that bear such properties (e.g. conscious subjects).

However, simply asserting one or more composition principles would amount to stating what is to be explained, and what is supposed to explain it, without guaranteeing the success of that explanation. Even a strong emergentist posits composition principles – they claim that these principles are basic laws of nature. For the same reason, it would be insufficient to simply claim that macrofacts supervene with nomological necessity on microfacts (cf. Horgan 1993, pp.559-560). Reductively explanatory composition principles must be, we might say, 'intelligible': they not only state that certain macrofacts obtain when certain microfacts obtain, but allow us to 'see why'. Some kind of criterion is needed which would separate out properly intelligible composition principles, so my investigation into what 'explanation' requires can be re-framed as an investigation into what criteria a composition principle must meet to count as intelligible, and thereby to underwrite suitable explanations.

Section 1: *A Priori* Entailment

¹ Combinationists are happy with multiple realisability. When there are many different sets of F s and R s that would yield the same G , we could equally well say that there are many different true composition principles, or that the antecedent of the true composition principle is (perhaps infinitely) disjunctive.

My first proposed criterion for suitably explanatory composition principles is that they be knowable *a priori*. On this view, to say that microphysics explains macrophysics is to say that an idealised reasoner could infer all the macrophysical truths from the microphysical truths *a priori*, and thus could know *a priori* the truth of a composition principle connecting them. This is the primary criterion used by modern property dualists to establish the ‘explanatory gap’ between matter and consciousness.² It is implicit in ‘conceivability arguments’: if P explains Q it should entail Q *a priori*, and so the situation where P&¬Q obtains should be inconceivable, and so if that situation is conceivable, P does not entail Q, and hence cannot explain Q.³ It is also widely invoked in discussing ‘explanatory gap’ versions of the combination problem for panpsychism (e.g. Goff 2009a, Seager forthcoming).

Subsection 1.1: A Priori Entailment and Empirical Knowledge

What is knowable *a priori* is the conditional, not the antecedent or the consequent. In particular, if the antecedent is disjunctive (or if there are many different antecedents that all imply the same consequent), it need not be *a priori* which disjunct actually obtains. Composition principles are not supposed to reflect any *a priori* insight into the world’s actual nature, but rather our grasp of the concepts involved in the consequent: this grasp ensures that if we understand the antecedent, and reason ideally, we can apply those concepts to the hypothetical situation described. For instance, Jackson explains the existence of composition principles for ‘solid’ thus:

The story science tells about tables, chairs, pens and the like being aggregations of molecules held in a lattice-like array by various intermolecular forces...tell[s] us that these lattice-like arrays of molecules exclude each other, the intermolecular forces being such as to prevent the lattices

² Chalmers 1996, Chalmers & Jackson 2001, though note that Jackson endorses the criterion but denies dualism; for criticism see Block & Stalnaker 1998, Diaz-Leon 2011.

³ While conceivability and *a priori* are closely connected by many authors, there may be ways to understand *a priori* without reference to conceivability, as suggested by Wilson & Biggs 2013.

encroaching on each other's spaces. *And this is what it takes, according to our concept, to be solid.* (Jackson 1998, pp.3-4, emphasis added)

Horgan says similar things about the liquidity of water:

We understand well enough the essential features, or defining conditions, of liquidity: if a quantity of stuff is liquid, then it will neither spontaneously dissipate into the atmosphere nor retain a rigid shape when unconstrained... explaining why liquidity supervenes on certain microphysical properties is essentially a matter of explaining why any quantity of stuff with these microphysical properties will exhibit those macro-features... *it seems explanatorily kosher to assume a "connecting principle" linking the macro-features to liquidity, precisely because those features are definitive*; the connecting principle expresses a fact about what liquidity is. (Horgan 1993, p.379, emphasis added)⁴

This allows the standard of *a priori* entailment to be applied to the decidedly *a posteriori* practice of scientific explanation. Scientists often 'explain' some observation by postulating a contingent, *a posteriori*, mechanism or law. What is held to be *a priori* is not the existence of this explaining factor, but simply the conditional statement that, *given* this postulate, the observation must follow.⁵

When I say that the antecedent of a composition principle contains only 'micro-level facts', this includes both particular facts, like the existence and location of certain particles, and general facts, like the laws governing the properties and behaviour of those particles. The physical combinationist, for example, can include the laws of gravitational, electrical, and other forces, as well as the laws of motion and the geometry of spacetime, in their 'explanatory base'. However, the antecedent does not include 'bridge laws', which explicitly connect micro-level and macro-level kinds (e.g. "water is

⁴ Note that the 'connecting principle' is not itself a composition principle: it entails one along with a scientific explanation of why the micro-features of H₂O give it the macro-features mentioned in the connecting principle.

⁵ This model also allows for *a posteriori* identities, such as that between water and H₂O: we can explain some fact about water by appealing to some fact about H₂O, together with a description of H₂O that entails that it fits our idea of 'water'. For instance, our idea of water specifies that it is the abundant, colourless, potable liquid that falls from the sky as rain on earth; a full microphysical description of H₂O, including its distribution across planets, entails that it is the only substance that fits the conditions that define 'water', and thus allows us to derive its identity with water. Yet the identity is still ultimately *a posteriori*, because it is *a posteriori* that H₂O has the properties definitive of water.

H₂O”), since these explicitly mention macro-level phenomena (e.g. “water”). This is important to note, because the ‘deductive-nomological’ model of explanation (Hempel and Oppenheim 1948, Hempel 1965) purports to explain particular macro-level facts by deducing them from the conjunction of such bridge laws with properly micro-level laws and conditions. However, the two models of explanation can be made equivalent, if the purely micro-level facts allow us to deduce the bridge laws themselves, (as when the bridge law “water is H₂O” follows logically from the microphysical fact that H₂O molecules are what ‘plays the water role’). Of course, bridge laws cannot be deduced from purely *positive* micro-level information. Any set of facts about the properties and distribution of H₂O (or any other substances) leaves open the possibility that there might be some *other* substance that fits the specifications involved in our notion ‘water’ better than H₂O does. Similarly, any list of laws of nature will fail to entail anything about macroscopic forces, if they do not rule out the existence of other laws of nature. This leads some authors (e.g. Chalmers & Jackson 2001) to include a ‘that’s all’ clause, stating that there is nothing *else* in the world than what is stated in the micro-level facts, and what is constituted by them.⁶ By allowing for the derivation of bridge laws, the ‘that’s all’ clause permits a convergence of the *a priori* entailment model with the classical deductive-nomological model (cf. Carruthers 2004).

Subsection 1.2: The Case of Mass Additivity

A particularly extensive defence of an *a priori* entailment analysis of physical combinationism is given by McQueen, who defends analyses of this sort for basic composition principles involving mass, charge, conductivity, and quantum phenomena. His most fully explored example is:

⁶ Does the ‘that’s all’ clause beg the question in favour of combinationism, by assuming from the outset that nothing exists but the microfacts and what they constitute? No, because combinationists are not claiming that the clause is true, but that from its supposed truth, all actual macro-level truths can be fully deduced.

Mass Additivity: The mass of a composite entity is equal to the sum of the masses of its parts. Now, mass additivity itself is not *a priori* – indeed, mass additivity is empirically false, at least for rest mass, in Einsteinian physics.⁷ Rather, McQueen’s claim is that mass additivity (or the similar principle of ‘relativistic mass additivity’) *follows a priori* from the physical laws operative in classical physics (or in Einsteinian physics). Note two particular features of this derivation (see McQueen 2014a for the full discussion).

First, McQueen’s derivation (modelled after derivations often given in physics textbooks) assumes, as one of the *microphysical* facts, the principle that the total force on an entity is the sum of all forces exerted on it by other entities, even though some writers (e.g. Mill 1843, p.458 ff) regard this as itself something like a composition principle, involving the ‘composition of causes’. Insofar as we think of the vector-additivity of forces as itself a composition principle, it seems hard to regard it as *a priori*. But this principle meets plausible requirements for being considered a micro-level law: it “come[s] into play at or below the atomic level of organization” (Wilson 2002, p.74), among microscopic entities. It makes no mention of any composite thing: it is simply that multiple forces are exerted on one (perhaps simple) entity, by multiple other (perhaps simple) entities. Thus there is some plausibility in treating it, as McQueen does, as an *a posteriori*, microphysical, law of nature.

Second, McQueen explicitly presupposes that it is *a priori*, for some objects, that they compose a whole which is located wherever they are. This is what lets him say that the mass of the whole, which is to be explained, can be identified with whatever property is appropriately mathematically related to the average acceleration of the parts (which, he shows, is the sum of the masses of the parts). If *a priori* deducibility is an appropriate criterion in the physical case, we must be allowed to draw on general metaphysical doctrines about the nature of the part-whole relation, and treat them as holding *a priori*.

⁷ More precisely, it fails for ‘rest mass’, as opposed to ‘relativistic mass’, due to the ‘binding energy’ required to form a complex out of smaller particles; given the equivalence of mass and energy, this alters the rest mass of the whole.

Section 2: Structural Simplicity

We might take the British emergentists' classical distinction between emergent and mechanistic composition principles to rely on on *a priori* entailment, since Broad, for instances, describes emergence as occurring when:

The characteristic behaviour of the whole *could not, even in theory, be deduced from the most complete knowledge* of the behaviour of its components, taken separately or in other combinations, and of their proportions and arrangements in this whole. (1925, pp.59, emphasis added)

That is, we might define emergent phenomena as 'not predictable' or 'not deducible' from their emergence bases, even in principle (Broad 1925, p.61, McLaughlin 1992, Kim 1999 p.5ff, Wilson 2010, pp.32-38), and conclude that since strong emergence is usually taken to lack the sort of explanation we are interested in, it makes sense to focus on *a priori* entailment as what is missing in such cases. The distinguishing feature of the emergentist's composition principles, on this analysis, is that they cannot be derived from anything else: they are brute, fundamental, laws of nature that can only be discovered by empirical investigation.

However, Broad explicitly denies that any composition principles, even mechanistic (aka 'reductive') ones, are knowable *a priori*:

...in no case could the behaviour of a whole... be predicted *merely* from a knowledge of the properties of these constituents... and of their proportions and arrangements... Whenever this seems to be possible it is because we are using a suppressed premise which is so familiar that it has escaped our notice. The suppressed premise is the fact that we have examined other complexes in the past and have noted their behaviour; that we have found *a general law connecting the behaviour of these wholes with that which their constituents would show in isolation*; and that we

are assuming that this law of composition will hold also of the particular complex whole at present under consideration. (1925, p.63)

So while Broad does deny that emergent properties are deducible *a priori*, even with perfect knowledge of the microphysical base, this is true of *all* properties of wholes. What distinguishes emergent composition principles from non-emergent ones is not their epistemology, but their structure of application: they are specific only to a particular collection of components in particular relations, rather than providing a general rule for deriving the properties of a wide variety of composites.

This gives us a second criterion for explanatory composition principles: their pattern of application. As Broad puts it, the non-emergent, or ‘reductive’, or ‘mechanical’, composition principles are those in which the features of the whole are “compounded in a simple and uniform way” from those of their parts (p.44), providing a “general law” that is not restricted to any specific configuration. The task for the experiential combinationist, then, is to find similarly simple, general, and uniform experiential composition principles. In contrast to section 1’s approach, these principles might be *a posteriori* laws of nature, holding in the actual world but not in all other possible worlds. Thus combinationism might be true in the actual world, but not necessarily true: its truth does not require the necessity or *a priority* of its composition principles, but merely their generality.⁸ Presumably, we would need empirical evidence that we live in a world where combinationism is true, but this might be supplied by the two observations that i) we are conscious, and ii) we are to all appearances composite things, fully material and to that extent explicable through our parts. Thus we may distinguish two approaches to experiential combinationism, distinguished by their view on what is required for combination in general: *a priori* combinationism and *a posteriori* combinationism.

⁸ *A priori* entailments still feature in this model: given all the laws of nature, including cross-level ones (*a posteriori* composition principles), along with all the micro-level facts, the macro-level truths follow. But the composition principles themselves are not *a priori*: they must be included in the antecedent of any *a priori* conditional.

Subsection 2.1: Simplicity in the Set of Principles and in Individual Principles

The idea of a structural criterion of intelligibility could be developed in two directions, focusing either on the overall set of principles, or the internal structure of particular ones. On the first approach, we would think of a set of composition principles as properly explanatory when they are few in number and general in application, each entailing a broad range of macrofacts; ‘emergent’ composition principles are manifest only when certain specific conditions obtain, so that we make little explanatory progress in our attempts to unify phenomena. The sense in which emergent composition principles are less ‘explanatory’, on this account, is that they do less well at accounting for many observations by progressively fewer, or progressively better-understood, mechanisms, so as to minimise the number of distinct unexplained phenomena (the fundamental laws and initial conditions), and weave observations together into a single connected web by their reflecting the same underlying principles (cf. Friedman 1974, Hempel 1965).⁹

Rather than looking at the overall set of laws, we might alternatively look at the internal structure of individual composition principles, as suggested by the recurrent talk of their being ‘linear’ (alternatively, “additive and subtractive only”, Morgan 1923, p.3, or “broadly additive”, Wilson 2013-a, p.201). But while such talk makes it clear that a formula involving simple addition ($a=b+c$) meets the criterion, it does not make clear how to apply the criterion beyond that. Vector addition and scalar addition are already somewhat different mathematically, and the addition of other mathematical forms like squares, logarithms, or integrals hardly seems a mark of emergence. The problem intensifies when we consider phenomena that are not readily quantified, like qualitative consciousness; moreover

⁹ This goes beyond the much more modest claim often expressed as ‘Occam’s Razor’, that simpler explanatory schemes are preferable *other things being equal*. We are considering the proposal that simplicity is not only a virtue, but is the yardstick by which to measure whether one thing can properly be said to explain another.

the role of non-linear equations in modelling chaotic systems, generally taken as straightforwardly non-emergent, threatens to undermine the criterion in any form (as argued by Wilson, 2013-a). Thus it seems that the best way to capture the appeal of the ‘linear combinations’ criterion, understood as constraining the internal structure of a composition principle, may be to interpret it as simply requiring combinationists to find, in Broad’s words, some “simple and uniform way” (1925, p.45) of compounding the experiential properties of a thing’s parts, mathematically or otherwise, to yield those of the whole.

Section 3: Continuity of Natures

A third approach holds that a composition principle is intelligible when the concepts used in the antecedent and consequent are in some sense akin to each other: this imposes *a priori* constraints without requiring *a priori* deducibility. For example, it might be intelligible for one set of movements in space to explain another set, but not for a set of non-spatiotemporal facts to explain an occurrence in space and time. More specifically, we might require that explained and explaining properties be determinates of the same basic determinables, related in the way that different shades of red are related to each other through the determinables ‘red’ and, ultimately, ‘coloured’.

Strawson (2006, p.13ff) uses this standard to argue that intra-mental explanations are more intelligible than the arising of experience from a non-experiential basis, for only in that case do we “move wholly within a completely conceptually homogeneous (non-heterogeneous) set of notions” (p.15). Mørch (2014) elaborates this as requiring that intelligible explanations involve different *forms* inhering in the same *matter*, the same ‘basic stuff’. As James, whom she quotes, says: “all the new forms of being that make their appearance are really nothing more than results of the redistribution of the original and unchanging materials.”(1890, p.147). Mørch argues that this requirement lies behind

the demand for *continuity* in our explanations of consciousness, voiced by authors such as James and Clifford: that each change must be relevantly gradual, rather than a discontinuous “irruption... of a new nature” (James 1890, p.148).¹⁰

For instance, the explanation of macroscopic liquidity by appeal to microscopic physical and chemical facts about bonding and charges may, on this approach, involve *a posteriori* composition principles, but what makes it intelligible is that both what explains and what is explained deal essentially with how things in space move in space in response to the movements of other things in space. The task for the experiential combinationist would then be to provide composition principles that are similarly homogeneous with regard to basic determinables.

Subsection 3.1: The Problem of the Bonding Relations

Continuity of natures may seem an easy standard for the experiential combinationist to meet. Microexperiential facts are clearly akin to macroexperiential facts, both being experiential. However, there is a way to strengthen the continuity of natures requirement that makes it much more challenging for combinationists to meet, by extending it to the relations involved in the antecedents of their composition principles, which we may call ‘bonding relations’. We might require the natures of these bonding relations themselves to be continuous with the properties they bond; as Coleman puts it:

Such relations [should] visibly *flow from* the intrinsic natures of the *relata*... [just as] it is because of the relative looseness of hydrogen electrons, coupled with the convenient gap in the oxygen’s outer shell, that electron sharing happens so readily in the constitution of water. For phenomenal bonding to work, we would need some analogue of this... taking into consideration

¹⁰ Mørch also takes the idea of preserving the same ‘matter’ to require quantitative conservation of some underlying quantity. However, as far as I can tell this criterion makes sense primarily in relation to the particular sort of ‘causal combination’ that she discusses, in which a group of simpler consciousness at one time are entirely replaced by a simple consciousness at the next moment, the explanatory relation between them being diachronic.

the intrinsic features of microsubjects [should] sugges[t] to us the mechanism for their phenomenally bonding. (forthcoming, p.18)

So the experiential combinationist should look for a relation between distinct subjects that ‘flows from’ the intrinsic features of those subjects, so as to be conceptually akin to the experiential properties of the whole thereby created. Yet it is hard to see what relation satisfies this description: the apparent privacy of experience makes it hard to see relations among distinct subjects as essentially experiential.

Of course, we can trivially define an inter-subject relation intelligibly connected with the experiential properties of the whole those subjects compose, by speaking of relations such as ‘jointly composing a conscious subject’, or ‘jointly giving rise to phenomenal unity’. But this fails to connect them intelligibly with the properties of the components: we still need some grasp of what the relations in question *are*, beyond being simply ‘the relations which give rise to such-and-such an experiential composite’. Call this the *independent grasp* requirement. Physical combination easily meets this requirement, for we can understand spatial relations, causal relations, energy-transfer relations, and so on at least as well as we understand any of the complexes they enter into. For instance, we can understand a whole’s shape in terms of distances between parts, and these distance relations meet the independent grasp requirement; relations such as ‘jointly composing a whole with such-and-such a shape’ do not.

So while the experiential combinationist can easily satisfy the ‘continuity of natures’ criterion when it comes to the intrinsic properties of parts and wholes, they need to also satisfy it regarding their bonding relations by providing some micro-level account of what these relations are. It has been argued that this cannot be done:

Human beings are able to have neither introspective nor perceptive experience of relations between subjects of experiences qua subjects of experience. We are unable to perceive [them] through the senses simply because we are unable to perceive subjects of experience (qua

subjects of experience) through the senses... [and since] we can introspect only one subject of experience [ourselves]... we cannot introspect how subjects of experience qua subjects of experience are related, for [then] we would have to be able to introspect more than one subject of experience. (Goff forthcoming, pp.9-10)¹¹

So we cannot grasp experiential relations through external perception, for that does not reveal anything distinctively experiential, or through introspection, for that reveals only one subject, and thus cannot reveal the links between multiple subjects: thus we cannot grasp experiential relations at all. Consequently Goff concludes that while there may be such relations, and while they may play the metaphysical role that constitutive panpsychists (and other combinationists) claim, they are of no explanatory use; the link between experiential parts and wholes remains mysterious.

There are at least three ways for combinationists to address this challenge: they might claim that the experiential bonding relations are graspable via external perception, or via introspection upon only one of the relata, or via introspection upon both relata. The first of these options is probably open only to the physicalist combinationist, who might hold that experiential properties, being ultimately physical, can be directly observed in others. They might claim that neural synchrony (cf. Crick & Koch 1990), for instance, or information integration (cf. Tononi 2012) is the experiential bonding relation, and that we learn of it by observing brain structure. But primitivists about consciousness, such as the panpsychists at whom Goff's and Coleman's arguments are directed, will find this unsatisfying.

The second approach is to claim that we can grasp the experiential bonding relations through standing in them. For instance, perhaps in some social interactions, the distinctive way in which our experiences represent those of our conversational partners or co-operators allows those experiences to constitute a conscious group mind. Or perhaps some constant background feature of our conscious

¹¹ Compare Coleman forthcoming, p.18: "it is because panpsychists *cannot* see how subjects could come together, given their intrinsic properties, that the supplement of phenomenal bonding is broached. We have here a relation devised precisely to supply the obvious defects of its putative *relata*."

‘being in the world’ constitutes the cosmic composite mind. However worked out, this implies that we can grasp the nature of a relation through being *one* of the relata, contra Goff’s implicit claim that we grasp a relation only if we are acquainted somehow with *both* relata. Moreover, it implies that we sometimes stand in the experiential bonding relation to something else: we actually stand to external things in some of the relations which bond our component subjects into a composite subject. This virtually guarantees that combinationists must countenance what in chapter 1 I called ‘mega-subjects’, with all the attendant difficulties. Call this the ‘outward-looking’ approach to grasping the bonding relations.

A final way to satisfy the independent grasp requirement is to claim that we are introspectively acquainted with the relations among our component subjects, because they are somehow present to us in our own experience. In particular, combinationists might point out that we do seem to be introspectively acquainted with relations among *experiences*, and that these might constitute relations among subjects that meet Coleman’s requirement that bonding relations ‘grow out of’ the natures of their relata. This helps only if our experiences belong not only to us but also to our parts. That is, combinationists might seek to meet the independent grasp requirement by endorsing the sharing of token experiences: the experiential relations among our parts are mediated by relations among their experiences, which are then introspectively revealed to us as internal relations within our own experience.

Call this the ‘inward-looking’ approach: it aims to identify the ways that elements of experience are related to each other, and abstract those relations from the particular configurations of them that we find, hoping to extrapolate to their holding between distinct component subjects and thereby satisfy the independent grasp requirement. This commits combinationists both to regarding the

bonding relations among subjects as mediated by those subjects' experiences, and to regarding experiences as shareable. This latter claim is a significant commitment, as the next section discusses.

Section 4: Sharing, Inheritance, and Connections

A final account claims that in genuinely intelligible combination, the properties of the whole are not really distinct from the properties of its parts: they are either the very same token properties, shared by whole and part, or are reducible to some set of properties that are thus shared. We can make this idea more precise by distinguishing three sorts of principles (with two alternative variants of the second sort):

Token-Sharing of x : Particular tokens of property x can belong simultaneously to two different entities.

(Conditional) Inheritance of x : A whole has property x whenever one of its parts does (and when, moreover, that part is appropriately related to its other parts), simply in virtue of the part having x (and being appropriately related to the other parts).

x - y Connections: Something can have property y simply by having properties $x_1, x_2, \text{etc...}$

Call these three sorts of principles 'token-sharing principles', 'inheritance principles', and 'connection principles'; they are logically independent, yet fit naturally together. If all three hold for some properties, then a whole may have those properties and yet there be no properties involved that are not, ultimately, grounded intelligibly in the parts. Yet these principles are far from trivial: they commit combinationists to some strong metaphysical claims. It would be implausible to demand that the experiential combinationist must fit experiential combination into this schema, if there were no examples of its holding in the physical realm. However, there are some apparent examples, as I will show.

Subsection 4.1: Physical Cases of Token-Sharing and Conditional Inheritance

It is intuitively plausible that token physical features are often shared between parts and wholes. For instance, an object with an uneven surface seems to share that instance of unevenness with its surface (and all sections of it that include that surface). A red surface seems to share, with each of its red subsections, their particular instances of redness. A car may be dented when its roof is dented, and this seems to involve only one instance of the property ‘being dented’: similarly for being perforated, or wounded. If I can be said to be ‘metabolising alcohol’ when my liver is, it seems wrong to think of this as two instances of that activity.¹² In all of these cases it seems natural to say that there is a single instance of the property in question, which can be truly ascribed both the whole and the part.¹³ These cases are at least intuitively different from cases where there are two separate and independent instances of the same property (two red things side-by-side, two distinct dents or wounds, etc.).

Obviously the interpretation of these cases will depend on one’s metaphysics of properties, for token-sharing is meaningful only if there are such things as particular instances of properties. If our basic ontology admitted only objects and universals, we might have to construct the idea of an instance, for instance by saying that there is an instance of a property for every triple of an object, a property, and a time such that that object instantiates that property at that time. Depending on how we construct this idea, we might find that the sharing of instances between objects is simply ruled out. However, I find the above examples compelling enough, and widespread enough, to consider it a requirement on any theory of properties that it be able to do justice to the distinctive form of ontological dependence involved. Even if we individuate instances in such a way that the dented car has a distinct instance of

¹² On the other hand, it does not in general seem that two discrete substances can simultaneously share a property instance. If we try to imagine, for instance, a single instance of redness belonging to two discrete surfaces, it seems that all we can imagine is a token of redness some parts of which belong to the one surface, and other parts of which belong to the other.

¹³ Hellie raises the following objection: the instances must be distinct, by Leibniz’s Law, for there will always be properties that one has and the other lacks. For instance, with the liver example, it seems my instance will possess the property of being an instance of metabolising alcohol *by its bearer’s liver*, while the liver’s instance will not (on the assumption that livers do not have themselves as livers). But this begs the question, for if there really is a sharing of instances, then the phrase ‘its bearer’ will fail to refer.

dentedness from each of its dented parts, we should still recognise that the dentedness of the car is nothing over and above that of its roof. The current proposal is that this sort of ‘nothing over and aboveness’ is part of what makes physical combination intelligible, and that the experiential combinationist should seek to employ it in their own theory. This proposal does not depend on how we characterise this ‘nothing over and aboveness’, whether in terms of a single shared instance, or as a *sui generis* sort of compositional grounding, or in some other way.¹⁴

Many of the above examples are also compelling cases of ‘conditional inheritance’, with the whole ‘inheriting’ the relevant properties from its parts, having them simply as a consequence of containing a part which has them. Yet not all wholes would inherit these properties: many mereological fusions of which I and my liver are parts (e.g. ‘the fusion of me and the CN tower’) should not be said to ‘metabolise alcohol’, for they lack a metabolism. The property is inherited by the whole only if the part instantiating it is integrated with the other parts of that whole so as to give the whole the appropriate sort of structure. Consider an example from Kriegel:

By punching Johnny's nose, Jimmy punches Johnny, but does not punch the galaxy, because Johnny's nose is integrated into Johnny in a way it is not into the galaxy: the interconnections between Johnny's nose and the rest of Johnny are very tight relative to the interconnections between Johnny's nose and the rest of the galaxy. (2009, p.227)

Here the property ‘being punched’ is conditionally inherited: the whole (Jonny) having these properties is both intelligibly explained by, and metaphysically grounded in, the possession of those properties by the part (Jonny’s nose). Yet these properties are not inherited by the whole galaxy, because the relations among the nose and the rest of the galaxy are not suitable.

¹⁴ Perhaps the grounding relation involved in these cases could be used to permit even those who individuate instances by their bearers to accept talk of shared instances, through the following definition of a ‘stack instance’: a set of unshared property instances which stand to each other in the sort of grounding relation that examples of apparently ‘shared’ instances do. Readers who consider shared property tokens definitionally impossible could read my talk of them as really speak of ‘stack instances’ whose members belong to different objects.

Subsection 4.2: Unconditional Inheritance - Location Properties

Are there examples of unconditional inheritance of physical properties? Several authors have endorsed the following principle:

Location Inheritance: A whole 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.

This principle is affirmed by Van Inwagen (1990, p.44) Lewis (1991, p.85), Sider (2004, p.52 and throughout), Bennett (forthcoming, p.10), and McQueen (2014a), with Sider and Bennett both expressing it as ‘inheritance’ of location properties. These authors take this principle to be obviously true: Van Inwagen calls it “self-evident.”(p.54), Sider says that “everyone accepts” it (p.75), and Bennett calls it “fairly uncontroversial and not really in need of argument... [something] that can legitimately be taken as data [and is] hard to deny.” (p.11)¹⁵ We can attribute this plausibility in part to the plausibility of token-sharing for location properties: if both part and whole occupy a particular location, that location is not occupied twice over, but only once; the same instance of occupying it belongs to two entities.

Moreover, these authors do not regard location inheritance as expressing an idiosyncratic fact about space – rather, they accept it as just one instance of a more general mereological phenomenon, though their other examples diverge somewhat. Sider pairs “Inheritance of location” with “Inheritance of intrinsicity: If property P is intrinsic, then the property *having a part that has P* is also intrinsic”

¹⁵ My use of the phrase ‘inheritance principles’ is based primarily on this literature: other examples are Bennett (2011, p.12ff), who introduces the principle of ‘slot inheritance’ for her distinctive notion of a ‘parthood-slot’, and Helle (2013, p.54ff), who uses the term for any principle of the form “if a part *x* of some whole *y*, is F, then *y* is F”, arguing that the Stoics endorsed inheritance principles for rationality and sentience apparently held by the Stoics. Kim (1998, p.54) and Philips (2014, p.139 ff), speak of ‘inheritance’ in metaphysical but non-mereological senses, between realising and realised properties or between the stream of consciousness and its contents.

(p.70), while Van Inwagen accompanies it with principles governing mass (a whole's mass is the sum of its parts') and surface area (a whole's surface area is equal to or less than the sum of its parts') (p.44). Lewis speaks more generally of "the ease of describing fusions"(p.85): if you "describe the character of the parts [and] describe their interrelation... you have ipso facto described the fusion" (p.81).¹⁶

Of course, location inheritance is defensible only if its meaning is properly qualified. There is an obvious sense in which I am located in a different (larger) area than, say, my foot. Sider (2007, Fn4, p.52) replies that there is an ambiguity in 'located': it can mean 'wholly located', and this is not inherited, but this sense can be defined in terms of another, more non-committal sense: to be 'wholly located' in some region is to be 'located', in this basic and non-committal sense, at all and only the points in that region. And location in this sense is inherited. Here we see a pattern that will be repeated: something can be located somewhere in a sense which is sensitive to its total set of location properties, and in a sense which says nothing about where *else* it is located.¹⁷ Call properties of the first sort 'systemic', and properties of the second sort 'additive'. Systemic properties are sensitive to the overall set of properties belonging to their bearer; additive ones are not.

It might be objected that many composite objects occupy large regions of space even though their parts occupy only point-sized parts within this region. This illustrates a different ambiguity in 'located'. To be located is to occupy some amount of space, and we might analyse 'occupying' as meaning (roughly) exerting forces that exclude other, discrete, things from that space. Then something could be located in a lot of the 'empty space' between its simple parts. But in that sense we should not

¹⁶ Bennett glosses this as a supervenience claim: "each x's properties supervene on the properties of its parts" (p.11). But this is arguably too weak: as argued in Schiffer 1987, Horgan 1993, Wilson 2005, and elsewhere, supervenience claims are generally not enough to capture substantive and interesting metaphysical theses.

¹⁷ 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".

say that the parts are located only at points, since they exert forces much more widely than that. In the sense in which they occupy only points, the whole does as well, according to the defender of location inheritance. We simply need to keep our notion of ‘location’ univocal.

One striking thing about location inheritance is how readily it suggests a set of ‘connection principles’ to complete the combinationist account of geometrical properties. Obviously many geometric properties are not inherited: something made of circular parts need not be circular, nor must something made of small parts be small. But location properties, which plausibly *are* inherited, provide a basis for determining other geometrical properties: by *a priori* principles, we can derive a thing’s shape, size, orientation, etc. from knowing which points or regions of space it is located at. For instance, if it occupies approximately those points whose distance from a single specific point is less than a certain value, then it is spherical. Different geometrical connection principles will place more or less importance on the relations among the inherited properties, and in the limiting case may be entirely insensitive to these: a thing’s volume depends merely on how much space it occupies, regardless of how that space is arranged, while its surface area depends crucially on how this space is distributed.¹⁸ But in both cases, fully understanding the whole’s property (total volume or total surface area) reveals how, given the other properties of that very whole, it cannot fail to have that property.

Subsection 4.3: Unconditional Inheritance - Causal Powers

Another candidate instance of the sharing-inheritance-connections schema would involve causal powers. We would first endorse:

¹⁸ In an infinitely divisible (‘gunky’) world, given the Axiom of Choice, the Banach-Tarski Theorem may falsify common-sense claims about a whole’s volume being the sum of its parts’ volumes: any whole is divisible into many parts with volume equal to the whole’s. Here I am content to simply observe that infinity makes *everything* weird.

Power Sharing: Particular token causal powers can belong simultaneously to two different entities.

Power Inheritance: A whole has a given causal power whenever any of its parts do, simply in virtue of the part having that power.

These could be glossed as saying that what a whole does is just all the things which its parts do.¹⁹

Connection principles would then allow us to deduce various other physical properties, such as brittleness, mass, density, and so on, from the whole's set of causal powers.²⁰ Many connection principles will be analytic, simply definitions of intuitively-understood properties; we might look for examples to the explanations of solidity and liquidity given by Jackson (1998, pp.3-4) and Horgan (1993, p.379), already quoted in section 1.

Several authors say things which seem to suggest acceptance of power sharing and power inheritance, at least for non-emergent physical cases. Merricks says that when a baseball shatters a window, “every atom arranged baseballwise causes something, and when what one of them causes is added to what each of the others causes, the ‘sum’ is the shattering of the window.”(2001, p.111)²¹ Mill's seminal discussion of ‘homopathic’ and ‘heteropathic’ composition of causes also suggests

¹⁹ 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”).

Even then, however, there is the problem that parts of X generally have powers that would be exercised in conditions where X does not exist, or does not contain those parts. This might be a good reason to prefer only a *conditional* inheritance of powers, or alternatively to restrict the unconditional inheritance to only those powers that are actually exercised on a given occasion. Alternatively, perhaps unconditional power inheritance could be endorsed as-is by someone who accepted *mereological essentialism*, possibly in perdurantist form.

²⁰ In examples like ‘brittleness’ (also solidity, liquidity, hardness, etc.) the relevant causal powers are those exerted on other parts of the whole; the inheritance principle then implies that the whole exerts a causal influence on its parts (in such a way as to be brittle, solid, liquid, etc.). This phrasing might seem to imply a kind of ‘downwards causation’: the whole exerts forces on its parts. But it exerts only those forces that its parts exert on each other; to quote Smart, “this is a sort of emergence that the most reductionist and mechanist physicalist will never have dreamed of denying.” (1981, p.111)

²¹ Merricks in fact denies that the baseball causes whatever its parts cause, because he denies the baseball exists. But he clearly thinks that *if* there were such physical composites, they would have the same causal powers as their parts; indeed this is crucial to his argument for their elimination.

something like this. In the homopathic (non-emergent) case, the combined effect is simply the two component effects, both existing and composing the joint effect just by existing. That is, in such cases “one cause never, properly speaking, defeats or frustrates another; both have their full effect” (1843, p.458). With heteropathic combinations, by contrast, the normal effects are absent, being replaced according to a new and different law.

There is a complication here in the fact that Mill is not talking of persisting composite objects like molecules, but only about ‘composite causes’. We could translate this into talk about the causal powers of composite things, by supposing that when there is an effect of two or more causes, there must be a single thing (‘the cause’) which may be attributed the causal powers of all the things producing effects, which can then be called its parts. But such a conversion goes beyond what Mill says, and is not uncontroversial; it seems to presuppose or imply unrestricted composition, insofar as it requires a whole containing any two things which both have effects on the same thing. If we did translate Mill’s distinction in this fashion, we might say that a whole’s properties are transparently explained by those of its parts when its causal powers are simply the combined causal powers of all of its parts (perhaps restricted to those which are actually exercised, or could be exercised compatibly with that part still composing that whole), which is just what power sharing and power inheritance claim.

Others might endorse only a *conditional* form of power inheritance. Some authors have defended the view that particular instances of mental states are nothing over and above the physical states that realise them, because they possess a proper subset of the realising state’s causal powers (Wilson 1999, pp.45-51, Shoemaker 2000, p.28). The realising state in turn may be regarded as the collection of physical states of the relevant brain parts (see Gillett 2002, p.319, Wilson 2013-a, Fn.3). Indeed, some advocate extending this to all special science entities, whose causal autonomy lies precisely in their lacking the fine-grained powers of their specific underlying basis (Wilson

forthcoming), yet whose ontological dependence on the physical lies in their sharing a subset of those causal powers.

Like location inheritance, power inheritance naturally suggests accompanying connection principles. In particular, any functionalisable property - any property which can be analysed as the playing of a certain ‘causal role’ - can be analysed as a certain set of causal powers. The earlier contrast between ‘additive’ and ‘systemic’ properties recurs here: a causal profile (the property of playing a certain causal role) is a ‘systemic’ property, dependent on a thing’s total set of properties, and might be lost by the gaining of new powers. By contrast, causal powers (like any inherited property) must be understood in an ‘additive’ sense, as having no implications about the other properties of their bearer.

A much more basic connection principle might be that exerting two or more forces on the same object is no different from exerting a resultant force equal to their vector sum on that object. Together with power inheritance, this implies that a composite exerts on any object the sum of the forces which its parts exert on that object: in McQueen’s derivation of mass additivity from Newtonian physics, he explicitly relies on this principle (2014a, p.13), though treating it as a microphysical law.

This connection principle could be rejected: perhaps we must distinguish exerting a single total force on something from exerting many smaller forces on it. For one thing, as has been noted repeatedly (Russell 1903, p. 477, Wilson 2009, p.541), component forces are not ‘parts’ of a resultant force in the usual sense of ‘part’, since a single force can be seen as having any number of sets of components – a diagonal motion as two perpendicular motions, or a greater motion in the same direction and a lesser motion in the opposite direction, and so on. So the idea of the component effects as still existing ‘in’ the joint effect requires elaborating the relevant notion of containment.²² However,

²² More radically, we might deny the existence of component forces altogether, holding them to be epistemologically inaccessible or ontologically mysterious (Cartwright 1980, 1983), or to threaten causal overdetermination (Wilson 2009). Yet component forces are commonly appealed to in scientific explanations, and arguments against their reality are generally taken to have radically revisionary implications, such as the laws of nature being strictly false (Cartwright

a number of accounts have been offered that vindicate the presence of component forces in combined circumstances, construing them in terms of their ‘contributions to an effect’ (Molnar 2003) or ‘tendencies’ (Bhaskar 1978) as opposed to effects proper.

Subsection 4.4: Relations among these Principles

So there are at least two cases - causal properties and spatial properties - where a physical combinationist might endorse unconditional inheritance. It is worth asking whether conditional inheritance can be analysed in terms of unconditional inheritance, specifically unconditional inheritance of the additive components of a property whose proper analysis shows it to be a conjunction of additive and systemic components.

Recall Kriegel’s example, ‘being punched’. Someone might maintain that there *are* some properties that the galaxy instantiates simply in virtue of someone punching Jonny’s nose: the galaxy’s basic physical features (mass, charge, etc.) come to be differently distributed by the compression, changing its causal powers in various miniscule ways.²³ These physical properties, if formulated in an appropriately additive way, might be claimed to be unconditionally inherited. And if this does not suffice for us to say that the galaxy has been punched, the reason might be that to be punched is not only for punching-events to take place in you, but for them to play some sort of overall systemic role in you. For instance, perhaps they need to be in some sense prominent among the things taking place in you: the punching events involve both Jonny and the galaxy, but involve Jonny *more heavily*. The small changes in the nose are likely to result in much greater downstream changes in the other parts of

1980, 9183) since in most circumstances they only specify component forces. Moreover, the component/resultant distinction may be itself relative to a reference frame (A. Wilson 2009). Thus the experiential combinationist may take the default view that there are such things as component forces, whatever they are.

²³ Plausibly, ‘being punched’ essentially involved certain historical second-order properties, i.e. ‘having undergone certain changes *as a result of* a certain kind of action performed by a certain kind of animal using a certain kind of organ (a blow from a human fist).’ Insofar as the changes are inherited, so too are these historical properties.

Jonny, on average, than in other parts of the galaxy, on average (even though all of the former changes are among the latter). The punching-events are not prominent in the overall goings on in the galaxy, since most of the galaxy is unaffected, whereas they are prominent in the overall goings on in Jonny. Or maybe for you to be punched, the punching-events need to impact your overall physical or biological integrity and self-maintenance, either positively or negatively. Or maybe the overall role is something else: the point is that any such role will be non-heritable simply because it is systemic, not additive.

Perhaps all cases of conditional inheritance admit of a similar analysis: the properties in question resolve into an unconditionally inherited ‘core’, which is additive in that its possession does not depend on a thing’s total set of properties, and a non-inherited systemic ‘role’ for that core to play. A composite ‘conditionally inherits’ the full property in that it instantiates it only when the inherited core plays the right overall role. Alternatively, perhaps conditional inheritance is just as basic as, or more basic than, unconditional inheritance. This question will recur in chapter 3.

There can be further variation among inheritance principles, beyond conditional and unconditional. What I have discussed so far, and will generally have in mind, is ‘upward inheritance’: inheritance by wholes from parts. We could also evaluate ‘downward inheritance’, where a part has a property because the whole has it. ‘Both-ways inheritance’ conjoins downward and upward inheritance into a biconditional, as would be particularly appropriate for supporters of composition-as-identity, who take wholes and parts to be symmetrically related.

Note that given an inheritance principle for x and an x - y connection principle, we can derive a composition principle for y : a principle saying that when certain things exist with certain properties and certain relations, a whole must exist and have certain other properties.²⁴ The inheritance-connections

²⁴ Strictly, inheritance and connection principles do not cover the mere existence of a composite entity: they require some answer to what Van Inwagen (1990) calls ‘the special composition question’ (SCQ): under what conditions do some things compose something? I set this topic aside because most commonly defended answers to this question are not specific to physical or mental composites, but apply generally across domains.

schema simply adds a ‘middle-man’ that the composition principle does not, in the form of the properties the whole inherits from its parts, and which themselves determine its possession of the properties its parts lack. Thus an explanation fitting this schema says more than one merely fitting the composition principle.²⁵

Subsection 4.5: The Schema Applied to Experiential Combination

For experiential combination to fit the sharing-inheritance-connections schema, we would have to suppose that token experiential properties can in some cases be shared by distinct subjects, that at least some are conditionally or unconditionally inherited, and that those which a whole possesses but its parts lack can be analysed as some set of more basic experiential properties connected together.

This proposal faces serious challenges, most obviously over the apparent conflict between the requirement that experiences be shared by part and whole, and the idea that each subject’s experiences are ‘private’, belonging essentially to that subject and never also to another subject. In physical combination, the difficult questions are generally about *properties*: can we explain properties like ‘being water’, ‘being alive’, ‘being liquid’, in terms of other, simpler, properties? There is little concern that the composite entities bearing those properties are themselves specially inexplicable, by nature cut off from the other entities which are their parts. But conscious subjects have been claimed by some to be ‘metaphysically insulated’ from each other by their basic nature: Coleman (2012) calls them “irrevocably separate”, “inviolable individuals” (p.146); James (1890) says that the breaches between

²⁵ We could construct explanations that formally fit this schema but are not in this way stronger than the composition principles they entail. We would simply make the inheritance principles tautologous by applying them to properties of the form “has a part which...”. Trivially, something whose part has a part of some particular sort, has a part of that sort, since parthood is transitive. The interesting work would then be done by the corresponding connection principle, which says that something with one set of properties (namely properties of having parts of certain sorts) must have certain other properties; such connection principles would be equivalent to composition principles.

subjects are “the most absolute breaches in nature.” (p.226) This seems to rule out token-sharing, which in turn makes inheritance hard to maintain.

The schema potentially also threatens to ‘overgenerate’, in at least two ways. First, any sort of *unconditional* inheritance of experiential properties is liable to entail consciousness in many systems we would normally have considered clearly non-conscious, such as the solar system, the galaxy, or any number of scattered mereological sums. This will apply also if the inheritance of experiential properties is conditional, but based ultimately in the unconditional inheritance of their additive ‘core properties’ (as was suggested for ‘being punched’). For then whatever core there is to experiential properties, not captured by their overall functional role, will be inherited by mereological fusions. For similar reasons, panpsychist combinationists will find that the multitudinous experiences of the micro-level subjects they postulate may be inherited by human beings: our experience will be, contrary to initial appearances, ‘over-stuffed’ with trillions of microexperiences. These problems were prefigured in the last chapter, in the discussion of mega-subjects and micro-subjects; they will be addressed in chapters 6 and 5.

Despite the difficulties just canvassed, there are also advantages in pursuing a sharing-inheritance-connections analysis of experiential combination. For a start, it provides a direct, and distinctly metaphysical, sense in which the experiences of the whole are nothing over and above, and thus may be explained by, those of the parts. Indeed, some critiques of combinationism (in the form of constitutive panpsychism) start from the assumption that it must involve something like the token-sharing of experiences if it is to be intelligible (James 1909, p.181, Basile 2010, p.108, Coleman forthcoming, pp.14-16). Moreover, it is parsimonious: explaining a composite mind as involving the same tokens of experiential properties as its component minds seems preferable to duplicating tokens at each mereological level.

There is even an advantage in being open to all the challenges discussed above: the sharing-inheritance-connections schema brings directly into focus what seems problematic about experiential combination, such as the conflict between token sharing and the privacy of experience, and the threat of overgeneration and its relationship to the possibility of phenomenal overflow. Other frameworks run together the transition from component subjects to composite subjects with the transition from one set of experiential properties to another: by separating these steps, this schema allows a clearer discussion of specific objections to each.

Finally, as noted at the end of the last section, the sharing of experiences between part and whole gives combinationists a way to meet the independent grasp requirement on experiential bonding relations: if our component subjects share our experiences, then our introspective grasp on the relations among those experiences gives us a grasp on at least those relations among our component subjects that are mediated by their experiences.

Summary:

We have reviewed four models of what makes physical combination intelligible, and what is therefore needed for experiential combination. These models are not necessarily opposed: they might all capture fragments of the truth. For instance, it might be that continuity of natures, a metaphysical criterion, dovetails with *a priori* deducibility, an epistemological criterion, because only when the explained and explaining properties are conceptually akin to one another can one conceptually entail the other. Or the idea of properties being shared and inherited might be judged more attractive because general in application: many different properties of many different wholes can be explained in the same way, given their presence in the parts.

Consequently I will not choose any single criterion, but keep all on the table. As we explore various challenges to combinationism - most especially its apparent explanatory gap, the topic of chapter 3 - we will be able to see the different implications of adopting different standards of intelligibility, and can evaluate which ones combinationism can and cannot meet. That said, this chapter has revealed some interesting choices in how we approach combinationism, which it will be useful to run over here. This discussion (of different approaches to making combinationism intelligible) complements the discussion in chapter 1 section 3 (of different versions of combinationism for different prior views of consciousness and composition).

First, there are *a priori* and *a posteriori* approaches: adherents of the former think that in order for microexperiential facts to genuinely explain macroexperiential facts, even just in our world, they must entail macroexperiential facts in every world. On this view the very natures of parthood and experience guarantee combinationism. On the *a posteriori* approach, only the laws of nature operative at our world allow microexperiential facts to explain macroexperiential facts: what makes these laws suitably intelligible is their generality of application, or the metaphysical relations between the experiential properties and relations they deal with.

We should also distinguish different approaches to our grasp of the experiential bonding relations. A physicalist combinationist need not see an especially pressing problem here: the relations involved in experiential combination need not be fundamentally experiential relations, and so might be detectable in fairly normal ways. But a primitivist combinationist might feel a need for distinctively experiential bonding relations, and these would have to be grasped in one of two ways: the ‘outward-looking’ way, via the phenomenology of our relationships with other macro-subjects, and the ‘inward-looking’ way, via the relations among our own experiences.

In fact, I believe that there is a suitable bonding relation which we can grasp in both the outward-looking and the inward-looking ways, a relation which I call, inspired by Husserl, ‘adumbration’, and which forms the centrepiece of chapter 4’s discussion of the unity of consciousness. The principal remaining difference between the two approaches is then the further commitments they imply: the outward-looking approach requires accepting mega-subjects, while the inward-looking approach requires accepting the token-sharing of experiences.

Finally, we can connect views on experiential inheritance with views on phenomenal overflow (phenomenal consciousness without cognitive access). If phenomenal overflow is possible, then there may be more space for a combinationist to accept unconditional experiential inheritance, and the resultant profusion of cognitively-inaccessible phenomenal consciousness both in mega-subjects and in human beings (if there are micro-subjects). Thus combinationists might pursue two divergent approaches: using the possibility of phenomenal overflow to argue that the consciousness of composite things may in fact *include* all or much of the consciousness of all their parts, or using the impossibility of phenomenal overflow to explain why so much of the consciousness of a thing’s parts is *excluded* from the consciousness of the whole. Call these the ‘inclusionary’ and ‘exclusionary’ approaches to combinationism: a combinationist who defends any unconditional inheritance of experiential properties must take the inclusionary approach, while others may take the more conservative exclusionary approach. This distinction will prove important in chapters 5 and 6, where I consider the problems posed by mega-subjects and micro-subjects.

Chapter 3: Conditional Inheritance, Experience Sharing, and the Problem of Subject-Summing

<u>Section 1: The Subject-Summing Problem</u>	p.59
<i>Subsection 1.1: Examples of people in groups</i>	p.60
<i>Subsection 1.2: Microexperiential zombies and Microexperiential Super-Zombies</i>	p.62
<i>Subsection 1.3: The Role of Sympathetic Imaginability</i>	p.63
<u>Section 2: The Easy Case and the Inheritance Proposal</u>	p.64
<i>Subsection 2.1: The Easy Case and the Homuncular Zombie</i>	p.65
<i>Subsection 2.2: The Inheritance Proposal</i>	p.68
<i>Subsection 2.3: Versions of the Inheritance Proposal</i>	p.69
<u>Section 3: In Defence of Experience-Sharing</u>	p.72
<i>Subsection 3.1: Strong and Weak Exclusivity, Strong and Weak Sharing</i>	p.74
<i>Subsection 3.2: Metaphysical Arguments for Exclusivity</i>	p.75
<i>Subsection 3.3: Exclusivity and Privacy, Sharing and Publicity</i>	p.77
<i>Subsection 3.4: Phenomenal Holism and the Incompatible Characters Argument</i>	p.78
<i>Subsection 3.5: Responding to the Incompatible Characters Argument</i>	p.81
<i>Subsection 3.6: Responding to Coleman's Argument</i>	p.85
<i>Subsection 3.7: The Targets of Sympathetic Imagination</i>	p.86
<u>Section 4: Components of Experiential Ownership</u>	p.87
<i>Subsection 4.1: Causal Analyses of the Ownership Relation</i>	p.89
<i>Subsection 4.2: Ontological Analyses of the Ownership Relation</i>	p.92
<i>Subsection 4.3: Distinctively Experiential Analyses of the Ownership Relation</i>	p.95
<i>Subsection 4.4: Putting it All Together</i>	p.98
<i>Subsection 4.5: Conditional Experience Inheritance and Basic-Experience Inheritance</i>	p.101
<u>Section 5: Making Basic-Experience Inheritance A Priori</u>	p.103
<i>Subsection 5.1: Motivating the Heritability of Fundamental Properties</i>	p.105
<i>Subsection 5.2: Basic-Experiential Properties Without Unity, Integration, or Control</i>	p.109
<i>Subsection 5.3: The Argument from Unimaginability</i>	p.111
<i>Subsection 5.4: Return of the Microexperiential Super-Zombies</i>	p.113
<i>Summary</i>	p.116

The experiential combinationist claims that understanding the consciousness of a thing's parts explains - renders intelligible, lets us understand - the consciousness of the whole thing. Yet it is far from obvious that it does anything of the sort, and a central strand of this apparent 'explanatory gap' is what I call, following Goff (2009b) and Chalmers (forthcoming-a, p.4) the 'subject-summing' problem, the problem of whether the mere being-conscious of a composite with conscious parts can be explained by the consciousness of those parts. Subjects, *qua* subjects, seem insulated from each other in a way that precludes any experiential facts about one subject from explaining experiential facts about another. So the experiential combinationist must confront a principled concern that the experiential properties of a thing's parts are necessarily irrelevant to its experiential properties, in a way that nobody would suggest the physical properties of its parts are necessarily irrelevant to its physical properties.

Other strands of the explanatory gap problem involve explaining particular features of the consciousness of a composite subject, such as its unity or qualitative diversity, or considering experiential changes in a thing's parts that seem compatible with experiential invariance in the whole. These directly connect with other problems, involving what I earlier called 'latent incoherence' arguments about phenomenal unity and the implications of 'micro-subjects' of the sort postulated by panpsychists. I address them in detail in chapters 4 and 5. But addressing these problems would be pointless if the more basic subject-summing problem remained unresolved, so in this chapter I consider that problem. I first outline the problem, and a minimal proposal to address it, involving the conditional inheritance and token-sharing of experiential properties. The final three sections then evaluate this proposal, arguing that given the right premises, versions of it can provide adequate explanations of macro-subjectivity given almost any background view about the nature of experience and the requirements for explanation.

Section 1: The Subject-Summing Problem

The idea lying behind the subject-summing problem is formulated by Goff thus:

No Summing of Subjects (NSS): It is never the case that the existence of a number (one or more) of subjects of experience with certain phenomenal characters *a priori* entails the existence of some other subject of experience. (2009a, p.302; a slightly different formulation appears at 2009b, p.130)

However, while the NSS principle is a good start, it is not quite the right principle to be evaluating, for at least two kinds of combinationist might happily endorse it. First, it does not make any mention of relations among subjects, and so a combinationist who assigned a crucial role to relations might accept it. Second, it is specifically formulated in terms of *a priori* entailment, and so a combinationist who

employed some other standard for intelligible explanations might accept it. For a claim that all combinationists must deny, we should consider a slightly adjusted version of NSS:

Explanatory Gap between Subjects (EGS): It is never the case that the existence of a number (one or more) of subjects of experience with certain phenomenal characters, standing in independently intelligible relations¹, renders intelligible the existence of some other subject of experience.

There is at least some plausibility in this claim: it seems that conceiving of any number of conscious subjects, aware of *their* experiences, does not entail any further subject, aware of *its* experiences. In this section I review some existing discussions of this subject-summing problem, and the sorts of examples and thought-experiments employed to support EGS.

Subsection 1.1: Examples of people in groups

The explanatory gap between subjects is sometimes dramatised by considering groups of human beings, and claiming that nothing said about the human beings individually seems to entail consciousness in the group itself.

For instance, Brentano and Plotinus both, in arguing for the simplicity of the mind, claim that sensory experiences in distinct parts of something could not be directly compared, because they would be analogous to experiences in two people, and pluralities of people cannot directly compare their experiences (Plotinus 1956, p.346, Brentano 1987, p.293). Louis de Courcillon makes the same point, saying that “it is as impossible for the one [part] to sense what the other senses, as it is impossible that

¹ ‘Independently intelligible relations’ are those which satisfy the ‘independent grasp’ requirement discussed in subsection 3.1 of chapter 2.

we sense in this room the pleasure that those who are at the Opera are presently sensing.” (1684, trans. Schachter 2002, p.250) James makes a similar claim in arguing against a certain form of panpsychism:

Take a hundred [feelings], shuffle them and pack them as close together as you can (whatever that may mean); still each remains the same feeling it always was, shut in its own skin, windowless, ignorant of what the other feelings are and mean. There would be a hundred-and-first feeling there, if, when a group or series of such feelings were set up, a consciousness belonging to the group as such should emerge. And this 101st feeling would be a totally new fact... Take a sentence of a dozen words, and take twelve men and tell to each one word. Then stand the men in a row or jam them in a bunch, and let each think of his word as intently as he will; nowhere will there be a consciousness of the whole sentence. (1890, p.160)

One way to read such analogies is as presenting *reductio* arguments, with one conditional and one negative premise: if conscious parts produced conscious wholes, then groups of human beings would have a collective consciousness; but they do not; hence conscious parts do not produce conscious wholes. But this reading is uncharitable, because the first premise seems needlessly strong: it is hardly fair to assume that if something is possible, it must occur in all cases. Moreover, the second premise, while intuitive, is actually denied by James: he does not say that the ‘101st feeling’ definitely does not appear, but that even if “by a curious physical law” it did, it would not be explained by the component feelings (p.160).

The reading I prefer is that the analogy to a group of people simply *illustrates* the failure of explanation claimed by EGS: by considering the parts of composite subject as though they were many little humans, we see vividly that no specification of facts about them could by itself entail any consciousness in the group. After all, we routinely attribute consciousness to people while refusing to do so with groups of them.

Two things are worth noting about this argumentative strategy. First, the parts of a human being are related to each other differently from the human members of any known group: they are

bound together more tightly, are more sensitive to each other's state, transmit information more quickly and directly, and so on. Since this difference in relations may well play a role in the constitution of a composite subject, these analogies provide stronger support for the NSS principle than for the EGS principle. Second, James's discussion is not as specifically directed at *experiential* combination as has often been assumed (Cf. Shani 2010): he is in fact a mereological nihilist who holds that "Atoms of feeling cannot compose higher feelings, *any more than atoms of matter can compose physical things*" (1890, p.161, emphasis added). Nevertheless I include James's passage because it has taken on a life of its own, being repeatedly quoted by authors registering an intuitive difficulty specifically with experiential combination (e.g. Seager 1995, p.280, Goff 2006, p.54, Strawson 2006, p.26 Fn.48, Tononi & Koch 2014, p.6).

Subsection 1.2: Microexperiential zombies

A different way to show the explanatory gap is inspired by the 'philosophical zombie', a being physically and functionally identical to a human being but wholly lacking in consciousness, whose apparent conceivability threatens physicalism. The possibility of such an argument is already noted by Carruthers and Schechter (2006, p.38), but Goff (2009a) develops it at length, arguing for the conceivability of a 'microexperiential zombie', defined as a "physical duplicate of an actual organism which is such that there is something that it is like to be each of its ultimates [but which] does not have o-experience"(p.296), where 'o-experience' is "the conscious states pre-theoretical common sense attributes to organisms"(p.290). An example of such a zombie is:

My itchy twin... a physical duplicate of me such that each of its physical ultimates *feels itchy*... my itchy twin has no o-experience. If you stick a knife in him he will scream and run away, but he will not *feel pain*. My itchy twin successfully negotiates a three-dimensional world using his eyes, but he has no *visual experience* of that world."(p.296)

Now, the conceivability of microexperiential zombies in this particular sense does not actually support EGS, for it is no part of the definition of a microexperiential zombie that it lacks experience altogether and thus fails to be a subject; it may have the simple and disconnected experiences of its parts. Thus the conceivability of Goff's itchy twin does not establish that experiential facts about one subject cannot explain experiential facts about another - but merely that they cannot explain a certain sort of experience.

But we can define a being which is physically and functionally identical to a human being, wholly lacking in consciousness, but composed of conscious microscopic parts. Call this being, whose conceivability would support EGS, a 'microexperiential super-zombie'. As with microexperiential zombies, we can devise any number of specific types (itchy, pained, euphoric, etc.) by attributing different experiences to the parts. The difference is that when you stick a knife into an 'itchy microexperiential super-zombie', not only does it feel no pain, it also feels no itch.

Even the microexperiential super-zombie is a threat only to *a priori* forms of combinationism. *A posteriori* combinationists can accept the conceivability of such a being, but say that what is being conceived of is simply a world with different laws of nature - in particular, without the laws of nature that govern the generation of macroexperience by microexperience in our own world. The fact that a law is not conceptually necessary does not mean that it does not actually hold, though it does mean that it is not really explanatory by the standards of explanation employed by *a priori* combinationists.

Subsection 1.3: The Role of Sympathetic Imaginability

Here is another way to think about subject-summing. Plausibly, 'x is a conscious subject' is equivalent to 'there is something it is like to be x', and the latter might be equivalent to 'it is possible in principle

to *sympathetically imagine* being x ', where 'sympathetic imagination' is the mental process of "put[ting] oneself in a conscious state resembling the thing itself" (Nagel 1974, Fn11). If something has experiences, it should be possible to ask what it is like to be it, and imagine various answers. It may be that accurately imagining being some creatures is in practice impossible for us, as it is likely impossible for us to imagine the sensory experiences of someone or something with different sensory modalities from us, but there should be at least some possible act of sympathetic imagination which would count as accurate if directed onto the being in question.

Granting this equivalence of consciousness with imaginability, the question of subject-summing is whether, given some collection of subjects somehow related, each of which we could imagine being, some further act of imagination qualifies as imagining being the whole they compose. Does it make sense, given some conscious beings, to try and imagine 'being all of them'? What imaginative act could be appropriate to the group? It might seem impossible to make sense of imagining being all of them: as Barnett says, even if we can imagine being Descartes and can imagine being Hobbes, it makes little sense to imagine what it was like to be both, for "surely there is nothing it is like to be a pair of people" (2008, p.312).

Section 2: The Easy Case and the Inheritance Proposal

The examples considered in the last section tend to undermine the plausibility of combinationism, and in particular of explanatory relations between distinct subjects. However, there are other cases which may push us in the opposite direction, such as those which fit a certain abstract specification I will call 'the easy case'. Building off these cases, combinationists can develop a systematic account of explanatory relations among subjects that is, I shall argue, plausible enough to overcome the objections

canvassed in the preceding section. In this section I explain the easy case, the proposed account, and the versions of it appropriate to different combinationists.

Subsection 2.1: The Easy Case and the Homuncular Zombie

Suppose a group of entities compose a highly-integrated whole, and one of them is a conscious subject. Suppose that this one part's stream of consciousness plays a central role in guiding the overall activity of the whole - when that part wishes another part to move a certain way, the other part does so, and events happening in the other parts produces conscious perceptions in this part.

It seems plausible that, given these facts about its parts, the whole entity has experiences - more precisely, has exactly the same experiences (both type- and token-identical) as its one conscious part. It is at least overwhelmingly plausible that *if* the composite has experiences, its experiences are simply those of its one conscious part - at least, given something like a 'that's-all' premise, ruling out any extra emergent forces or extraneous factors. Moreover, if the whole shares the experiences of this part, it clearly has them *in virtue of* that part having them (and, perhaps, being appropriately connected to the other parts), and its having them is *explained by* that part having them (and being so connected).

This plausibility appears especially when we consider two particular instances of the easy case: first, the human body as described by materialists, which might be thought conscious in virtue of events going on in its brain, and second, the human being as described by substance dualists, a composite of soul and body which might be thought conscious in virtue of events going on in its soul. Even here, the explanatory claim a combinationist might want to make is not uncontroversial: a dualist might deny that the soul-body composite is really conscious, while a materialist might either deny that the whole body is really conscious, or affirm its consciousness while denying that of the brain. Nevertheless, I

think cases like this are the best place for combinationists to look for a starting point of intuitive plausibility.²

The plausibility of thinking that the whole in the easy case shares the experiences of its part is reflected in the fact that even Goff, in his defence of NSS, actually accepts it. The relevant instance of the easy case is a particular type of microexperiential zombie, which Goff calls the ‘homuncular zombie’. This is a physical and functional replica of a human, which lacks all complex human experience, but one microscopic part of which does have all the richness of human experience. Given that NSS (and *a fortiori* EGS) entails the conceivability of homuncular zombies, it is surprising that Goff holds such creatures to be inconceivable precisely because by attributing complex experiences to a part of the zombie one *ipso facto* attributes experiences to the whole zombie – i.e. because mental properties are inherited in this case, in spite of NSS.

This tension in Goff 2009a has not been widely recognised, but could be resolved in two ways (in correspondence with Goff both were mooted). On the one hand, one might weaken NSS to the different version given in Goff 2009b, which rules out subjects summing into a subject with *different* experiences from theirs. This version of NSS no longer concerns *subjects*, but only how to get the right experiences for those subjects. Thus it amounts to giving up the strict denial of explanatory relations between distinct subjects. On the other hand, one might qualify the inconceivability of homuncular zombies by saying that the homuncular zombie has experiences ‘in a merely derivative sense’, allowing therefore that while wholes may inherit experiential properties from their parts, they have those

² Won’t there be a difference in the content of the consciousness we can attribute to the whole and to one part - for instance, when the brain ‘wishes the arm to move’, won’t it think of the arm (an external thing) differently from how the whole person does (who thinks of the arm as a part of itself)? In this chapter I am ignoring this issue, since my concern is not the particular content of consciousness but whether, *supposing* that the part has whatever consciousness we wish to attribute to whole, its having that consciousness is enough for us to do so. In chapter 7, however, I directly address this question about awareness of self and other as such.

properties ‘in a merely derivative sense’. This demands a brief digression on the meaning of this phrase.

In one sense, saying that an inherited property is had in a derivative sense just means that the bearer *derives* it from something else, i.e. has that property in virtue of something else having it. But it could also mean something stronger: that the whole doesn’t *really* have that property at all. This would be like a demographer counting the children of Christians as Christians, regardless of their actual beliefs or practices: the children are ‘Christians’ in a merely derivative sense, which contrasts with actually being Christians. In the experiential case, we can approach this question by asking whether, for some intended meaning of ‘derivative’, we ourselves might have experiences ‘in a merely derivative sense’. If we might – if our acquaintance with our phenomenology does not rule that out – then it is no defect in a compositional explanation of our minds that it attributes us experiences in that kind of derivative sense. If, on the other hand, we can be sure that we do not have experiences ‘in a merely derivative sense’, because that would contrast with literally having experiences, then nothing I say about inheritance principles should be read as speaking merely in *that* kind of derivative sense.

In particular, any principle of inheritance that ascribed the inherited properties in a ‘merely derivative’ sense, where that contrasts with ascribing them literally, would be uninteresting, and perhaps even tautologous. For instance, if the claim “everything with a square part is ‘square’” uses the term ‘square’ in a non-literal sense, as meaning ‘has a square part’, then it is a boring tautology. The claim is substantive (and in this case, false) because it attributes properties in their literal, primary sense – though of course they may be ‘derivative’ in the other, weaker, sense of being derived from something else.

Subsection 2.2: The Inheritance Proposal

Extrapolating from the easy case, a combinationist might propose the following two theses to bridge the explanatory gap between distinct subjects:

Token-Sharing of Experiences (TSE): Particular experiences can belong simultaneously to two different entities.

Conditional Experience Inheritance (CEI): A whole has an experiential property whenever one of its parts does and that part is appropriately related to its other parts, simply in virtue of the part having that experiential property and being appropriately related to the other parts.

One immediate question is what ‘appropriately related’ means in Conditional Experience Inheritance: this will be explored in more depth in section 4, but can be summarised here as some combination of *phenomenal unity*, *causal integration*, and *intelligent joint control of behaviour*. Whatever relations are involved in the conditions specified by CEI will to that extent function as experiential ‘bonding relations’ of the sort discussed in chapter 2, subsection 3.1.

Another question is how we are to think of ‘experiences’, these particulars that TSE asserts can be shared; as noted in chapter 1, subsection 1.2, I have so far left their ontology open beyond the stipulation that we have experiences whenever we instantiate experiential properties. Their ontology will become more important now, because a key part of the motivation behind the above two theses is the idea that wholes do not generally have new experiences distinctive to them: rather, they inherit (under the right conditions) the experiences of their parts - these token experiences belong both to the whole and to the part. This allows combinationists to shift their focus away from explaining the presence of experiential properties *per se* in the whole, and towards explaining why, given the presence of certain experiences, the whole relates to them in the right way to qualify as ‘having’ them.

Insofar as the easy case makes it plausible that wholes can have the experiences of their parts in at least some cases, combinationists may hope that it provides an initial motivation for the these two theses, which I will refer to collectively as ‘the inheritance proposal’. But a more compelling case for this proposal demands a more thorough investigation of the nature of the properties under discussion. Thus the remainder of this chapter provides a three-step defence of the inheritance proposal. In section 3 I clarify and defend TSE, arguing that it can be made consistent with most views on the ontology of token experiences, as well as with a qualified doctrine of ‘privacy’ for experiences. In section 4 I then undertake a conceptual analysis of the relation between subjects and their experiences, arguing that several candidate accounts of the relation entail CEI. Finally, in section 5 I consider the remaining accounts of this relation, on which it is primitive and fundamental, arguing that given certain background views on the part-whole relation, fundamental properties are unconditionally inherited by default, supporting Conditional Experience Inheritance.

Subsection 2.3: Versions of the Inheritance Proposal

My proposal for a combinationist account of the explanatory connections among subjects is compatible with many versions of combinationism, though they may require subtly different interpretations of its two component theses.

A first difference is between what in the last chapter I called ‘inclusionary’ and ‘exclusionary’ approaches to combinationism. The former excludes many of the parts’ experiences from the consciousness of the whole, while the latter includes them as a form of ‘phenomenal overflow’. An inclusionary combinationist will explain Conditional Experience Inheritance in terms of some kind of *unconditional inheritance* - of experiential properties, or more plausibly of those components of experiential properties which are not conceptually tied to any systemic role, and which are thus such as

to be logically capable of being unconditionally inherited (what I earlier called ‘additive’ properties). This parallels the last chapter’s suggestion that ‘being punched’ requires *both* having certain punching-type events transpire in you (an inherited, additive property) and also having those events be prominent in your internal goings-on (a non-heritable, systemic property), and is thus inherited by a person from their nose, but not by the universe from said nose. The exclusionary combinationist, on the other hand, may take Conditional Experience Inheritance to be prior to any sort of unconditional inheritance.

A second important difference arises between subject-first combinationists, who can endorse CEI straightforwardly, and experience-first combinationists, who must consider experiential combination (the combination of subjects) in relation to what they consider more basic, the combination of sets of experiences. For one subject to be part of another, on their view, means for it to be constituted by a subset of the experiences which constitute the other. Reading CEI with this notion of ‘part’ in mind, it amounts to saying that when some set of experiences constitutes a subject with certain experiential properties, any set of experiences which contains that set, and whose members are appropriately related to one another, will also constitute a subject with those same experiential properties.

Experience-first combinationists can also take CEI and TSE in a secondary sense, as speaking not of subjects strictly so-called, but of enduring material things which indirectly constitute subjects by generating (or ‘underlying’, or ‘constituting’, etc.) experiences. Since the only ‘experiential properties’ such things can have are ‘generation properties’ (the property of generating, realising,, giving rise to, etc. an experience of a certain sort), the analogues of CEI and TSE would state that any composite, some part of which generates an experience, also generates that experience if that part was suitably related to the others, and that a single instance of a generation property can belong to two distinct bases.

A third difference is between the *a priori* combinationist and the *a posteriori* combinationist. While both may regard TSE as a mere statement of metaphysical possibility, they will differ in their understanding of CEI (as well as any claims of unconditional inheritance that, on the inclusionary approach, are meant to explain it): the *a priori* combinationist will treat it as a conceptual necessity, a mere consequence of the nature of experience and composition, while the *a posteriori* combinationist may treat it as basic law of nature, nomologically necessary but metaphysically and conceptually contingent. For the former, the explanatoriness of the inheritance proposal lies in its *a priority*; for the latter it lies in something else, such as the metaphysical connection forged by TSE, the conceptual continuity of antecedent and consequent, or the simplicity and explanatory power of the posited laws.

Fourth, while the physicalist combinationist and the primitivist combinationist may endorse the same principles about experiential properties, they will understand those properties differently: for the physicalist they are ultimately reducible to some other sort of properties, while for the primitivist they are not. This will affect what sort of explanation they can offer for CEI - if experiential properties can be analysed in non-experiential terms then it is reasonable to try and establish CEI through a conceptual analysis of the sort explored in section 4, but if not this may prove impossible: establishing CEI might then require the different approach pursued in section 5.

Finally, combinationists with different views of composition will understand CEI differently: the pluralist combinationist will take the whole's having the relevant properties to be grounded in the part's having them, while the monist will reject the 'in virtue of' clause of CEI, instead endorsing the related principle CEI_M :

Monistic Conditional Experience Inheritance (CEI_M): A whole has an experiential property whenever one of its parts does and that part is appropriately related to its other parts, and that part has that experiential property simply in virtue of the whole having that experiential property, and the part being appropriately related to the other parts.

The supporter of composition-as-identity will accept both CEI and CEI_M (I take grounding to be reflexive insofar as identity claims are one way to substantiate a claim of grounding). A nihilist combinationist will take the whole notion of the whole's inheriting these properties as a circumlocution for CEI_N:

Nihilistic Conditional Experience Inheritance (CEI_N): Some things have an experiential property collectively whenever one of them does so individually and that thing is appropriately related to the other things, and the things have that experiential property collectively simply in virtue of the particular thing having that experiential property individually, and being appropriately related to the other parts.

What all these parties agree on is that the part and the whole having the property in question are not two independent facts; my defence of CEI will apply just as well, *mutative mutandis*, to CEI_M and CEI_N.

Section 3: In Defence of Experience-Sharing

TSE could be contested in two styles. First, it might be rejected as incoherent, because the way we individuate token experiences do not allow us to make sense of their being shared - as on Bayne's 'tripartite account', on which experiences are individuated by their phenomenal character, subject, and time (Bayne 2010, pp.24-29). But this kind of objection to experience-sharing parallels the way that some ontologies of properties might refuse to countenance the sharing of tokens in general, as discussed in subsection 4.1 of chapter 2. However, we do have compelling examples of a distinctive form of intelligible grounding between property instances in wholes and in their parts (e.g. the redness of a surface and its sections, the torn-ness of a coat and its sleeve), and this grounding is naturally expressed by saying that a single property token belongs to both. If our ontology or our definitions do not allow us to say this, we will need some other account of this sort of grounding, and claims like TSE should be construed in terms of that other account.

Moreover, we have significant freedom in how we individuate experiences, and could employ different criteria for different theoretical purposes: as Bayne says, “Counting experiences is arguably more like counting the number of objects in a room... [than] like counting the number of beans in a dish... the idea that there is only one way in which to proceed is somewhat farcical.”(2010, p.24) Thus I will put aside this kind of technical objection to TSE, assuming that we can if necessary define or construct some way of formulating what is important in the claim, namely that a whole’s consciousness could be ‘nothing over and above’ that of its part in the same way that a whole’s being striped could be ‘nothing over and above’ a certain part of it being striped.³

The interesting question is whether there is some substantive reason for thinking that experiences cannot be analogous to material property instances in this regard. Claims of such a difference are not hard to find: TSE conflicts with the intuitive idea that experiences belong to their subjects ‘exclusively’, an intuition expressed emphatically by James:

Each of [our] minds keeps its own thoughts to itself. There is no giving or bartering between them. No thought even comes into direct sight of a thought in another personal consciousness than its own. Absolute insulation, irreducible pluralism, is the law... The breaches between such thoughts are the most absolute breaches in nature. (1890, p.226)

Dainton gives a more precise formulation of this idea, writing that "it seems plausible to think that subjects and experiences are governed by an Exclusivity Principle along these lines:"

If an experience e_1 belongs to a subject S_1 , it belongs ONLY to S_1 , so e_1 cannot also (and simultaneously) belong to a distinct subject S_2 . (2011, p.246)

James and Dainton both regard this ‘exclusivity’ as a problem for experiential combination, which it clearly is if combinationists defend the Token-Sharing of Experiences. In this section I consider what can be said in support of exclusivity, arguing that while there may be reason to rule out sharing

³ In particular, a slightly amended version of the tripartite account is compatible with TSE. On this amended account, experiences are individuated by their time, quality, and *subjects*, that is by the set of all subjects they belong to.

between discrete subjects, they do not extend to sharing between overlapping subjects: either those which fully contain each other as parts, or those which overlap partly with or without forming any overarching whole.⁴

Subsection 3.1: Strong and Weak Exclusivity, Strong and Weak Sharing

Exclusivity concerns sharing experiences between one subject and ‘another’, but there are two ways to read this ‘another’: we might mean simply a ‘distinct’, i.e. non-identical, subject (as in Dainton’s formulation), or we might mean ‘discrete’, i.e. non-overlapping, subjects. That yields the following four theses:

Strong Exclusivity (SE): A single experience cannot belong to multiple distinct subjects.

Weak Exclusivity (WE): A single experience cannot belong to multiple discrete subjects.

Weak Sharing (WS): A single experience may belong to multiple distinct subjects.

Strong Sharing (SS): A single experience may belong to multiple discrete subjects.

WS and SS are two versions of TSE, the Token-Sharing principle discussed in the last subsection. WS is the negation of SE, while SS is the negation of WE; SE entails WE and SS entails WS. But, crucially, WE and WS are compatible – experiences might be weakly shareable yet also weakly exclusive. I will argue that we have good reasons to reject SS, but not WS, so that combinationists can

⁴ We could get a hypothetical example of overlapping subjects that did not form any overarching whole by extending the real-world phenomenon of craniopagus twins, conjoined twins fused at the skull. Such twins can have nervous tissue connecting their brains, and there is no reason in principle that there could not be shared brain parts, connected with and integrated into both brains. There might then 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 currently only 7 years old. Cf. Langland-Hassan 2011.

endorse TSE as long as they understand it as restricted so as to be compatible with WE, i.e. as long as it is read as WS, not SS.

Subsection 3.2: Metaphysical Arguments for Exclusivity

Both subject-first and experience-first views of the metaphysics of subjects provide resources to argue for exclusivity: however, under scrutiny they turn out to support only WE, not SE.

On the one hand, the subject-first view permits the following argument: experiences are ‘adjectivally dependent’ on their subjects, in a manner so metaphysically intimate that they lack the independence needed to belong to any other subject. Experiences are simply modifications of, or acts of, or exercises of the powers of, subjects. This is the view Noonan expresses when he writes:

The concept of *someone’s having* a perception is logically prior to the concept of *a perception*... the relation between the self and its perceptions is analogous to that between the sea and its waves. The waves are modifications of the sea and perceptions are modifications of the self. (Noonan, 1990, p.71)⁵

This may be a good argument against Strong Sharing, for it is hard to see how two discrete, independent, substances could share a modification. However, all this is compatible with Weak Sharing, because the intimacy and ontological dependence of the experience-subject relation is also present in the whole-part relation. For an experience to be ontologically dependent on a subject does not conflict with its also being ontologically dependent on some part of that subject, as becomes obvious when we consider that the very examples used to convey ‘adjectival dependence’ seem to admit of sharing. There is no incoherence in a wave depending on multiple overlapping parts of the sea, or (to employ an example from Shoemaker) a dent depending on multiple overlapping surfaces. So if

⁵ Compare Shoemaker, 1985: “Examples given of experiences are typically examples of *experiencings* and *experiencings* are patently adjectival [on subjects].”

sharing does not conflict with the priority of material substances over their modifications, it should not conflict with the priority of subjects over their experiences. So I conclude that the subject-first view provides *prima facie* reason to reject Strong Sharing, but not Weak Sharing.

The experience-first view, on the other hand, seems to make sharing of experiences impossible, because if subjects are constituted by their experiences they may well be individuated by them too, making it hard to see how two subjects sharing experiences could really be *two*. Here we should distinguish ‘total’ and ‘partial’ sharing: the former involves different subjects who share *all* their experiences, while the latter involves different subjects who share some but not all of their experiences with each other.

Now, *total* sharing seems impossible on the experience-first view, for if two subjects have all the same experiences, and are nothing over and above their experiences, then surely they will be the same subject. Moreover, *partial* sharing between *discrete* subjects also seems impossible, because for entities made up of their experiences, to ‘overlap’ simply means to be made up of overlapping sets of experiences, in which case any subjects who share experiences will presumably both be partly constituted by the shared experience, and just on that basis will not be discrete.

Even if these arguments are accepted, however, they leave ample room for the kinds of experience-sharing that combinationists would want to postulate. For a start, they clearly allow for partial sharing between distinct-but-overlapping subjects. Admittedly it does not allow for total sharing, even between overlapping subjects, but it allows for a straightforward work-around: shared generation properties. Wherever a subject-first combinationist would posit total sharing (e.g. a human being and their head), the experience-first combinationist can say that two entities (the human being and the head) share all the same instances of properties of generating experiences - that is, they generate all the same experiences. These experiences constitute a single whole subject, which has both body and head as

bases.⁶ Thus insofar as we read ‘experiential properties’ broadly, as including generation properties, the experience-first view is compatible with both total and partial sharing of experiential properties, at least among subjects which are distinct but not discrete.

Subsection 3.3: Exclusivity and Privacy, Sharing and Publicity

Plausibly, it is definitional of conscious experience is that it is in some sense ‘subjective’, not ‘objective’, and this plausibly entails that experience is ‘private’, directly knowable by only one subject. This distinguishes it from the world of matter, all facts about which are ‘public’, directly knowable by all subjects equally. Obviously much depends on what is meant by knowing ‘directly’, but it will not matter here what account we give of this distinctive form of knowledge. What matters is that we can argue from privacy to exclusivity: plausibly, my undergoing an experience is both necessary and sufficient for my being in a position to know it directly, while others, who are not me and hence cannot undergo my experiences, cannot know them except indirectly.⁷ We can express this as:

Knowledge by Ownership (KO): Having an experience is necessary and sufficient for being able to know it directly.⁸

Given KO, TSE seems to violate privacy, making experience public by allowing multiple subjects to know the same experience directly. Now, there might be ways to break this link between exclusivity and privacy, somehow barring all but one of the subjects which undergo an experience from knowing it directly. But a much simpler approach is available: just as we distinguished Strong Exclusivity from

⁶ Note that generation properties may well be adjectivally dependent on their bearers, in the same way as waves are dependent on the sea; in that case they likely could not be shared between *discrete* entities.

⁷ Indeed, the theses may not be distinguished at all; for instance, Unger 1990 describes the thesis of ‘the privacy of experience’ thus: “Except for that particular subject himself, nobody else and nothing else can *have* that conscious experience that he has. As another gloss on this idea, nobody else, and nothing else, can be *directly conscious of* the experience of that particular subject.”(p.40, emphasis added)

⁸ Note that this does not say that we *will* know all our own experiences, or be able to know them *easily*; ‘able’ here is meant to convey whatever sense of ability is employed in stating the privacy of experience.

Weak Exclusivity, and observed that denying the former is compatible with maintaining the latter, we might distinguish the following two theses:

Strong Privacy: A single experience cannot be directly known by multiple distinct subjects.

Weak Privacy: A single experience cannot be directly known by multiple discrete subjects.

Combinationists can try to capture the intuitive force behind Strong Privacy by endorsing Weak Privacy in its stead. In particular, they can point to the sharp contrast that remains between Weak Privacy and the publicity of physical fact; to parallel the distinction between Weak Sharing and Strong Sharing, we can distinguish Weak Publicity from Strong Publicity, as the negations of Strong and Weak Privacy:

Weak Publicity: A single experience may be directly known by multiple distinct subjects.

Strong Publicity: A single experience may be directly known by multiple discrete subjects.

Physical facts are *strongly* public, while experiential facts are merely weakly public. Whereas a physical fact can in principle be known equally well by any subject, an experience is directly knowable only by those intimately and directly involved in its occurrence. The fact that more than one distinct subject may be intimately and directly involved in a single experience's occurrence need not render this contrast less significant.

On the other hand, making experiences strongly public would undermine this contrast, and to that extent seems to go against a defining characteristic of experience. Given KO, Strong Sharing would imply Strong Publicity, and so to that extent we might think we have reason to deny Strong Sharing and maintain Weak Exclusivity, in addition to the reasons noted in the last subsection.

Subsection 3.4: Phenomenal Holism and the Incompatible Characters Argument

Some authors claim that we really have only a single experience at any one time, and what we distinguish within it are just aspects, not parts (Searle 2000, Tye 2003, Ch.1-3, Raymont 2005). Had my present total experience been qualitatively different in any way, it would be a distinct total experience, and rather than saying that *one* of my experiences would have been different but *others* the same, we should say that my experience would have been a different one that resembled our actual experience in some ways. We can abstract out these resemblances but should not construe them as particular things which would have been present in both cases, and which could be experienced alone by a subject.

If I have only one experience at a time, then partial sharing seems to be impossible: I cannot share simpler experiences with another subject, discrete or merely distinct, for I have no simpler experiences to share. Parts of me like ‘my head’ might still share my *total* experience, but we could not explain my total experience as a complex of simpler experiences shared between me and various parts of my brain.⁹ So we must ask what is supposed to motivate adopting the one-experience account. Some motivations are easy enough for combinationists to accommodate. Some might be attracted by the idea that the whole experience is ontologically prior to its ‘parts’ (e.g. Searle 2000, Raymont 2005); combinationists can retain this idea of priority while accepting the existence of parts (Cf. Bayne 2010, pp.225-249), especially if they are monist combinationists who make component subjects correspondingly dependent on the whole. And Tye alleges that any multi-experience account would be unable to make sense of phenomenal unity without facing a vicious regress from the need to unify the unified whole with its parts (2003, p.21), an objection that can be met by specifying that the unity

⁹ Perhaps combinationists can accommodate the one-experience view by adjusting their description of what is shared. They might claim that aspects themselves, and not just whole experiences, might be shared, so that component subjects ‘have’ just some aspects of the whole’s experience. This grants the claim that experiential elements are mere abstractions, but adds that they are abstractions from a multi-subject state of affairs. Whether this is coherent depends on the exact construal we give to ‘aspects’, and thus depends on what the one-experience view is meant to amount to (Cf. Chudnoff 2013).

relation is reflexive, so that a total conscious state trivially unifies itself with its component experiences (Bayne 2010, pp.29-30).

The most challenging motivation for the one-experience view, however, is what I call the ‘Incompatible Characters’ argument. This argument has been particularly developed in the literature, with Basile (2010) giving the clearest presentation. He offers it as a problem for constitutive panpsychists, supposing that they need “an experience...[to] be felt by two different subjects” (p.109). Sharing is precluded by an argument which he attributes to James, involving two premises:

PHENOMENAL ESSENTIALISM: ... for an experience, to be is to feel a certain way... in the case of experience, ‘appearance’ and ‘reality’ are one and the same.

PHENOMENAL HOLISM: ...within a person’s total psychical whole, the nature of a single identifiable experience... is essentially determined by the other experiences occurring alongside it... within the whole... (p.107)

Phenomenal essentialism implies that experiences cannot be numerically the same while feeling different, but Phenomenal Holism implies that an experience will feel different when unified with different sets of other experiences. Hence an experience cannot simultaneously be part of two different sets of experiences, and thus cannot be shared by two experientially different subjects. The experiences of the parts *as experienced by the parts* cannot be among the experiences of the whole *as experienced by the whole*, because the whole would have to experience them both with and without the changes in phenomenal character which come from being united with each other.¹⁰

The same argument is made by Coleman (2013) who imagines two subjects, named ‘Blue’ and ‘Red’, having experiences respectively “pervaded by a unitary blueness” and “pervaded by a unitary... redness” (p.15), and combining into a composite subject:

¹⁰ Note that this argument rules out partial, but not total, sharing. It allows subjects to share their entire sets of experiences (as Basile recognises, pp.110-111).

To say these points of view were present as components in the experiential perspective of the uber-subject... would therefore be to say that [it] experienced a unitary phenomenal blueness and a unitary phenomenal redness, i.e. had synchronous experiences as of each of these qualities alone, to the exclusion of all others. For it is these qualities each on their own that characterise, respectively, the perspectives of the original duo. Experience excludes, as well as includes. (p.15)

I construe this as a version of the Incompatible Characters argument because Coleman's claim that "experience excludes, as well as includes", and that each part's experience is best captured as a certain quality "to the exclusion of all others", amounts to an assertion of phenomenal holism, emphasising the role of absences. Implicit in the claim that experience e 's phenomenal character depends on the experiences it is unified with is the claim that its phenomenal character would be different, had that set included additional members, and hence that its actual phenomenal character is contingent on the absence of those additional experiences.

Subsection 3.5: Responding to the Incompatible Characters Argument

I believe that the Incompatible Characters argument involve an ambiguity at a crucial point. To see this most clearly, consider the following formulation, which draws heavily from Basile's version but adds a third premise to mark the argument's specific relevance to partial, rather than total, sharing.

Phenomenal Holism (PH): The phenomenal character of an experience depends partly on its phenomenal context, i.e. by the set of other experiences it is unified with.

Phenomenal Essentialism (PE): The phenomenal character of an experience is essential to it.

Different Contexts (DC): If a whole, all of whose experiences are unified, shares only some of its experiences with a certain part, then each experience of the whole will have a different phenomenal context from any experience of that part, i.e. will be unified with a different set of other experiences.

C1: *Therefore* If a single experience were shared by part and whole, it would have a different phenomenal character relative to the part and to the whole (by PH and DC).

C2: *Therefore* If a single experience were shared by part and whole, it would differ from itself in its essential properties, and would thus be, impossibly, numerically distinct from itself (by C1 and PE).¹¹

C3: *Therefore* A single experience cannot be shared (by reductio).

How might combinationists resist this argument? Denying PE is difficult, for as soon as the denier of PE claims that a single experience is ‘experienced differently’, with different phenomenal characters, the suspicion arises that it is these phenomenologically different ‘experiencings’ that we should be talking about, and referring to as ‘experiences’. On the other hand, there is considerable room for denying PH, as Basile himself admits:

[W]hile there are very obvious illustrations of [holism] (a glass of wine has a better taste when enjoyed in a pleasant surrounding), the principle does seem somewhat implausible in other circumstances. Would the red of the book’s cover in front of me have a different feel if I were not hearing music at this moment? (2010, p.110; for further discussion of holism see Gurwitsch 1964, p.120ff, Dainton 2000, pp.181-213, and Chudnoff 2013)

The claim that *every* experience I have is sensitive to *all* my other experiences does seem very strong.¹²

Yet I think combinationists should grant PH for the sake of argument, partly just to maintain neutrality on substantive phenomenological claims, but also because even more limited examples of phenomenal *interdependence*, such as Basile’s glass of wine, or the elements of a Gestalt perception, can support a similar argument. If we replace PH with a more modest claim about some particular experience being altered by particular others, then we can run an otherwise identical argument that those two experiences

¹¹ We might dispense with PE and simply observe that nothing can differ from itself, even in non-essential ways. But without PE it remains open to claim that the experience ‘merely appears different’ to different subjects, i.e. has multiple phenomenal characters relative to each. PE is what ensures that the experience’s *prima facie* self-differing cannot be analysed away.

¹² Dainton (2010, pp.133-139) advances a different argument for holism, arguing that the mere unification of my experiences with each *should be counted as* a feature of their phenomenal character. But this simply makes PH true by definition, and we might simply deny PE for *this* sort of phenomenal character.

cannot be shared with different parts, and this may be enough to rule out certain theoretical applications of sharing (especially for constitutive panpsychists).

Rather than denying PE or PH, I believe the best option for the defender of sharing is to argue that while a form of PH is true, and a form of DC is true, they are true for different senses of ‘phenomenal context’. I earlier defined an experience’s phenomenal context as “which experiences it is unified with”. But there are two ways to read this appeal to unification, which are easy to conflate because if Strong Exclusivity is true, they are equivalent.

On the first reading, ‘unified’ is relative to a subject, so that an experience counts as unified with another only relative to a subject which experiences both together. A single experience shared by multiple subjects may be unified with certain experiences relative to a subject which experiences them all together, but simultaneously not unified with those experiences relative to a subject which does not experience them. Call this the ‘subject-relative’ reading of phenomenal context:

Subject-Relative Phenomenal Context: The phenomenal context of an experience e for a subject s = the set of experiences of s with which e is co-conscious for s .

Alternatively, ‘unified’ could be read as pertaining to how two experiences are related, irrespective of which subject or subjects we consider them relative to. If some subject experiences one but not the other, they still experience the one in the same phenomenal context – in the context of another experience which that subject does not share. Call this the ‘absolute’ reading of phenomenal context:

Absolute Phenomenal Context: The phenomenal context of an experience e = the set of experiences with which e is co-conscious.

It is the subject-relative reading of phenomenal context which is needed for DC. On the absolute reading, the part and whole have experiences with the same phenomenal context – it is just that one

experiences the context and the other does not. Hence the defender of Weak Sharing can argue as follows: the argument from Incompatible Characters is equivocal, because PH is true on the absolute reading, but not on the subject-relative reading. That is, 'experience is holistic' means simply that each experience is phenomenally altered by the other experiences which, in that concrete situation, are unified with it.

This response requires that Phenomenal Holism is true for absolute phenomenal context, but false for subject-relative phenomenal context. Is this plausible? Most discussions do not distinguish the subject-relative and absolute readings, because they assume it makes no sense to consider one experience relative to different subjects. Moreover, the phenomenological observations that are used to support holism work equally for both readings: when an author says 'consider the experience of x – wouldn't it feel different if you experienced it in a different context?', the defender of the subject-relative reading can say 'yes, it would feel different relative to a different subject's total experience', and the defender of the absolute reading can say 'yes, it would feel different if it occurred in a different situation'.

There may nevertheless be theoretical reasons for preferring one reading over the other, in particular the following: the absolute reading requires making sense of the sort of phenomenal character which experiences acquire through being part of a unified whole, inhering in the experiences of a subject who does not experience most of that whole. And this may seem hard to make sense of: what could it be like to have only a few experiences, but for their character to somehow reflect their unification with a whole host of other experiences? How could the sort of infusion and interpenetration we (supposedly) find amongst our various experiences somehow obtain between an experience someone has, and an experience they do not? But this is no longer a deductive argument for impossibility: it is an explanatory challenge, which will be met if the combinationist's account of unity

adequately captures the way that (according to PH) the phenomenal character of each experience reflects that of the others. Thus the incompatible characters argument reduces itself to a constraint on combinationist accounts of conscious unity, the topic of the next chapter.

Subsection 3.6: Responding to Coleman's Argument

Coleman claims that if two subjects composed another, we could not “understand the original two points of view as components of the third”, because that would require the composite to “experience red-to-the-exclusion-of-(blue-and)-all-else [and] blue-to-the-exclusion-of-(red-and)-all-else... both together.”(2013, p.15) Since it is clearly incoherent to have such experiences together, subjects cannot combine. Insofar as this is a version of the Incompatible Characters argument, the response explained above can be extended to cover it. Combinationists can insist that characterising each component subject's experience as ‘something-to-the-exclusion-of-all-else’ is ambiguous, and that when properly understood, we can see that when in the composite, each part's experience is altered by the other's so as to lose its ‘exclusive’ character.

If ‘to the exclusion of’ denotes a phenomenal character, something experiential, something positively present in experience, then combinationists may accept that it is impossible for the whole to experience it in the way described, but go on to say that the parts do not experience it either. Perhaps each does, *when isolated*, experience their redness or blueness as excluding all else, but once they are connected into a unified whole, their experiences are changed, and lose this exclusive character. Just as the whole experiences the red *as* unified with blueness, so does the part: it experiences its red *as* unified with something else, rather than *as* excluding all else. This is what Phenomenal Holism implies, on the absolute reading of ‘phenomenal context’.

On the other hand, the phrase ‘to the exclusion of’ may simply report the negative fact that nothing other than red (or blue) is being experienced. We cannot deny this of the parts – the one is not experiencing anything other than red, the other is not experiencing anything other than blue. But it is not true of the whole, and this is unproblematic, for there is absolutely no reason to expect a whole to inherit *this* sort of property from its parts. It is a trivial logical point that a whole may have some parts which are not X, but nevertheless be X (a house can have parts which are free of asbestos, yet contain asbestos because it has other parts which contain asbestos). And if we read ‘to the exclusion of’ in this way, we are dealing with trivial logical points, not with phenomenology.

Coleman’s argument turns on going from the *logical* claim “this component subject experiences no blue” to the *phenomenological* claim “this component subject experiences something *as* excluding blueness”. This is a substantive inference, which combinationists can and should reject.

In essence, my response to the incompatible characters argument is to point out that the character of an experience for one subject might be altered by its relation to another experience, even though that subject does not have that other experience: there is no need for me to be conscious of everything that affects the phenomenal character of what I am conscious of. The full vindication of this claim, however, must wait until the next chapter, where I discuss at length the relations among unified experiences and the construction of a unified total experience.

Subsection 3.7: The Targets of Sympathetic Imaginability

TSE lets combinationists begin to address the challenge of subsection 1.3: what act of sympathetic imagination is appropriately directed onto a collection of subjects? Combinationists can now say that the experiences that one is to imagine having are simply (some of) those belonging to the parts. So to imagine being all of the parts, we simply imagine being each of them, all at once. That is, we perform

all the imaginative acts we would perform to imagine the parts, but simultaneously as a single act of imagination.

It might be worried at first that this makes no sense, because we will not be performing any single imaginative act, but several. What makes this a single act, and differentiates it from simply performing many acts, imagining being each part, in turn, separately? While the nature of imagination is complex and difficult, we can distinguish at least three stages or moments. There is the ‘simulation’ of whatever state one takes the target to be in, but if this occurred on its own it would not be imagining *them*, but simply entering some mental state. There must also be the right sort of intention - the intention to imagine being that particular target, which confers on the resulting simulation an intentional relation to an external being. And, once this intention has led the mind to simulate the appropriate mental state, there must be something like a judgement that *this* is what it is like to be the target.¹³ Given this three-stage analysis, combinationists can distinguish the imaginative act they prescribe to target the whole from a sequence of imaginative acts targeting the parts: the former involves a single intention and a single judgement, both directed onto the whole as such, while the latter involves multiple intentions and judgements directed onto distinct parts. Thus the claim of combinationists who accepts the inheritance proposal is that to imagine being a composite requires simply imagining having certain of the experiences of the parts (those which meet the conditions specified in CEI), tied together by a single intention and a single judgement.

Section 4: Components of Experiential Ownership

Even if TSE, the token-sharing of experiences, is accepted, it remains to justify CEI, the conditional inheritance of experiential properties. But accepting TSE significantly changes what combinationists

¹³ Of course one or more stages may be brief, or unconscious, or occur in some other order: the point is just that something in the subject’s mental life must accomplish these three things: intending, simulating, and judging.

have to do to make CEI intelligible. Rather than explaining why there are experiences there, they need only explain why, given the presence of certain experiences, the composite qualifies as their subject. Thus their focus shifts from experiential properties *per se* to the relation of ‘having’ that subjects bear to experiences. Note that subjects ‘have’ experiential properties, and ‘have’ experiences, in different senses of ‘have’. They ‘have’ experiential properties in the sense of instantiating them, just like any property is ‘had’ by its bearers. But subjects ‘have’ experiences in a different sense, which I will label ‘experiential ownership’.¹⁴ In this section I analyse this relation to see whether it is such that a composite automatically bears it, given certain conditions, to the experiences belonging to its parts.

It is clearly beyond the scope of this work to definitively establish the nature of experiential properties, so I shall not argue for any particular analysis of the experiential ownership relation. Rather, I will review a number of candidate analyses, while remaining neutral about which of them, or which combination of them, is correct. My aim in this section is to show that, if TSE can be assumed, then many of the available analyses will make CEI an intelligible truth, while also identifying those analyses which do not: for the former the explanatory gap problem, at least as it concerns subject-summing, is resolved, while for the latter it requires the further arguments of section 5 to be resolved.

Subject-first and experience-first combinationists have different perspectives on the experiential ownership relation, reflecting their different views of subjects. For the subject-first combinationist, it makes sense to consider some entity (a human head, a planet, a neurone, etc.), and some experience, and ask whether they are related in the appropriate way. For the experience-first combinationist, however, this will seem an awkward framing, for most ways of picking out a particular subject will refer to its experiences, and so will presuppose facts about ownership. A more natural

¹⁴ There may be further structure here – in particular, experiences may be the event of a subject entering into a certain relation of ‘awareness’ with some sort of object, whether that is an external real thing, an intentional content, or a phenomenal quality. But for my purposes it is enough to distinguish subjects, experiential properties, experiences, and the relations of instantiation and ownership.

framing of an equivalent question would be this: given the existence of some set of experiences, do they constitute a subject which then ‘has’ each of them? What is really in question here is something like ‘collective self-ownership’: the experiential ownership of each experience by a subject constituted by them together. My discussion of the question ‘when does some entity stand in the experiential ownership relation to some experience?’ is meant to also cover the question ‘when do some experiences together constitute something to own them?’

The candidate analyses of ownership that I will review fall under three headings: causal factors, ontological factors, and distinctively experiential factors.

Subsection 4.1: Causal Analyses of the Ownership Relation

The first sort of factor that may be involved is causal: having an experience may require, reduce to, or confer certain patterns of causal powers. Functionalist accounts of experience make this central to their analysis of experiential properties, but even non-functionalists might assign importance to some sort of causal role in their account of the ownership relation.

Consider, for instance, why it seems so natural to ascribe experiences to a human being but not to various large wholes which contain that human being, such as the galaxy. Knowing, for instance, that an experience of enjoyment occurred in my brain while watching a certain film, lets you predict much more about the what my body will do than it does about what the galaxy will do. The fact that I enjoyed the film lets you infer many interesting facts about what I will do – the expressions of pleasure or displeasure I will make during the film, what I will say when asked about it, how likely I am to do things I believe will cause me to see it again, and so on. By contrast, that same fact tells us little about what the galaxy will do – we cannot expect animals worldwide to flock to cinemas, or the rocks of the

solar system's asteroid belt to arrange themselves into "four out of five stars". Perhaps for us to say that something 'has' an experience, that experience must control its behaviour so as to be useful in predicting that behaviour.

Of course, the enjoyment *does* let us predict some of what the galaxy does – namely, the things that it does with one particular part, me. (Just as the tapping of my foot can be counted as behaviour of the whole of me, even though most of my body stays immobile, so my walking back to the cinema to see the film a second time could be counted as one of the many things the galaxy does.) What this shows is that ownership depends not just on control of *some* behaviour, but control of *enough* behaviour – or, to put it another way, control of *overall* behaviour.

There are many approaches we could take to capturing this idea: we might speak of Turing machines and machine-tables, or some richer and less formal notion, perhaps invoking evolutionary concerns of 'proper function'. We need not to venture too deep into the details, and may instead employ some deliberately intuitive and loosely-defined notions.

First, for the experience-first combinationist the relevant causal relations will be those among experiences themselves. Let us say that some set of experiences are 'causally integrated' to the extent that various kinds of information-sensitive interactions go on among them - interactions in which the particular content and character of each experience is important. This would cover the way that, for instance, beliefs and desires jointly serve to produce volitions when and only when the successful execution of those volitions would, were all those beliefs accurate, promote the fulfilment of those desires. It would also cover the way that conscious beliefs and perceptions interact to produce other beliefs whose content follows from theirs, or images whose content is associated with theirs.¹⁵

¹⁵ Of course, no agent is perfectly rational or perfectly efficient, and so the prediction that actions will occur which would fulfil the desires, were the beliefs true, will not hold universally. But if it is true enough of the time, it will be sufficient to warrant the judgement that agency is on display.

Note that these patterns of interaction are holistically defined, in that each individual action, representation, or goal plays the appropriate role only relative to a certain set of others. Adding a hundred new desires to such a structure can potentially bring it about that its original members no longer play their roles. For instance, those desires might entail that actions which previously made sense in light of certain beliefs no longer do so, unless the representations are also changed to compensate.

The experience-first combinationist might say that a set of experiences constitutes a subject, and are therefore owned by that subject, only when they are sufficiently causally integrated relative to each other. But the subject-first combinationist will want a relation which obtains between experiences and underlying material entities, which requires expanding on the above notion. Say that an experience ‘controls’ some entity to the extent that the events going on in that entity are guided by a set of experiences containing that experience, in the kind of goal-driven, information-sensitive way that events in my brain influence other events in my brain or body - that is, in the way characteristic of causally-integrated experiences. For instance, we could first label some events in that entity as ‘sensory’ or as ‘actions’, and then observe the degree to which conscious beliefs, perceptions, desires, and so on can be systematically matched up with these events in rationalising ways. We might then propose that the entity ‘has’ the experience only if that experience sufficiently controls its behaviour.¹⁶

Both integration and control are clearly a matter of degree. Moreover, to say that some experiences are integrated, or that they control an entity, will be a *vague* claim, since there does not seem to be any non-arbitrary threshold: for instance, not all of the events occurring in my body are affected by my experiences, but still it seems true to say that my experiences control my body, while they do not control the galaxy. But in this regard these notions are no worse off than a great many of

¹⁶ This might have a competitive structure (‘the entity is controlled by the experience to a *greater degree than any other entity is*’), or merely a threshold structure (the entity is controlled by the experience to a *sufficiently high degree*’).

the notions we regularly employ. This is no barrier to considering them as possible analyses of experiential ownership, which is important for combinationism because neither notion is heritable.

The example of the galaxy illustrates this well. There are far more events happening in the universe than in me – the events in me are a miniscule fraction of those in it. And so even if most of the events in me are influenced, in the right way, by my enjoyment, it does not follow that most of the events in the universe will be. In the terminology employed in chapter 2, it is a ‘systemic’ property, not an ‘additive’ property, and so not heritable. However, the non-heritability of control and integration is not problematic for combinationists, since they need only defend the *conditional* inheritance of experiential properties, and can easily take control, integration, or some similar causal notion, as one of the conditions for inheritance. That is, precisely because ownership of an experience may require being (e.g.) controlled by it, a whole will not inherit ownership of their parts’ experiences unless those parts are related to their other parts so as to let that experience control the whole; control and integration, along with any other systemic component of ownership, will have to be treated as an experiential bonding relation. Combinationists must, of course, maintain that the relations appealed to are ‘independently intelligible’, and can themselves be adequately explained. But causal properties clearly meet this criterion.

Subsection 4.2: Ontological Analyses of the Ownership Relation

The above discussion of control spoke of influencing the events occurring ‘in’ an entity. But they did not require experiences themselves to occur ‘in’ the entity which they controlled. Hence an entity might be controlled by a set of mental events occurring ‘elsewhere’. For instance, consider a human brain housed in an immobile container, communicating remotely *via* radio signals with its brainless body, which travels around under its control (Cf. Dennett 1981). In this case the body is controlled by certain

mental events which, if we are materialists, we will likely regard as identical to or grounded in certain neural events which are clearly occurring ‘in’ the brain, and thus ‘outside’ the body.¹⁷ Should we say, in this case, that the brainless body experiences anything? We might, but we might not: there is some intuitive plausibility in the idea that an entity cannot be the subject of a mental event unless the event occurs ‘in’ it, and so the brain but not the body is a subject in this case. If so, there is an ontological component to experiential properties, which we could express by saying that to *have* an experience you must *underlie* it.

This is particularly plausible if one thinks of experiences as simply modifications of a subject, or otherwise constitutively dependent upon a subject. This might be the whole account of experiential properties for Cartesian dualists, for whom a subject’s nature is so essentially experiential that *anything* it underlies can only be an experience. But other theorists might also think it significant: for instance, a mind-brain type-identity theorist, who takes experiences to be brain events, might well suppose that to have an experience is to ‘have’ a brain event, i.e. to have a brain ‘in which’ those events happen.

All of this presupposes a subject-first view: the experience-first combinationist, insofar as they make subjects ontologically dependent upon experiences, will probably deny that the subject underlies those experiences. But they have an equally straightforward ontological criterion for ownership available: they can say that to have an experience one must be *partly ontologically dependent on it*. An experience can belong to me only if it is involved in constituting me. Obviously this is a trivially weak criterion, for obviously any subject constituted by some set of experiences will be partly ontologically dependent on them, and thus will satisfy this requirement for ownership.

Suppose we accept the subject-first view, and the importance of ‘underlying’: can we make this notion more precise, and will it be heritable? First, observe that the word ‘in’ suggests an analogy with

¹⁷ Enactivists will likely reject this example, holding that the mental events are not brain-bound; indeed they may well regard control and underlying as inseparable.

a property which is clearly and uncontroversially heritable: spatial containment. If some part of something contains X (at a time), it must be the case that the whole contains X (at that time). Underlying might be simply identified with containment, or more plausibly it might be presented as an ontological analogue to it: rather than containing something spatially, what underlies X ‘contains X ontologically’, in that nothing ‘outside’ that thing is constitutively involved in X . We might express this by saying that the underlying thing is ‘ontologically sufficient’ for X . And just like spatial containment, ontological sufficiency for any X is clearly unconditionally inherited.

Leaving this notion of ontological sufficiency rough and intuitive for now, here are two doubts as to whether it captures our target notion of ‘underlying’. First, perhaps subjects must be ontologically both necessary *and* sufficient, so that something cannot be counted as the subject of an experience if it has any additional parts beyond those necessary to produce that experience. Then, for instance, the universe does not count as underlying my experiences, because most of the universe is unnecessary, given one part, for those experiences.

However, this principle would yield some strange results when applied to human subjects, since it implies that any experience that does not depend on every single part of me cannot be mine. For a start, then, any experiences which my brain underlies cannot be attributed to the whole human being. For another thing, if two experiences are underlain by slightly different brain regions, they cannot be attributed to the same subject. It might be maintained that in fact every single neurone in my brain, or even every peripheral nerve, is strictly necessary for a given experience to occur: without that neurone, it would be a numerically different experience, even if that neurone makes no appreciable difference to its character. But this seems a rather forced claim, since our only basis for identifying as necessary those neurones, and not some other things which also make no appreciable difference to the

experience's character, is a desire to preserve our prior ideas of what should count as underlying the experience. We provide no independent criterion of underlying.¹⁸

Second, subjects might not be ontologically sufficient for their experiences if the identity of an experience depends also on its *objects*. Perhaps a visual experience, for instance, depends not only on its subject but also on the visible items which it presents – without these, there would have been a different token experience (even if a phenomenologically indiscernible hallucination). But if we accept this view of perceptual experiences, we are unlikely to regard underlying as necessary for ownership, unless we think that everything we see thereby becomes part of us.

Consequently, I believe that the ontological component of experiential properties, if any, is best captured (at least on a subject-first approach) by the notion of ontological sufficiency, and sufficiency is heritable: a whole is sufficient for anything that its parts are sufficient for, and hence underlies whatever its parts underlie. Of course, if ownership has components other than underlying, such as the above notion of control, then ontological sufficiency for an experience will be only necessary, not sufficient, for ownership of it.

Subsection 4.3: Distinctively Experiential Analyses of Experiential Ownership

Causal and ontological analyses employ topic-neutral concepts, but perhaps the experiential ownership relation involves something distinctively experiential. If so, then the heritability of experiential properties is not settled by the above reflections about causal and ontological factors.

¹⁸ Consider also the Wada test, in which language-laterality is ascertained by injecting anaesthetic into one carotid artery, so as to 'turn off' one hemisphere while leaving the other unaffected. Here, presumably, the anaesthetised hemisphere is not involved in supporting the experiences that occur in the human being, and the entity composed of the lower brain and the un-anaesthetised hemisphere stands out as the necessary-and-sufficient basis for those experiences. But surely those experiences still belong to the same person.

One way this might be true is through a ‘unity requirement’: all the experiences a subject owns must display ‘the unity of consciousness’, in some sense of that phrase. Bayne defends a position of this sort, recommending that we “think of selves as... entities whose identity is determined by the cognitive architecture underlying a [unified] stream of consciousness.”(2010, p.289) If a subject necessarily has unified consciousness, then to have an experience (i.e. to be a subject) requires that it be unified with all of one’s other experiences. Dainton (2008) also defends an account of the self on which it is individuated by a unified stream of consciousness; in a more equivocal vein, Nagel 1971 argues that the apparent lack of conscious unity in the split-brain syndrome poses a problem for regarding the split-brain patient as a single subject.¹⁹

A second option involves the sort of quasi-epistemic ‘awareness’ we seem to have of our experiences. Intuitively, experiences are essentially the sort of thing which we know in a special, subjective way – we can only know ‘what it is like’ to have an experience if we have ourselves had that experience (or, perhaps, a very similar one). This sort of knowledge is derived from the basic ‘acquaintance’ we have with the experience when we experience it (Cf. Nagel 1976, Chalmers 1996, Siewert 2013, Kriegel 2009). Conscious experiences, we might say, are essentially things that we are conscious *of*, where ‘conscious of’ indicates this distinctive epistemic relation (theorists who deny that we are strictly aware *of* our experiences might place similar importance on the way that an experience confers on a subject awareness of its content: I *have* only those experiences that do this for me).

This sense of awareness is clearly epistemic on some level, but it cannot be sufficient by itself for the kind of introspecting powers which we find in human beings. If awareness is a component of experiential ownership itself, then all conscious things must be capable of it. But plausibly many

¹⁹ The example of the split-brain patient can be supported by the hypothetical case of a multi-headed and hence multi-brained animal, such as Cerberus, a mythical three-headed monster. Bayne argues, regarding Cerberus, that even if (in my terms) all three streams of consciousness exert some control over the whole organism, and even if it underlies all three, still it is more appropriate to regard each stream as belonging to a separate subject, rather than to attribute all three to a single subject, for a single subject cannot have multiple disunified experiences.

conscious animals lack the concepts to introspect as we do. Hence, while awareness of an experience may be necessary for reflecting on it, it cannot be sufficient.

Reductive analyses of awareness and unity can be attempted (see Kriegel 2009 for a particularly developed example of the former), on which they are ultimately nothing more than certain causal, correlational, or informational patterns among physical goings-on. But equally, maybe unity and awareness are among the basic ingredients of reality - or maybe experiential ownership itself is a fundamental primitive. What would follow about the heritability of experiential properties?

A unity requirement on experiential properties will rule out their being unconditionally inherited, for having unified consciousness is a systemic property. If two parts of something are each the subjects of some unified experiences, which are not unified with those of the other part, the whole would inherit both sets, and be the subject of disunified experiences, in violation of the unity requirement. But as with causal factors, this is not a problem for CEI, since conscious unity may be one of the conditions which must be met before the property can be ascribed to the whole, one of the bonding relations.

On the other hand, the heritability of fundamental monadic experiential properties themselves, or the primitive awareness involved in them, is left undetermined by any formal or logical considerations. No straightforward contradiction follows from either affirming or denying inheritance principles for them. Indeed, the same holds if experiential properties are not properly viewed as involving any sort of relation to 'an experience', but as fundamental properties with no further structure to discern. Conceptual analysis would then be powerless to explain their conditional inheritance, even given token-sharing.

Subsection 4.4: Putting it All Together

This investigation of the experiential ownership relation allows us to flesh out the idea of conditional experience inheritance, first introduced thus:

Conditional Experience Inheritance (CEI): A whole has an experiential property whenever one of its parts does and that part is appropriately related to its other parts, simply in virtue of the part having that experiential property and being appropriately related to the other parts.

Now we can understand ‘is appropriately related to the other parts’ as meaning one or more of ‘has only experiences causally integrated with those of the other parts’, ‘has only experiences which jointly control the whole together with those of the other parts’, or ‘has only experiences unified with those of the other parts’. Just as I have remained neutral on the proper analysis of experiential ownership, I will now remain neutral on the proper filling out of the conditions in CEI: if we think that such factors are essential to ownership, that provides a clear and direct rationale for incorporating them.

Is CEI now a sufficiently intelligible claim, i.e. one which provides an adequate explanation of the presence of experiential properties in the whole? We have supposed that the experiences the whole inherits are numerically the same as those of the parts, and that the relations among the parts provide the whole with the necessary overall structure for those experiences to be owned by it. Does this suffice to explain why the whole does in fact own them? That depends on the other components of experiential ownership, those which are not systemic, not a matter of how each part relates to every other part, but are what I called in chapter 2 ‘additive’.

The physicalist combinationist has good prospects here, because for them there is nothing irreducibly experiential in the ownership relation, and so it is natural to think that, if there are any additive components of experiential ownership at all, they will be primarily ontological - having an

experience will just mean underlying it and having it play the right overall role. Since underlying is unconditionally inherited, this provides the physicalist combinationist with an explanation of CEI.

It is not only the physicalist combinationist who can make CEI intelligible based only on what has been said so far. A primitivist combinationist must think that there is something irreducibly experiential in experiential *properties*, but they might locate this entirely in the experience which is had, and not in the ownership relation that subjects bear to it. This would be particularly natural for an experience-first combinationist, who thinks that the basic experiential reality is a vast array of experiences, with larger aggregates of these constituting subjects just when their members are suitably related.

But what about the subject-first primitivist combinationist? In particular, what if they think that experiential ownership, or some component of it, cannot be explained in terms of anything else? They still seem to face the problem of the explanatory gap over subject-summing: no analysis of what it is for something to be conscious, to undergo experience, can reveal why a whole should inherit consciousness from its parts.

But even for this type of combinationist, the presence of an explanatory gap problem depends on what criterion of explanation one adopts. In particular, the *a posteriori* combinationist might endorse CEI as basic law of nature, arguing that its intelligibility can come from the conceptual continuity between the properties ascribed in its consequent and antecedent, or from the sharing of token experiences, or from its being an inheritance principle.

Can the *a posteriori* combinationist claim that CEI is intelligible for its simplicity and generality? Perhaps not, for though it is explanatorily powerful (any number of different experiences in any number of different composites might follow from it) it suffers from a significant degree of complication if it incorporates any sort of causal conditions, such as integration or control. At least,

pending a systematic, unifying account of the so-far rough and intuitive notion of causal integration, it will seem unsatisfying to build it into a fundamental law of nature. On the other hand, the *a posteriori* combinationist who incorporated only conscious unity into the conditions in CEI, and who moreover took unity to be a more or less primitive relation, has a good case for CEI being a simple, and therefore intelligible, law.

Finally, observe the following possibility, which will become important in chapter 5, section 1. Some analyses of experiential ownership leave open the possibility that a set of experiences might satisfy them together but not individually. For instance, a set of experiences might collectively produce the right effects to meet some causal requirement for ownership by a certain subject, or might give that subject knowledge of overall facts about the set, even while none of them individually had the causal powers to do so, or could be known about by that subject. We might then say that they composed a composite experience that belonged to that subject, even though none of them individually did (assuming some meaning can be given to the idea of a composite experience with experiences as parts). Conversely, other analyses of ownership, such as one based on underlying, do not seem to allow for this: I could not be metaphysically sufficient for a composite experience without being metaphysically sufficient for its parts.

This opens up the possibility of what I will call ‘quasi-inheritance’: a whole having an experience in virtue of its parts having, not the property itself, but its parts. Here no token experience would be directly shared: the relationship between the whole’s experience and the parts’ experiences would be composition, not identity. Either of the above two approaches to justifying CEI - by reducing ownership entirely to systemic factors like causal role, or by making it a basic law of nature - can be extended to cases involving quasi-inheritance rather than full inheritance, in any situation where the independently intelligible relations among experiences and subjects are such that the parts’ experiences

satisfy the conditions for inheritance collectively but not individually (whether those conditions come from conceptual analysis or *a posteriori* laws of nature). By contrast, the approaches to justifying CEI discussed in the next subsection do not allow for quasi-inheritance, a fact that will be important in chapter 5, section 1.

Subsection 4.5: Conditional Experience Inheritance and Basic-Experience Inheritance

Here is another approach, that may appeal both to the *a priori* combinationist and to the *a posteriori* combinationist concerned with simplicity, if either is convinced of the irreducibility of (some component of) experiential ownership. They might explain CEI by reference to a mixture of conceptual analysis (explaining the complex causal conditions it imposes by reference to our complex intuitive notion of ‘having an experience’) and the unconditional inheritance of what I will call ‘basic-experiential properties’. Basic-experiential properties, properties of ‘basically-having’ experiences, abstract away from whatever components of experiential ownership are systemic. They involve bearing to an experience whatever in the experiential ownership relation is primitive and additive.²⁰ To emphasise the contrast, we can call any experiential properties that *do* involve a systemic component ‘full-experiential properties’, involving ‘fully having’ an experience.

It bears emphasising that basic-having and full-having are schematic notions, to be filled in according to different views of experiential ownership. Physicalist and primitivist combinationists will thus differ on what they come to - indeed, a functionalist might deny that there is even such a thing as basic-ownership, because once we abstract from (systemic) functional roles there is nothing left to

²⁰ I might have used the term ‘proto-experiential’ (or ‘proto-phenomenal’) for these properties, but that term has already been employed for “properties that are not phenomenal [experiential] but that can collectively constitute phenomenal [experiential] properties” (Chalmers forthcoming-b, p.14). It is not definitional that basic-experiential properties are not experiential properties: if experiential properties do not essentially require any systemic structure or role, then they are themselves basic-experiential. But if experiential properties do involve such a requirement, they are distinct from, but imply, basic-experiential properties, which will themselves count as proto-experiential properties.

experiential ownership. But on the other hand, a primitivist about consciousness might take basic-ownership to be simply the same as experiential ownership, because they deny that any systemic factor is essential to experiential properties.

So we are considering the following unconditional inheritance principle, governing these ‘basic-experiential properties’ that some combinationists might regard as non-existent, and others might regard as simply experiential properties by another name:

Basic-Experience Inheritance (BEI): A whole has a certain basic-experiential property whenever one of its parts does, simply in virtue of the part having that property.

Suppose that basic-ownership is a coherent notion, and is distinct from full-ownership. Then since having an experiential property is simply a matter of having a basic-experiential property involving an experience which is then connected with one’s behaviour, and the other experiences one basically-has, in the necessary ways (if any), BEI would entail CEI. However, BEI faces a number of problems. One is that, since systemic features like control and unity cannot be unconditionally inherited, BEI implies that a being might basically-have an experience but not have it in the normal full sense, i.e. might in some sense enjoy ‘phenomenology’ which was not reflected in its behaviour or fully unified. If this kind of inert and disunified phenomenology is not possible, and if this is not because phenomenology *reduces to* control and unity, then BEI must be false. Moreover, BEI would imply that anything with human beings as parts, like the galaxy or James’s 12-man group, has basic-experiential properties, and to that extent is ‘conscious’, albeit in an inert and disunified sense. If basic-experiential properties are just experiential properties, the same worries arise even more directly, without a need for quotation-marks around ‘phenomenology’ and ‘conscious’.

But quite apart from these problems with the possibilities or actualities that BEI entails, there is still the concern that it does not appear to follow from any discernible fact about the nature of

basic-experiential properties. Why should BEI hold? Why should wholes have this primitive relation to experiences just because their parts do? The *a posteriori* combinationist need not be worried by this: they could claim that it is simply one of the laws of nature in our universe, and deflect charges of emergentism by noting its other virtues. But the *a priori* combinationist cannot rest content with this: BEI is still a claim connecting different mereological levels, and it demands explanation. If it cannot be justified *a priori* then it does not explain the consciousness of composite subjects.

So I conclude this section with the following claim: the explanatory gap remains, even given TSE, and even in light of a systematic analysis of the experiential ownership relation, *if we focus on a particular sort of combinationist*. The combinationist in question is a primitivist, most likely a panpsychist, regards conscious subjects as ontologically independent of their experiences, takes experiential ownership to involve some primitive, irreducibly experiential relation, and uses *a priori* entailment as their standard for intelligible explanation. Though highly specific, this version of combinationism is not marginal: believers in the *a priori* entailment criterion are particularly likely to reject physicalism and posit experiential primitives. However, I believe the explanatory gap challenge can be met even on this particular sort of combinationism, as argued in the next section.

Section 5: Making Basic-Experience Inheritance *A Priori*

Basic-Experience Inheritance is not *a priori* in the way that, say, location inheritance, is evident simply through consideration of the properties involved. Yet it may be *a priori* in a more indirect way, in virtue of a general background view of composition. That is, the *a priori* necessity of BEI may derive from the nature of the part-whole relation rather than the nature of basic-experiential properties. In this section I will argue that if any of the ‘level-connecting’ views of composition is true (i.e. if either parts or wholes ground the other), then all *fundamental* properties will be unconditionally heritable by

default, i.e. heritable unless their inheritance is demonstrably impossible. Since it is only the view that basic-experiential properties are fundamental and additive that faces the explanatory gap problem, a response which shows that precisely such properties are inherited would resolve that problem. More precisely, I will argue for the intelligibility of the following conditional:

Heritability of Fundamental Properties (HF): *If a property is fundamental, and if one of the level-connecting views of composition is true, and if the heritability of that property would not yield demonstrable incoherence, then that property is both-ways inherited.*

When I speak of properties whose heritability would yield demonstrable incoherence, I primarily have in mind systemic properties, which are sensitive to the total set of properties possessed by their bearer. But I also have in mind ‘properties’ whose precise nature depends on their bearer. For instance, the property ‘releases acidic fluids’ is implicitly the property of releasing them *from oneself*, and so will mean different things when ascribed to an organ and to an organism – someone might have an organ which releases acidic fluids, but only into another of their organs, so that they themselves do not release such fluids. There may be other ways for inheritance to yield demonstrable incoherence. The point of HF is to shift the burden of argument: rather than needing reasons to think a given fundamental property heritable, an adherent of one of the level-connecting views needs reasons not to.

HF will only be plausible given certain views about what the fundamental properties are. I take it as definitional that ‘fundamental’ properties are those which ground all other properties; thus it is not definitional that fundamental properties are all and only those which are instantiated by fundamental entities (implying, e.g., that ‘mass’ as a determinable quantity is not fundamental, but only ‘having the mass of one electron’, Cf. Gillet & Rives 2005, Wilson 2012). Moreover, it will be hard to accept HF if we think that the fundamental properties are simply those ascribed by fundamental physics, for properties like spin and quark-flavour do not seem to characterise macroscopic wholes.²¹

²¹ What about properties like mass that sum, or properties like charge that can ‘cancel out’? There seems to be at least a form of quasi-inheritance here, but since a whole with a positively-charged part and a negatively-charged part might

Instead, we should take the fundamental properties to be the conceptually primitive ones, those on a grasp of which our grasp of all other properties depends. This means that the fundamental physical properties are not ‘mass’ and ‘charge’, or some specific determinates thereof, but rather the notions of causal power, spatial location and extension, and time: we grasp the properties ascribed by the equations of physics only because we grasp what it is for some property to govern causal interactions and movements in space. And there is, I think, no antecedent implausibility in thinking that wholes inherit their spatiotemporal and causal powers from their parts - indeed, this seems to be a necessary presupposition for deducing *a priori* the additivity of the properties ascribed by physics (see McQueen ms, 2014a; cf. Chapter 2 subsection 4.2 and 4.3).

I do not have space here to defend this view of fundamentality: my aim is just to illuminate what seems to me an available route to *a priori* combination, and the substantive premises it requires. I will simply try to illuminate why we should think, given a level-connecting view of composition, and given the suggested view of which properties are fundamental, that all fundamental properties are heritable by default.

Subsection 5.1: Motivating the Heritability of Fundamental Properties

The case for HF is most straightforward for one particular level-connecting view, composition-as-identity. Both-ways inheritance would then not go much beyond Leibniz’s Law. If the whole just is its parts together, then any property belonging to the latter must belong also to the former.

itself be electrically neutral, they do not seem to be directly inherited. Yet a whole might be positively-charged just in virtue of one of its parts being so. A defender of HF might be able to accept the fundamentality, and heritability, of charge, by saying that we habitually confuse additive fundamental ‘component charge’ properties with systemic ‘net charge’ properties, which are reducible to component charges in the same way resultant forces are reducible to component forces. A composite with charged parts may have a net charge of 0 if its charges cancel, but does not therefore lack charge-properties altogether.

Of course, this cannot apply straightforwardly to all properties, since not all properties are heritable. This reflects the fact that for composition-as-identity to be defensible, it needs an appropriate semantic apparatus for explaining how, as Lewis says, “It does matter how you slice it – not to the character of what's described... but to the form of the description.”(1991, p.87) Without wishing to wade further into this debate than is necessary, we can observe that for many properties, the conditions for their possession distributively by many parts diverge from the conditions for their possession collectively by those parts, i.e. by the whole *qua* whole.

For instance, composition-as-identity should not entail that parts being spherical entails the whole being spherical – not because the whole is anything other than them, but because it requires different things for sphericity to be instantiated distributively or collectively by the one thing involved. And we can see why by analysing sphericity as ‘occupying all and only points within a certain distance of a certain point’. The ‘only’ makes it a systemic property and so not heritable, and moreover the reference to ‘a certain point’ indicates how the conditions for distributive and collective possession can diverge (in the first case there may be a different centre-point for each part; in the second there is a single centre).

If we are considering a fundamental additive property, however, we cannot scrutinise the analysis of the property to find some element that can be interpreted differently (as with, e.g., the ‘centre’ for sphericity), because there is no such analysis available. If there were some positive argument that we must deny heritability, it would have to show a difference between collective and distributive possession. But without that, why should we believe in such a difference? It would seem an unmotivated and unparsimonious posit, and insofar as we should avoid such posits, I think that

adopting composition-as-identity gives good reason to make heritability our default assumption for fundamental properties.²²

But how is HF to be supported on the other views, which treat parts and wholes as distinct but hold that one is entirely grounded in the other? I think it is harder to construct a really satisfying argument, or provide a really rigorous motivation. But we could give an unsatisfying argument and a rough and sketchy motivation, and hope that together they are sufficient.

The rough and sketchy motivation is similar to that based on composition-as-identity: wholes (or parts, for the monist) are nothing over and above their parts (or wholes), and so anything they have they must get from their parts (or wholes). While there may not be identity here, neither is there independence, and so if a truly fundamental property is instantiated, we cannot ascribe it to one but not to the other. There is, so to speak, *nothing else there* for the fundamental property to belong to, save the parts (or whole), or to put it another way, there are not separate ‘truthmakers’ available for the two things’ instantiating the fundamental property.

In attempting to provide a rigorous argument for HF, we will need to distinguish motivations for upwards heritability and for downwards heritability. First, consider how a pluralist might support *downward* inheritance for some fundamental property F – the principle that if a composite has F, at least one of its proper parts must. According to pluralism, all facts about wholes are grounded in some fact about their parts – so the fact that a whole has F must be grounded in some fact about its parts. But if F is a fundamental property, then its instantiation cannot be grounded in the instantiation of any other property save F. Combining these two constraints, we can conclude that the fact of the whole’s having

²² The link between composition-as-identity and inheritance is noted by Sider (2007), who claims that the independent plausibility of location inheritance provides a reason to accept (a modest version of) composition-as-identity, which explains location inheritance better than its rivals (pp.74-76). Basic-Experience Inheritance would not, in the same way, provide an argument for composition-as-identity, since it lacks the independent plausibility of location inheritance. But support does flow in the opposite direction; composition-as-identity explains HF, and thereby BEI.

F must be grounded in some fact which a) involves the property F, and b) pertains to the parts of that whole. The natural ground would be the fact of some part itself having property F, and so we may conclude that some part has F, just as downward inheritance claims.

A monist can make a precisely analogous argument in support of *upward* inheritance. Facts about parts must be grounded in facts about wholes, and facts about property F can be grounded only in other facts about property F. Hence a part having property F can be grounded only in the whole having property F, from which it follows that the whole has property F, in accordance with upward inheritance (ideas along these lines are attributed to some Stoics by Helle, 2013).

What about supporting downward inheritance for monists, or upward inheritance for pluralists? Here we should recall the condition that nothing positively rules out heritability for property F; this entails that it is at least possible for both wholes and parts to instantiate fundamental property F, even if one or the other of them is non-fundamental.²³ But if instantiation of F by non-fundamental entities is possible, something must still in each instance ground it. By the arguments given above, this will have to be the instantiation of F by one of the corresponding fundamental entities (wholes on monism, parts on pluralism). But plausibly, grounds necessitate whatever they ground, and so the corresponding inheritance principle will be necessarily true: given the property's possession by the fundamental entity, it follows that the non-fundamental entity possesses it, for if that did not follow, there could be no explanation for the non-fundamental entity's possessing that property.²⁴

This argument is somewhat unsatisfying on its own, because it simply notes that, *given* the theory that parts (or wholes) ground wholes (or parts), we should expect that instantiation of

²³ It certainly does not seem to be true in general that only fundamental things can instantiate fundamental properties: even if the mass of a whole is derived from that of its parts, a pluralist need not therefore maintain that the whole does not *really* weigh anything, or that mass is not a fundamental property.

²⁴ These arguments confirm the point made briefly in chapter 1, that *explanation* (a cognitive relation involving one fact making another intelligible) can run in *both* directions, even when *grounding* (a metaphysical relation) runs in only one. A ground both entails what it grounds, and can be inferred from knowledge of what it grounds.

fundamental properties by parts (or wholes) would ground and explain their instantiation by wholes (or parts). This is a little like the way that physical combinationism, to be *a priori*, must incorporate a ‘that’s-all clause’ into its explanatory base: a key part of the explanation for certain facts is the unavailability of any alternative explanation. Establishing in this way that one fact must explain another does not tell us *how* it does so. But for this we should return to the rough and sketchy motivation provided above: wholes and parts are so intimately connected, and so wholly dependent (in whichever direction), that there is no room to ascribe fundamental properties to one independently of the other.

Subsection 5.2: Basic-Experiential Properties Without Unity, Integration, or Control

HF implies heritability for properties unless there is positive reason to deny heritability, and in section 4 I noted two possible reasons to deny heritability for basic-experiential properties. One of these, the threat of ascribing basic-experiential properties to mega-subjects, will be discussed in chapter 6. But the other will be confronted here: given that systemic properties and relations like unity, integration, and control cannot be unconditionally inherited, BEI entails that basic-experiential properties might come apart from them: that there could be something like phenomenology, but inert and disunified, with no tendency to produce any of the effects we are used to associating with it.²⁵ And someone might deny that this is possible: they might maintain that the very nature of basic-experiential properties is that, while they do not definitionally *require* such systemic relational structures (i.e. it is not part of what we mean by speaking of them), they always confer them on any being that possesses them. In virtue of basically-having an experience, that is, we can always and automatically access and employ that experience for various purposes.

²⁵ It might still be that, for there to be experiences, they must play that role for *some* subject, but it is not guaranteed that they play this role for every subject that possesses them.

This objection relates to a question we have already encountered, about the possibility of phenomenal overflow. BEI entails the possibility not only of everyday sorts of cognitive overflow (e.g. hard-to-notice visual experiences) but of a far more radical sort, as exemplified in the case of mega-subjects like the galaxy, who enjoy the core of my conscious phenomenology but are not guided by it in the way I am. So to address the explanatory gap problem, I believe that an *a priori* subject-first primitivist combinationist must also be a pro-overflow combinationist, or as I put it in chapter 2, an ‘inclusionary’ combinationist. I do not see any way to address the explanatory gap problem from the position of an anti-overflow, *a priori*, subject-first, primitivist combinationist.

Fortunately for the *a priori* subject-first primitivist combinationist, there are several reasons for them to be sceptical of the anti-overflow position that phenomenology is neither reducible to nor dissociable from control and unity. For a start, this claim amounts to an *a priori* limitation on the possible forms which consciousness can take, even in entities quite different from us. We might be sceptical of such claims in general, suspecting them of being over-generalisations from our own parochial nature. Second, it involves a somewhat awkward middle road between reducing phenomenology to its overall causal behaviour, and making it independent of its overall causal behaviour. The intuition that there is more to experience than its functions seems to push us further, to saying that this ‘something more’ might occur without those functions.

Third, if the anti-overflow position is right, and yet experience is irreducible to the physical, then consciousness has causal powers which are hard to square with naturalism. Any being which basically-has an experience will immediately find that experience unfailingly connecting and interacting, in deep and sophisticated ways, with *all* its behaviour and other experiences. This kind of power would be in certain respects non-local and unlimited, since it works however many experiences and behaviours there are to connect with, and wherever they are. We might think it implausible that such powers are operative in the natural world. Indeed, we might think that plenty of actual examples –

simple animals, or neurologically damaged or abnormal humans, even ourselves when we are absent-minded or drugged – already show us that unity and behavioural control are fallible and imperfect, and accompany consciousness only ‘under normal conditions’, i.e. under conditions which are normal for creatures like human beings.²⁶

Subsection 5.3: The Argument from Unimaginability

Here is a final argument against BEI: the dissociation of basic-experiential properties from control, integration, and unity is impossible, because combinationism’s own recipe for sympathetic imagination breaks down in such cases. In subsection 3.7, I proposed that to imagine being a composite required simply imagining all the experiences one would imagine if imagining being the parts, at once, tied together by a single intention and a single judgement. But it might be alleged that we cannot do this for disunified composites: to bring many experiences that are in fact not unified, into a single imaginative act, must either collapse into impossibility or else falsify their real nature. But if we cannot sympathetically imagine being a disunified composite, we should not attribute even basic-experiential properties to it, for sympathetic imaginability goes with phenomenology.

Of course, combinationists might deny that final claim, accepting that disunified composites are not suitable targets of sympathetic imagination but maintaining that this does not preclude their having basic-experiential properties. But their position is stronger if they can accept the link between phenomenology and imaginability, saying that we can, in principle, imagine having disunified experiences, though it is so difficult as to be in practice impossible.

²⁶ Bayne (2010, chapters 6-9), considering the split-brain syndrome, epilepsy, hypnosis, schizophrenia, and others, argues that actual cases show only the breakdown of various kinds of functional and representational integration, not breakdowns of phenomenal unity. But to maintain that there is phenomenal unity in these cases we must allow phenomenal unity to come apart from other forms of unity, and so give up on the claim that a single being’s consciousness always displays ‘unity’ in a thicker, more multi-faceted sense. This kind of ‘bare’ phenomenal unity might extend much further than we typically suppose; in chapter 6 I discuss the position that phenomenal unity extends universally in our world, but often without thicker sorts of unity.

Consider Barnett's example once again. Suppose we say that imagining being Descartes and Hobbes is a matter of both imagining being Descartes, and imagining being Hobbes, at the same time, prompted by a single volition, and feeding into single final judgement. What makes this seem implausible is that this act of imagining being two people seems to be not veridical but fantastic, because what we are doing *would be appropriate* if Descartes' and Hobbes' bodies were in fact just two organs of some kind of supra-personal mind-linked entity, with a single unified consciousness, seeing through four eyes and responding with four arms. Then, this act of imagination would be reasonable and successful. But since they are not, it is misguided.

Combinationists can accommodate this point by saying that *in practice*, someone imagining being this composite would most likely do so inaccurately. They would most likely integrate the two sets of experiences they imagined, just as they would normally integrate different experiences they imagined as had by a single human being. This would involve things like noticing comparisons or connections, identifying interactions, being able to draw contrasts, linking desires to affordances, and so on. That is the mistaken imaginative act that the objection is thinking of. We could accurately execute the intention to imagine being both only by scrupulously refraining from integrating Descartes' and Hobbes' imagined experiences. Of course, Descartes' should be integrated with Descartes', and Hobbes' with Hobbes', but no member of either group should be connected with any member of the other group. If the simulations were kept dissociated, the defender of BEI can claim, that would be a successful and accurate imagination of what it is like to be the fusion of Descartes and Hobbes.

In practice this kind of 'quarantining' is difficult, and perhaps impossible, simply because our brains tend to actively connect whatever we are thinking of. Just having the two sets of imagined experiences at once would tend towards connecting them; it has been more evolutionarily useful for us to have a brain that connects things constantly and compulsively, than to have a brain capable of holding two complex notions in mind without them having any interaction. Moreover, the demands of

being a single act of imagination will force them together – they must both be connected in the right way to our intentions and to our judgements, and so there will be a further structural tendency pushing them together. Thus the perspective of a disunified composite is in-practice unimaginable because its structure conflicts with the structure that our brains tend to demand and impose. But it is not unimaginable absolutely or in-principle.

Subsection 5.4: Return of the Microexperiential Super-Zombies

For the *a priori* combinationist to address the subject-summing problem, they must show that failures of their preferred basic principle (CEI or BEI) are inconceivable. In section 1 I introduced the figure of the microexperiential super-zombie to dramatise this issue: what should the *a priori* combinationist say about these creatures? That will depend on how we read the experience ascriptions used to define them. If the super-zombie does not *fully have* any experiences, but its microscopic parts do, then combinationists can allow that it is conceivable. For it might be that the relations of causal control and of unification that would be needed for it to fully have any of the experiences of its parts in fact fail to hold.

But if the super-zombie is a composite which does not even *basically have* any experiences, but whose microscopic parts do, then the *a priori* combinationist should say that such creatures (call them microexperiential super-duper zombies) are inconceivable. Their inconceivability is a special case of the more general inconceivability of a composite lacking the fundamental properties of its parts, which in turn is a special case of the inconceivability of both affirming and denying certain properties of something.

This analysis is most obvious if composition is identity, since then there is literally a single thing(s) under discussion. But a pluralist can get the same result from their premise that the composite

is nothing in addition to, or nothing ‘over and above’, its parts. For then talking of the whole is not talking about anything over and above the parts, and to that extent is a way of indirectly talking of the parts. Conversely, the monist, who thinks that the parts are nothing in addition to, nothing ‘under and below’ the whole, can also rule out super-zombies as instances of both affirming and denying certain properties of something.

An opponent of BEI may be unconvinced. A microexperiential super-duper zombie may still seem to them positively conceivable, especially if they are strongly attracted to the NSS principle. Indeed, they may regard the apparent conceivability of the super-duper zombie, together with the reasoning presented in this section, as providing a good reason to reject the level-connecting views of composition. I think the best that the defender of BEI can do is to present the following two reasons why the microexperiential super-duper zombie may seem conceivable even though it is not.

First, the distinction between full-having and basic-having allows the defender of BEI to identify something which *is* conceivable, and which might easily be confused with the microexperiential super-duper zombie. This parallels the strategy by which many philosophers have claimed that the apparent conceivability of, say, water having the chemical structure H_2N , arises from a conceptual confusion. It seems conceivable that water might have been H_2N because we confuse the essence of water with the features of water by means of which we fixed our reference to it. Consequently, when we conceive of some chemical other than water having all the features which fixed our reference to water, we mistakenly describe that as *water* having a different chemical structure (Kripke 1980, Chalmers 2009). Similarly, microexperiential super-duper-zombies seem conceivable because we fail to distinguish different components of ownership. Consequently, when we conceive of a composite whose parts *fully* have certain experiences, but which itself only *basically* has those

experiences (because they are, for the composite, disunified and inert), we mistakenly describe that simply as ‘a non-conscious composite with conscious parts.’

Second, the defender of BEI can suggest that a tacit commitment to Strong Exclusivity lies behind the apparent conceivability of microexperiential super-duper zombies, and behind the NSS intuition more generally. If Strong Exclusivity is true, then any experiences belonging to the whole zombie would be distinct entities from those belonging to its parts, but if Weak Sharing is true they may simply be the same experiences, shared by another, intimately connected, subject. If the zombie’s experiences are shared with its parts, but we try to conceive of it while retaining Strong Exclusivity, we may wrongly perceive the fact that there need be no *additional* experiences belonging to the whole as the fact that there need be *no experiences at all* belonging to the whole.

The *a priori* combinationist attracted to BEI could claim that these two strategies suffice to cast doubt on the apparent conceivability of microexperiential super-duper zombies, and that the earlier arguments from the level-connecting views suffice to judge them inconceivable. This inconceivability is, however, heavily qualified. First, it is qualified by the need to presuppose one of the level-connecting views of parthood. If that is not what parthood is – if parthood in fact has a different nature, a nature which allows for autonomy between mereological levels, with each having and lacking properties in their own right - then super-duper-zombies are probably conceivable. But a combinationist is likely to already endorse a level-connecting view; my question in this work is whether a generally combinationist approach to reality can be extended to the experiential realm in particular.

The second qualification is that the above reasoning is defeasible, if the metaphysical structure of experience or parthood turns out to involve complexities which explain why this particular fundamental property is not heritable. Thus while the above reasoning can be part of an overall attack

on the combinationist's explanatory gap, it cannot in itself defeat any positive argument against such inheritance.

Summary:

How can facts about one subject - one conscious being, one phenomenal perspective - explain facts about a distinct subject? Surely we can give a full description of one subject's experiences without specifying anyone else's experience? This is the problem of subject-summing, a central component of the explanatory gap that seems to face combinationism.

When we consider cases like James's twelve men standing in a circle, it seems compelling that their experiences do nothing to explain any the group might have, but when we consider other cases, those fitting the schematic description I labelled 'the easy case', it seems more plausible that a composite (like the human body, or human body-and-soul) can 'inherit' the experiences of a part of it as long as those experiences are suitably connected to the composite's overall behaviour and other experiences. Extrapolating from this intuition, I considered the 'inheritance proposal', consisting of two claims, TSE and CEI, respectively asserting the token-sharing and conditional inheritance of experiences.

In section 3 I defended the coherence of token-sharing, arguing that sharing between overlapping subjects can be made compatible with (slightly adjusted versions of) most views of the epistemology and metaphysics of experience. This allowed me, in sections 4 and 5, to focus on the relation of experiential ownership, which connects these shared experiences to each of their subjects. In the end, four distinct responses to the subject-summing problem emerged.

The first holds that, as a matter of conceptual analysis, to own an experience means simply to be causally and ontologically related to it in the right ways, and so once the relevant causal structures are built into the conditions in CEI, it becomes an *a priori* necessity. This approach is open both to thorough-going physicalists, and to non-physicalists who locate the irreducibly non-physical element of experiential properties in the experiences, not their relation to subjects.

The second holds that CEI, though not *a priori*, is still true and explanatory: it is a basic law of nature, which qualifies as explanatory in virtue of its simplicity and elegance, its moving within a homogeneous set of concepts, or the ontological parsimony it gains through conjunction with TSE. To substantiate a claim to simplicity, the conditions in CEI might be restricted to a fundamentally experiential ‘unity’ relation, rather than any complex causal role.

The third and fourth strategies both explain CEI in terms of a more basic principle BEI, the unconditional inheritance of ‘basic-experiential properties’, which are like experiential properties except that the relation they assert between subject and experience need not involve any systemic features, such as control, integration, or unity. One approach would be to take BEI as an *a posteriori* law of nature, preferring it to CEI on the basis of its simplicity and then deriving CEI from it by conceptual analysis of the remaining components of experiential ownership. The other would be to derive BEI from HF, a general principle that fundamental properties are inherited if they logically can be inherited, on the basis that wholes and parts are not fundamentally distinct things. In either case, combinationists commit themselves to the surprising conclusion that phenomenology (or something like it) is inherited not only in easy cases (e.g. from the brain to the human being) but in all cases (e.g. from each member of a twelve-man group to the group, if it exists), entailing both the existence of ‘mega-subjects’ and the possibility of radical phenomenal overflow.

If one of these four approaches can be made to work, combinationists can claim to have addressed the problem of subject-summing. But explaining the existence of a composite subject is not enough, on its own; there are other things to explain, in particular the unity of the experiences of certain subjects. The next chapter is focused on this issue.

Chapter 4: Explaining the Unity of Consciousness

<u>Section 1: The Boundary Problem</u>	p.123
<i>Subsection 1.1: Two Boundary Arguments</i>	p.124
<i>Subsection 1.2: Containment-Boundedness and Epistemic Boundedness</i>	p.127
<i>Subsection 1.3: Rebutting the Second Boundedness Argument</i>	p.129
<i>Subsection 1.5: Returning to the First Boundedness Argument</i>	p.131
<u>Section 2: What is the Unity of Consciousness?</u>	p.133
<i>Subsection 2.1: Dispositional Forms of Unity</i>	p.133
<i>Subsection 2.2: Phenomenal and Representational Unity</i>	p.136
<i>Subsection 2.3: Attention and the Phenomenal Field</i>	p.139
<i>Subsection 2.4: Four Choice-Points for Combinationists</i>	p.142
<u>Section 3: Amodal Perception and Perceptual Adumbration</u>	p.144
<i>Subsection 3.1: Introducing Amodal Perception</i>	p.145
<i>Subsection 3.2: From Amodal Perception to Perceptual Adumbration</i>	p.147
<i>Subsection 3.3: Presence, Informativeness, Salience, and the Possibility of Pure Openness</i>	p.150
<i>Subsection 3.4: How should Perceptual Adumbration be Analysed?</i>	p.153
<u>Section 4: Adumbration as an Explanation of Conscious Unity, Given Transparency</u>	p.157
<i>Subsection 4.1: Adumbration and the OA-Relation</i>	p.158
<i>Subsection 4.2: The OA-Relation and Conscious Unity</i>	p.159
<i>Subsection 4.3: Representational Unity and the Informativeness of the Revealed</i>	p.162
<i>Subsection 4.4: Dispositional Unity, Informativeness, and the Salience of the Concealed</i>	p.164
<i>Subsection 4.5: Phenomenal Unity and Pure Openness</i>	p.165
<i>Subsection 4.6: Transitivity and Reducibility</i>	p.167
<u>Section 5: Adumbration as an Explanation of Phenomenal Unity, Denying Transparency</u>	p.168
<i>Subsection 5.1: Can Experiences Adumbrate Experiences?</i>	p.169
<i>Subsection 5.2: Experiential Adumbration and the EA-relation</i>	p.172
<i>Subsection 5.3: Attention and the Contingent Structure of the Human Mind</i>	p.174
<i>Summary</i>	p.175

Each of us, at each moment, experiences many things, but we do not experience them separately. Rather, we experience them together, which may be captured by saying that our many experiences are linked by a relation of ‘co-consciousness’ or ‘unification’, so that our total experience (or ‘phenomenal field’) displays ‘the unity of consciousness’. Combinationism has to account for this unity: a common concern is that while combinationism may explain the multitude of diverse experiences we have, it cannot explain the intimate connection amongst them. If this were true, an explanatory gap would remain even once all concerns about subject-summing were resolved: a gap between merely having many experiences and having them together. Moreover, if unified consciousness is a precondition for subjecthood, chapter 3’s response to the subject-summing problem will not be complete without an

explanation of unity. Unity would then function as a crucial ‘bonding relation’ that served to forge a single subject out of many component subjects.

Part of what makes it a challenge to explain unity is what I called, in subsection 3.1 of chapter 2, the ‘independent grasp requirement’. We could always postulate a relation that, by definition, entails unity in the whole when it holds between two subjects, but this would give us no idea what this relation is, except in terms of a whole with unified experience. If we are to grasp the relation in a way not parasitic on our grasp of the thing it is meant to explain, we need some idea of what the relation is at the level of the parts, not just at the level of the whole.

But does combinationism’s explanation of unity need to be *distinctively* combinationist? There is already research in both psychology and philosophy on how to explain the unity among different experiences (e.g. Treisman 1980, Shoemaker 2003b, Tye 2003, Bayne 2010), and if some relation is successfully shown by non-combinationists to explain conscious unity (e.g. the synchronisation of neural firing rates in different brain areas), why couldn’t combinationists just adopt that account into their own theory? In particular, if the full explanation of unity involves some form of complex causal interaction among brain regions - which seems very likely - combinationists are not barred from appealing to that sort of causal interaction, since causal relations are ‘topic-neutral’ and thus part of the acceptable explanatory base for any sort of combinationism.

One initial response is that combinationists have a distinctive question to answer: unlike others, they have component subjects to think about, and insofar as those subjects share experiences with the whole, they will have experiences that are unified with other experiences they do not share. Thus there arises the question ‘what is it like to have one’s experiences unified with other experiences one does not have?’ i.e. ‘what is it to be a component subject in a unified composite subject?’ Only combinationists have to answer this question.

But perhaps it is not like anything in particular to be a component subject in a unified composite subject. Perhaps, that is, the relation among component subjects that explains unity in the whole does not make any special difference to their experiences. In this case nothing about component subjects would play an explanatory role in combinationism's account of conscious unity, and then this chapter would be largely unnecessary, since there would be no distinctive problem for combinationism. Call this 'minimalism':

Minimalism About Conscious Unity: Conscious unity among the experiences of a composite subject does not imply anything about the experiences of its component subjects, but is entirely explained by the topic-neutral or physical relations among them.

There are, however, two reasons to resist minimalism. One is specific to primitivist combinationists: if purely physical processes fail to explain consciousness, why should purely physical relations suffice to explain the unity of consciousness? That is, primitivists about consciousness might find physicalist accounts of unity unsatisfying just as they find physicalist accounts of other aspects of consciousness unsatisfying: they would then naturally look for distinctively experiential facts about component subjects and their relations that might provide a better explanation.

The second reason to resist minimalism is not specific to primitivists about consciousness, but can appeal to physicalists too. We have already encountered it in subsection 3.4-3.6 of the last chapter, where we considered Basile and Coleman's versions of the 'incompatible characters' argument: Strong Exclusivity follows from Phenomenal Holism (PH), because experience-sharing would require a single experience to have two different phenomenal characters, reflecting its unity with the different phenomenal fields of the whole and the part. The combinationist response I defended there accepted the key premise PH, the claim that an experience's phenomenal character is dependent on its 'phenomenal context', the other experiences it is unified with. But my response maintained that this premise was true

only for the ‘absolute sense’ of phenomenal context: the other experiences a given experience is unified with irrespective of which of its subjects we consider.

If PH is true, then it cannot be right to look for an explanation of a composite subject’s unified consciousness that has nothing to say about the consciousness of its parts. For those parts will have experiences that are unified with those of other parts, and according to the absolute reading of PH this will change the phenomenal character of these experiences. The character of each component experience will somehow reflect that of all the others, and so each component subject will have an experience that is somehow ‘suffused with’ many other experiences that it is not having. Call this ‘maximalism’:

Maximalism About Conscious Unity: When there is conscious unity among the experiences of a composite subject, the experiences of its component subjects are phenomenally interdependent, and this experiential interdependence is crucial to explaining the unity of the whole’s experience.

Of course, PH might be false, and primitivism with it, and then minimalism might be a perfectly reasonable approach to conscious unity, in which case my defence of combinationism in this chapter succeeds before it has even begun. But many combinationists may be drawn towards maximalism, particularly because, while very strong forms of PH may be counter-intuitive (e.g. the claim that every experience I have makes a difference to the character of every other experience), it is much more plausible that some experiences I have make a difference to some others, that my many unified experiences are not simply there beside each other but interpenetrate to some degree.

So for the sake of defending combinationism as fully as possible, I will assume in this chapter that PH is true at least in that the character of each experience I have is impacted by the character of

some other experience I have. This yields a distinctive challenge: characterising the phenomenology of component subjects whose experiences interpenetrate with experiences they do not themselves have.

In fact, there may be a further challenge, in the form of what has been called ‘the boundary problem’. The core observation here is not the unity among experiences but the limits of this unity, the essential ‘boundedness’ of a subject’s phenomenal field. According to some authors, this boundedness is incompatible with conscious unity extending beyond that field, connecting the experiences of one subject with those of another. But in that case whenever there is a composite subject with unified experiences, we would have to deny the boundedness of each component subject’s phenomenal field, and thereby deny something essential to their being genuine subjects. I must admit that I have never found it very obvious what the boundary problem is supposed to be: everything depends on the meaning of the term ‘bounded’, and it is hard to give this term a meaning that both fits with what proponents of the boundary problem say about it, and substantiates a genuine problem. Nevertheless it needs to be discussed, especially because on some construals it imposes constraints on combinationist accounts of unity that complement the constraints imposed by PH; this discussion occupies section 1. In section 2 I then review the different phenomena which fall under the umbrella term ‘unity of consciousness’. In section 3 I introduce what I call ‘perceptual adumbration’, a phenomenological structure in which we are aware of something being outside our awareness. In the final two sections I argue that perceptual adumbration meets all the desiderata for a combinationist explanation of conscious unity.

Section 1: The Boundary Problem

Two distinct arguments can be placed under this heading, but the first does not threaten combinationism *per se* and the second suffers from a crucial ambiguity in the term ‘bounded’, an

ambiguity which turns it from an argument against the possibility of combinationism into a constraint on acceptable combinationist accounts of phenomenal unity.

Subsection 1.1: Two Boundary Arguments

One version of the boundary problem is simply a worry that combinationist accounts of unity may over-generate, producing conscious unity between experiences which we know to be disunified (such as those of different humans). Here is how Chalmers expresses this worry:

To yield human consciousness, we presumably want [experiential combination] to bond a limited multiplicity of microsubjects associated with the human organism, without bonding these to microsubjects elsewhere. It is not at all easy to see what sort of fundamental microphysical relation has this character. (Chalmers forthcoming-a, p.23)

Why do we ‘presumably’ not wish experiential combination to bond human minds and their parts with other minds? Note that this is not the question ‘why does a given consciousness not include more experiences than it does?’ - that is a question about individuation that I discuss in chapter 6, subsection 3.3. My concern here is a question about unity, about the relations among experiences. So, what is our evidence that our own experiences are not unified with experiences that are not ours? Equivalently perhaps, what is our evidence no larger composite containing a human being (the type of entity I have been calling a ‘mega-subject’) enjoys unified consciousness? One answer is that mega-subjects, observed from the outside, typically do not display the kind of integrated and intelligent behaviour characteristic of beings with unified consciousness; this observation is discussed at more length in chapter 6.

But a second answer, more relevant to this chapter, is that we know that unity relations do not run beyond the limits of a human individual because because our own experience is ‘bounded’. For this

to imply anything about conscious unity with things outside us (such as would yield unity in the consciousness of mega-subjects), we would need an additional premise, to the effect that no set of experiences is 'bounded' when its members are unified with other experiences not belonging to that set. Then we could argue from the boundedness of our own experiences to the lack of unity between them and any other experiences. We would then need only a premise to the effect that for the experiences of a composite subject to be unified, the experiences of its component subjects must be unified with each other, and the conclusion that no mega-subject could have unified consciousness would follow. Call this the first boundedness argument:¹

Human Boundedness (HB): Each normal human has a bounded set of experiences.

Unity Incompatible with Boundedness (UIB): A subject's set of experiences is bounded only if its members are not unified with experiences outside that set.

C1: *Therefore* our experiences are not unified with any experiences that are not ours.

Unity in Part and Whole (UPW): For the experiences of a composite subject to be unified, the experiences of its component subjects must be unified with each other,

C2: *Therefore* we are not component subjects of any subject with a wholly unified experience.

However, once HB and UIB are accepted, they may suggest arguments for a stronger conclusion. For instance, Rosenberg argues as follows:

It can seem that the flow of interaction in the universe is inherently unbounded, and no merely abstract pattern presents a natural condition for containing it... This makes for the possibility of... some kind of cosmic consciousness. Unfortunately, no room exists for the more mundane, middle-level boundaries *necessary* for human consciousness to *exist*... this view

¹ I am uncertain of how far the first and second boundedness arguments I discuss map onto concerns raised in the literature: they may simply be original ideas which I have found worth considering. In particular, Chalmers in correspondence expresses caution about endorsing UIB. But he summarises what he did intend as "if unity were rife, all our experiences would be unified with lots of (most, all) other experiences. But our consciousness is strongly bounded so they're not." For the 'so they're not' to follow, boundedness must imply a failure of unification, and this is what UIB attempts to capture.

[combinationism] *banishes middlelevel individuals from existence*. (2004, p.88, emphasis added)

The concern here is not simply that if whatever relation accounts for our unified consciousness were to also connect us to external things, so that we and they together compose a subject with unified consciousness, then we would not have the kind of experience we do ('bounded'), but that we would not even exist ('banished from existence'). This makes sense only if we human subjects do not just happen to have bounded consciousness, but do so essentially. Tononi provides a theoretical rationale for this: by his 'Exclusion Principle', no part of a conscious being can itself be conscious. The primary motivation he offers for this is the "Phenomenological Axiom" that "experience is exclusive – in that it has definite borders, temporal, and spatial grain." (2012, pp.59-60). On this view, boundedness is essential to conscious subjects as such. We might thus formulate a second boundedness argument:

Essential Boundedness (EB): The set of experiences belonging to any subject is bounded.

Unity Incompatible with Boundedness (UIB): A subject's set of experiences is bounded only if its members are not unified with experiences outside that set.

C3: *Therefore* no experiences of any subject are unified with any experiences that do not belong to that subject.

Unity in Part and Whole (UPW): For the experiences of a composite subject to be unified, the experiences of its component subjects must be unified with each other.

C4: *Therefore* no component subjects can compose a composite subject with unified consciousness.

Unlike the first boundedness argument, this one is a threat to combinationism in general, not just those forms that imply mega-subjects. Combinationists might respond by denying UPW, but doing so is incompatible with the last chapter's defence of TSE, the token-sharing of experiences, especially combined with this chapter's working assumption that phenomenal character is impacted by absolute

phenomenal context (rather than by subject-relative phenomenal context, or not by any form of phenomenal context). Given these prior commitments, combinationists must regard unity relations among a whole's experiences as being absolute across subjects, rather than relative, and as connecting the very same experiences as belong to the parts. Thus combinationists should focus their response on the notion of 'boundedness', which plays a central role in HB, EB, and UIB.

Subsection 1.2: Containment-Boundedness and Epistemic Boundedness

It is surprisingly hard to get clear on what Rosenberg and Tononi actually mean by speaking of experience as 'bounded' or 'exclusive'. For instance, Rosenberg speaks of the fact that "Only some experiences are *part of* my consciousness; most experiences in the world are not" (2004, p.80), and Tononi says that "an experience encompasses what it does, and nothing more" (2012, pp.59-60). This suggests that 'boundedness' means simply excluding some things, so that for a set of experiences to be bounded is simply for it to not contain all the world's experiences (or to inconsistently contain experiences it also excludes). Call this the 'containment' sense of boundedness:

Containment-Boundedness: A set of experiences is containment-bounded just if there are some experiences it does not contain.

But this is an implausible reading, since this sense of boundedness is trivially easy to satisfy - anything except the entire universe, or a self-contradictory object, is containment-bounded, since anything 'contains what it contains, and does not contain what it does not contain.' And by setting such a low bar, this sense of boundedness makes UIB very *implausible*. Why should it follow, from the members of some set being related to other things, that the set must include those other things? This is not the case with other relations: the left half of a table is bounded, for it does not include the right half (among other things), but it is also related to the right half (causally, spatially, electrically, etc.), so as to

compose a whole table. My liver and lungs are both bounded, but closely related to each other. Even overlapping things are bounded in this sense!

Alternatively, we might read ‘boundedness’ as being an essentially epistemic notion - not so much ‘excluding some things’ as ‘knowing that some things are excluded’. This is suggested by Tononi’s appeal to the fact that “No matter how hard I try, I cannot become conscious of what is going on within the modules in my brain that perform language parsing... [and] while I can interact with other people, I have no access to their internal workings.” (2012, p.59)

One epistemic reading would focus on the *inaccessibility* of certain information: my set of experiences is bounded in that it gives me no knowledge about the experiences beyond it. Call this the ‘negative epistemic’ sense of boundedness. But what sort of ‘knowledge’ is meant here? Clearly, after all, our perceptual experiences give us some idea of what other people are thinking and feeling, so our experience cannot be ‘bounded’ in any sense that would rule that out. We might say that boundedness involves a lack of *direct* knowledge, but given the Knowledge-by-Ownership principle discussed in subsection 3.3 of the last chapter (by which we directly know only our own experiences), this would collapse into equivalence with the containment sense of boundedness. In between ‘no knowledge’ and ‘no direct knowledge’ we might try to carve out an intermediate criterion, such as the following:

Negative Epistemic Boundedness: A set of experiences is negatively epistemically bounded just if its members confer on their subject no knowledge of other experiences outside that set which is not dependent on knowledge of something non-experiential.

This criterion allows for my experiences to give me knowledge of my friend’s emotions, and still count as bounded, as long as this knowledge is dependent on some non-experiential knowledge (of their utterances, their facial expression, their habits and so on). What boundedness rules out is knowledge of experiences, in virtue of experiences - which differs from the sort of ‘direct knowledge’ restricted by

doctrines of privacy in that the experience known, and the experience that gives knowledge, may be distinct.

But maybe we should focus not on whether or not a subject can know about particular individual experiences, but on whether they can know general statements about which experiences they do and do not have. In particular, maybe references to ‘including only some experiences’, that seemed to suggest the containment reading, are really about being able to *know* what is included and what is excluded. This gives us a third reading, which we may call the positive epistemic sense of boundedness:

Positive Epistemic Boundedness: A set of experiences is positively epistemically bounded just if any subject who experiences all and only those experiences is thereby enabled to know that they experience all and only those experiences.

The phrase ‘enabled’ is meant to allow that this knowledge might not actually be achieved, and might not in practice be possible, if the subjects lacks the necessary concepts, or has insufficient processing power, or is in an unsupportive environment; this parallels the implicit qualifications on ‘able to know’ and ‘can be known’ used to state the principles of privacy and publicity in in subsection 3.3 of chapter 3.

Subsection 1.3: Rebutting the Second Boundedness Argument

Since the second boundedness requires both EB and UIB to be true, combinationists can escape it by showing that no single sense of boundedness makes both compelling.

On the one hand, as noted, the containment sense makes EB plausible but UIB very implausible, and thus does not get us the conclusion that no component subjects can compose a

composite subject with unified consciousness. To make UIB plausible, on the containment reading, we would need the premise that if my experiences are unified with some other experiences, then I *have* those other experiences: the unity relation confers both its relata on the subject of either. Thus when a human being's experiences are unified with those of external things, the human being comes to share all the experiences of those things, and so if co-consciousness runs universally within some whole, each component subject would become conscious of everything in the whole. And then, having all the same experiences as the whole, there would be no clear reason to consider them distinct subjects. Component minds would 'vanish' by apotheosis. But why should we accept this claim about unity? The only reason I can see is an over-reading of PH: going from the idea that each part of a unified phenomenal field somehow reflects the others, to the idea that anyone experiencing one part of the field thereby automatically experiences all the others. But we need not accept this: one experience can reflect other experiences without conferring them wholesale on those who experience it. The lesson of the second boundedness argument, on this reading of 'containment', is simply a constraint on acceptable accounts of unity: they must not be such that each experience confers on its subject all the experiences it is unified with.

On the other hand, the negative epistemic sense of boundedness makes UIB more plausible, at least if PH is accepted. If two unified experiences each reflect the other in their own phenomenal character, then it seems that the subject of one must have at least some idea of the other, and this might be independent of any knowledge of non-experiential things. However, even if combinationists accepted UIB for epistemic boundedness, they need not accept EB in the corresponding sense. Why should it be essential to a subject that their experiences should be bounded in the epistemic sense? If my experiences were highly informative about other experiences, why should it follow that I do not really have any experiences? Admittedly, combinationists need to elucidate, at least in broad outlines, the phenomenology of this kind of state: what would it be like to have experiences that told us about

other experiences in this manner? But they are already under this burden, given PH, and if the argument of the next three sections is successful, they can discharge it.

Finally, is positive epistemic boundedness incompatible with unity - would conscious unity between one's own experiences and other experiences undermine one's ability to know which was which? I do not see why it would: that X and Y are somehow connected does not usually stop us from distinguishing one from the other.

Even if it positive epistemic boundedness was incompatible with unity, must all subjects have positively epistemically bounded experience - could some subject be unable to ascertain which experiences were its and which were not? It seems a strange possibility, but not obviously impossible, and arguably actual in certain pathological cases like schizophrenia and DID (cf. Lane & Liang 2011). Even if we thought, as exclusionary combinationists might, that it is a conceptual requirement for an experience to be called 'mine' that I be able to know that it is mine, EB for positive epistemic boundedness goes beyond this, claiming that it is necessary for an experience to *not* be mine that I be able to know that it is not mine, and it is not clear why we should accept this. However, the core issues here are not really about unity, but about knowledge, in particular introspective self-knowledge. The questions of whether component subjects in a unified composite would lack knowledge of their own boundaries, and whether this is implausible or absurd, is taken up at length in chapter 7, where I argue that we should deny EB for this sense of boundedness; some subjects are unable to know the boundaries of their own consciousness, though they can know the boundaries of the unified and harmoniously integrated consciousness they are part of.

Subsection 1.4: Returning to the First Boundedness Argument

Combinationists can reject UIB on its containment reading, and reject EB on its epistemic readings. The experiences of component subjects in a unified composite need not be epistemically bounded, and their containment-boundedness is compatible with the whole's unity. But what about HB, the claim of human boundedness that drove the first boundedness argument? Since we have rejected UIB in the containment sense, the argument requires that one of the epistemic readings of HB is true, and this is certainly much more plausible than either epistemic reading of EB. It does seem that the average human being's phenomenal field does not give them the knowledge of any other experiences, except in virtue of knowledge about non-experiential things like spoken words and facial expressions. And it does seem that human beings have some overall grasp of what is and what is not part of their consciousness, though whether this grasp is entirely accurate will depend on whether phenomenal overflow is possible: if there can be experiences of mine that I cannot cognitively access, and if my beliefs about my total set of experiences are based on some form of cognitive access to those experiences (as seems to be almost definitionally true), then those beliefs may be inaccurate in excluding certain 'peripheral' experiences. But insofar as these inaccuracies are likely to be minor, HB seems plausible.²

However, there is ambiguity not only in the term 'bounded', but also in the term 'unified'. The phrase 'unity of consciousness' is used in a range of ways, and not all of those meanings will equally support UIB even in one of the epistemic senses. In particular, if there is such a thing as *bare phenomenal unity*, it might be compatible with a significant degree of negative epistemic boundedness, which might then give an illusory sense of positive epistemic boundedness. This would undermine not just the second but even the first boundedness argument. That would mean that the experienced

² My proposals in chapter 7, section 3 imply that, strictly speaking, HB is false for positive containment boundedness, not because human beings cannot form accurate beliefs about their total set of experiences, but because they cannot *self-identity* relative to their own sophisticated parts, and to that extent cannot *know* that they are the subject whose total set of experiences they are judging about. But this is not because of the unity of their consciousness - it is because they and their parts must share their introspective faculties.

character of our experiences is compatible with their belonging to a mega-subject who (like some interpretations of the split-brain patient) has a phenomenally unified consciousness but weak and uninformative causal connections among many of their experiences, a possibility to which I return in chapter 6, section 2.

This shows the need for a preliminary investigation of what exactly we mean by ‘conscious unity’, so as to be clear about what the combinationist seeks to explain and what kinds of explanation might be appropriate; this investigation occupies the next section.

Section 2: What is the Unity of Consciousness?

Having announced my goal of offering a combinationist explanation of the unity of consciousness, I ought to say more exactly what ‘unity’ is. This is controversial, but by reviewing different phenomena grouped under this heading, I hope to at least let those who disagree over its nature track the import of their disagreements on my subsequent discussion.

The biggest distinction is threefold, between ‘phenomenal unity’, ‘representational unity’, and understandings on which unity is a matter of experiences being *disposed to interact* in certain interesting ways. My first two subsections review these three categories, while my fourth is taken up with the topic of attention, which is importantly related to both sorts of unity. In my fifth subsection I explain how various disagreements about the unity of consciousness will affect the specifics of the combinationist’s explanatory task.

Subsection 2.1: Dispositional Forms of Unity

On some accounts of conscious unity, being unified is like being ‘in touch’: when two people are in touch, there need be nothing they are actually doing at every moment, but their capacity and willingness to communicate constitutes their constantly being in touch. One good example is ‘access-unity’, which connects two experiences when “the conjunction of their contents is available for verbal report, reasoning, and the deliberate control of behavior” (Bayne & Chalmers 2003, p.10). A single mental state is ‘access-conscious’ whenever its content can be used by most or all of the subject’s major faculties, and two states are ‘access-unified’ when they are jointly accessible. But in fact the idea that experiences are unified insofar as they are disposed to interact in particular ways is broader than access-unity specifically.

For example, unified experiences rationalise actions together. Conscious beliefs and conscious desires give reasons to act only when unified: my belief that X will lead to Y, and your desire for Y, will not lead either of us to do X.³ Yet when those beliefs and desires are in the same unified mind they do so, with sufficient reliability that failures are noteworthy. The structure of consciousness somehow ensures this (see Shoemaker 2003b for an analysis on which this tendency is partly definitive of the unity of consciousness).

The relevant sort of interactions will depend on the sorts of experiences involved. Two conscious thoughts might interact by informing and revising each other’s content. Two conscious intentions might interact by mutual adjustment. Two experiences may interact by producing a new experience whose content and properties reflect both. Another sort of interaction, which might be seen as the failed version of another sort, is the generation of dissonance or tension. Two firmly-held but contradictory beliefs, for instance, often interact by producing, not a consistent belief, but a (vague or precise) awareness that there is a problem. As a result of this tendency for contradictory contents to

³ There is a way of speaking on which reasons are external, objective considerations – in that language, what I refer to here are instances of my recognising or having mental access to reasons.

produce tension, a unified mind will tend to have a single consistent worldview, and a single consistent will – conflicts get resolved by the jostling among experiences (Cf. the ‘problem of Inconsistent States’ discussed in chapter 6).⁴

Another example is the way that, as Nagel puts it, “the relations among experiences can be substantially captured in experiences of those relations” (1971, p.408). For instance, if I am aware of red and also of purple, I can usually become aware that they are different, contrasting, colours (the absence of this ability in split-brain patients is part of why Nagel finds them so challenging to the usual individuation of minds).

The sort of interactions in question might perhaps be summarised as experiences interacting in ways that reflect their fine-grained phenomenal characters and contents. They do not merely interact, but interact *qua* conscious states.⁵ Note that dispositional unity is a matter of degree: two experiences which are very prone to interact in the above ways, such as foot-sensations and leg-sensations, could

⁴ Some authors refer to this possession of a consistent total plan or worldview as ‘the unity of consciousness’ (e.g. Shoemaker 2003, Korsgaard 1997, p.176 – also Schechter 2010, who contrasts this kind of ‘coherence unity’ with ‘conscious singularity’, which I would call phenomenal unity); this is not how I use the term, and even if we used the term that way it seems that this sort of overall consistency is a usual consequence of dispositional unity, rather than a fundamental explanatory phenomenon, for it is often imperfect and its occasional small failures seem to be explained by failures of interaction among elements (such as when we believe a contradiction because we never think about some pair of beliefs together, and so never notice their incompatibility).

⁵ I have spoken of token experiences interacting with one another, and some ways of individuating experiences might make this description seem incoherent. If experiences last only for a moment, they cannot last long enough to interact with each other, and if their precise phenomenal character is essential to them, it seems they can never be said to have been changed by each other, but only replaced. But this problem dissolves when we allow for a plurality of ways to individuate experiences. There is clearly some everyday meaning in saying that a certain experience (e.g. a pain) lasted a long time, or saying that it ‘would have been worse’ had circumstances been slightly different. The defender of ‘thin’ experiences (tied to their particular time and exact quality) need not deny or reject this habit of speaking of ‘robust’ experiences: instead, they can simply hold that robust experiences are constituted by certain sequences of thin experiences, with each thin experience in the sequence tied to those immediately before and after it by some combination of causal continuity and qualitative similarity. This is how most everyday objects are constituted, on many accounts of their fundamental ontology (see, e.g., Lewis 1976, Nozick 2003, Unger 1990). We would then talk about experiences interacting, meaning by that these temporally-extended, causally-connected, streams: in terms of thin momentary experiences this would amount to saying that one experience had made a difference to the sort of ‘successor experiences’ that the other had produced.

thus be said to be ‘more unified’ than two which would rarely connect, like a foot-sensation and an abstract plan.

Insofar as dispositional unity is a causal matter, it is the form of unity that most easily fits into the minimalist approach. Combinationists can simply say that the intricate physical-causal structure of the brain explains the overall causal dynamics of human experience, including why certain experiences are able to produce certain effects on other experiences. But maximalists about unity might still think that a purely causal account will leave something out, though they will certainly not dispense with it entirely. For instance, they might demand a distinctively experiential account of what it is for a conscious belief to make an action seem reasonable, but defer to neuroscience and cognitive psychology to explain why *these* beliefs are able to interact in that way with *these* motor experiences.

Subsection 2.2: Representational and Phenomenal Unity

Even apart from my experiences being disposed to interact, they are always in some sense ‘there together’, occurring beside each other in some sort of shared field. Part of this is that they are experienced as representing the same objective world, and often as representing the same object; call this ‘representational unity’ (Cf. Bayne & Chalmers 2003, pp.3-4, Tye 2003, pp.73-75). For instance, my visual experience of the left half of a cup is representationally unified with my visual experience of the right half, as well as with my tactile experience of the parts I am touching, because they are experienced as directed onto a single object.

There is also a basic sense in which, possibly even apart from what they represent, experiences themselves are somehow connected, constituting or occurring within a single ‘field of consciousness’;

this is what has been called ‘phenomenal unity’ (Bayne & Chalmers 2003, Tye 2003, p.11). This most general and most nebulous sense of unity has been given a variety of treatments.

On one analysis, two experiences are phenomenally unified when there is some single thing it is like to have both, so that there is a third experience that ‘subsumes’ them. (Bayne 2010, Ch.2). This subsumption relation is one way of spelling out the idea that experiences are composites of other experiences, particularly through the principle that a subject who experiences the composite experiences automatically experiences the components that it subsumes.⁶

On another analysis of phenomenal unity, it consists in some very basic form of representation unity (e.g. experiential contents being ‘closed under conjunction’, Tye 2003, Hurley 1994, or representing a single spatial world - for some discussions of this proposal see Tye 2003, pp.76-78, Dainton 2004, pp.9-10, Bayne 2010, pp.262-266, and Roelofs 2014b, pp.91-93). This makes the most sense if all conscious states have content, and all conscious contents are in ‘the same format’, e.g. all propositional, all perceptual, all in the ‘language of thought’ (see Bayne 2010, Ch.3 for an argument against conjunctive closure based on denying this).

Both subsumptive and representational accounts of unity violate the independent grasp requirement. Neither gives us a grasp of how two unified experiences are related, except in terms of some overarching whole or conjunction that they (or their contents) both feature in. As Masrour (forthcoming) puts it, Bayne and Tye “are not primitivist[s about] unity... But they are primitivists

⁶ Note that if phenomenal unity involves subsumption, then unity between two subject’s experiences necessitates a third experience which, on the assumption that experiences need subjects, will be experienced by some third subject, who will thereby automatically experience the subsumed experiences of the initial two subjects. Thus unity between distinct subjects will entail experience-sharing.

about one-ness” (forthcoming, p.6), whether the one-ness of a total experience or the one-ness of a total content.⁷

Masrouf offers his own rival account, called ‘the connectivity view’, on which experiences are unified if they are either “connected by an experience of a specific relation” (p.7), or by a chain of such connections. Being connected by an experience of a relation is then defined in terms of a primitive sort of ‘attachment’, which relates an experience of an n -place relation with n experiences of objects, such that the objects are experienced as standing in that relation. The connectivity view is similar to the view I will defend in section 4, but as it stands it does not satisfy the independent grasp requirement. Our grasp on ‘attachment’ is in terms of what it is like to experience things as standing in relations, but this is still an experience we attribute to the composite subject who has *all* of the connected experiences. It does not yet tell us what it is like to be a component subject who has only *one* of these connected experiences.⁸

There are opposing views on the relations among representational unity, phenomenal unity, and dispositional unity. Perhaps phenomenal unity provides the basis for dispositional unity: my experiences can communicate *qua* experiences only because they share a field, so that phenomenal unity is the more basic phenomenon. Or perhaps phenomenal unity is not really any primitive or distinctive structure, but just a sort of vague impression we have of our experiences’ systematic proneness to interact in certain ways, so that dispositional unity is the more basic phenomenon. Perhaps phenomenal unity reduces to some particular form of representational unity, or perhaps it exists at a

⁷ Masrouf draws a distinction between ‘Leibnizian’ and ‘Newtonian’ accounts of unity which is close to my distinction between accounts which do and which do not satisfy the independent grasp requirement. ‘Leibnizian’ accounts ground the unity of a phenomenal field in the relations among its component experiences, while ‘Newtonian’ accounts ground those relations in the experiences sharing a field (the terms evoke an analogy with relativist and absolutist accounts of space).

⁸ As well as differing in this regard, my distinction (meeting or not meeting the independent grasp requirement) differs from Masrouf’s (Leibnizian or Newtonian) in being epistemic rather than metaphysical – I am investigating what is intelligible, he is investigating what grounds what.

different level - representational unity among the contents of experiences, phenomenal unity among the experiences themselves as bearers of that content.

Another question, especially important for evaluating the first boundedness argument's premise HB: since dispositional unity is a matter of degree, how much dispositional unity is guaranteed by phenomenal unity, and how much requires additional conceptual capacities, or contingent neural architecture, to realise? This question is posed especially by the possibility that split-brain patients have a single phenomenally unified consciousness, with two clusters of experiences that are highly dispositionally unified internally but not with each other (see Bayne & Chalmers 2003, pp.18-20, Tye 2003, pp.121-125, Bayne 2010, pp.197-198). If this account of the split-brain case is even possible, then phenomenal unity is compatible with quite radical disunities of function. It might even be compatible with the degree of negative epistemic boundedness found in human minds: this possibility will be taken up in section 2 of chapter 6.

Finally, note that the formal character of phenomenal unity has been explicitly discussed (e.g. Bayne & Chalmers 2003, pp.20-29), and it is generally taken to be symmetric and reflexive, though its transitivity is subject to dispute (Lockwood 1989, Tye 2003, pp.129-132, Dainton 2000, Bayne 2010, pp.36-45).

Subsection 2.3: Attention and the phenomenal field

A particularly important feature of our mind's structure is attention. What I take to be essential to attention is that it privileges some contents over others: attended content is processed faster, more accurately, and in the light of a greater range of considerations, and is more accessible to the mind's systems for outward action, internal planning, and memory formation (See, e.g. Eriksen & St. James

1986, Johnsto & Dark 1986, Corbetta et al. 1990, Treue & Maunsell 1999). Moreover, attention can shift, sometimes voluntarily and sometimes involuntarily, so that something can first be unattended, come ‘into attention’, and then pass ‘out of attention’.⁹

Attention relates to dispositional unity in that the interactions which define dispositional unity are much more likely to occur when one (or both) of the experiences in question is attended. What we are attending to is subject to greater influence from other experiences, and conversely exerts greater influence over them. Attending to, for instance, a conscious belief will greatly increase the likelihood that any conflicting beliefs or perceptions we have will come to mind and be used to revise it; attending to a percept makes us more likely to think of things we could compare with it, conclude from it, or do with it. Indeed, Nagel’s characterisation of conscious unity as allowing relations of experiences to be captured in experiences of relations, quoted above, is immediately followed by the qualification “if [the subject] attends to the matter”(1971, p.407). Moreover, we might count transferring attention as itself one of these interactions: two experiences are more dispositionally unified to the extent that attending to one makes the other more likely to enter attention.

Attention also relates to phenomenal unity through the ‘field’ image. Phenomenal unity is often glossed as experiences *sharing* a ‘phenomenal field’; attention is often glossed as having something at the *centre* of some kind of field. James, for instance, says:

In most of our fields of consciousness there is a core of sensation that is very pronounced. You, for example, now, although you are also thinking and feeling, are getting through your eyes sensations of my face and figure, and through your ears sensations of my voice. The sensations are the centre or focus, the thought and feelings the margin, of your actually present conscious field. (1899, p.18)

⁹ My notion of attention is primarily functional, rather than distinctively phenomenological.

Sebastian Watzl has attempted to define attention specifically in terms of this ‘structuring of consciousness’ (Watzl 2010), while other writers have tried conversely to identify phenomenal consciousness with attention (Prinz 2012). Whether or not one accepts such strong claims, it is plausible that the ‘field’ in the two cases is the same: our natural inclination to think of consciousness as an array laid out before us, and of attention as organising a less-attended periphery around a more-attended focus, reflect different aspects of the same basic impression.

The ‘field’ metaphor suggests some idea of proximity or distance between experiences, and in Roelofs 2014b I suggest a way to spell out this suggestion that seeks to connect attention, phenomenal unity, and dispositional unity. It starts from the idea that dispositional unity is a matter of degree: two experiences are more unified if they are disposed to communicate more readily. We might say that they are ‘closer’ when they are more unified, and ‘further apart’ when less unified. Directions could then be constructed out of the relations among distances.¹⁰

We would then say that the attentional focus is at the centre of the field in the sense that it is closer to the other points on average than any other point is, reflecting the above observation that attending an experience increases its dispositional unity with other experiences. And since transferring attention is itself one of the interactions constituting dispositional unity, the field would be a sort of attentional terrain, through which attention moves quickly and easily over short ‘distances’, and with more difficulty over long ones.¹¹ A combinationist need not adopt this particular way of envisaging

¹⁰ For instance, if distance $A-B = X$, and distance $B-C = Y$, then the dimensions of the space containing A, B, and C will depend on the distance A-C. If it is $X+Y$, we can arrange them on a single line, in the order A-B-C; if it is $X-Y$, or $Y-X$, we can again arrange them on a single line, with A and C on the same side of B. If it is some other value, we can arrange them in a triangle, thereby establishing two dimensions. Further points D,E,F... will either be fitted into this plane based on their relative distances from A, B, and C, or will have a set of distances requiring further dimensions, e.g. four points each at the same distance from all of the others would need to be represented as the corners of a tetrahedron.

¹¹ We need not distinguish the things that ‘occupy’ points in the field from the points which ‘compose’ the field; experiences can be both. When attention shifts, therefore, that is not something moving to the centre of a static field, but the field itself twisting and stretching so as to bring one part of itself into closer connection with the rest.

unified consciousness, but it is useful to have a picture that links phenomenal unity (sharing a field), dispositional unity (distance in the field), and attention (the centre of the field).

Subsection 2.4: Four Choice-Points for Combinationists

In trying to explain conscious unity, combinationists must bear in mind several choices about how to think of this unity. Each choice will impact what their eventual explanation has to look like.

1. Is phenomenal unity a matter of experiential *contents*, or of the non-representational features of experiences?
2. Is phenomenal unity grounded in dispositional unity, or vice versa?
3. Is phenomenal unity *transitive* in structure, or non-transitive?
4. Is conscious unity fully explicable in non-experiential terms, or irreducibly experiential?

The answers to these questions will determine the kind of relation combinationists need to explain unity: whether that relation is in terms of contents or not; whether it is primarily dispositional or primarily categorical; whether it is transitive or not; and whether it is fundamental or reducible to some underlying non-experiential relations.

These choice-points relate to divisions among combinationists we have already observed. For instance, the relationship of dispositional to phenomenal unity will reflect the relationship of access-consciousness to phenomenal consciousness, and thus the possibility of phenomenal overflow. If experiences can be phenomenally conscious without having the suite of causal powers that would make them access-conscious, perhaps they could be phenomenally unified without having the suite of causal connections that would make them access-unified, or otherwise dispositionally unified. Similarly, whether conscious unity is reducible to non-experiential relations will reflect the reducibility of experience in general. Physicalists will naturally suppose experiential unification relations to be

ultimately reducible to non-experiential relations, while property dualists and Russellian monists may (but need not) regard them as basic ingredients of reality.

The most important division for this chapter, however, is the first: is phenomenal unity a matter of representational or non-representational features? This reflects a broader question about how to conceive of consciousness. Many theorists have been attracted to the idea that consciousness is, in general, ‘transparent’ or ‘diaphanous’, in that the only things manifest in experience are the external objects and features which experience represents (Harman 1990, Tye 1992, Dretske 1994, Hellie 2010). Combinationists attracted to this sort of view will naturally look to experienced contents for a unifying relation. Masrour’s connectivity account, mentioned in subsection 2.2, is an example of a view intended to be consistent with transparency (pp.7-8).

Conversely, other theorists (e.g. Block 2003) have defended the view that experiences have manifest non-representational properties (and may sometimes lack representational properties). Also at odds with transparency are those who take all experiences to represent (or better, to ‘present’) themselves, being “phosphorescent, like tropical sea-water, which makes itself visible by the light which it itself emits.” (Ryle 1949, p.159, meant as parody but endorsed “while cancelling his mocking tone” by Strawson 2013, p.9, who attributes this view to “Aristotle, Dignāga, Descartes, Arnauld, Locke, Brentano, Sartre and many others”, p.1). While self-representation is not strictly a non-representational property, it cuts against transparency by making the subject aware of the experience as such (though see Kriegel 2009 for a self-representationalist account which seeks to retain transparency). A combinationist who endorses either of these views will naturally look for a relation which holds between experiences themselves, not their external objects.

Section 3: Amodal Perception and Perceptual Adumbration

Let us recap what has been established so far. Combinationists are looking for a relation among subjects that will unify their experiences; if they accept either primitivism about consciousness or the phenomenal interdependence of unified experiences, then their explanation must involve an account of the phenomenology of these component subjects, the way that their consciousness is changed by their becoming part of a unified composite subject. Moreover, the relation appealed to must meet the independent grasp requirement, and in subsection 3.1 of chapter 2 we saw three ways to do this. First, combinationists might look for some physical or functional relation which intelligibly explains conscious unity, and which can be grasped entirely from a third-person point of view. Second, they might look for a distinctively experiential relation that we ourselves bear to things outside us (this is the ‘outward-looking’ approach). Third, they might take the unity relation among our own experiences, and seek to abstract it from that context (this is the ‘inward-looking’ approach).

I think the third of these is the best to focus on, because its major theoretical commitment - the sharing of experiences between part and whole - has already been established as a major plank of the combinationist scheme I have been suggesting. By contrast, the other two approaches both carry contentious further commitments - the first to reductionism about unity, the second to the existence of mega-subjects. However, the approaches are not mutually exclusive: indeed, the relation I discuss in this section, though I will claim it obtains amongst our experiences, is most easily recognised in the phenomenology of our perceptual relation to external things.

In section 1 of this chapter, and section 3 of the last chapter, I refuted two arguments against combinationism (the incompatible characters argument, and the second boundedness argument), from which emerged two constraints on an adequate account of phenomenal unity. First, it must do justice to the interpenetration among unified experiences, the way that each is reflected in the phenomenal

character of the others. Second, it must do this without implying that the subject of one experiences the other, so as not to violate the containment-boundedness of each subject's consciousness. In section 2 of this chapter I then distinguished some elements of the conscious unity that combinationists hope to explain, and noted some live questions about it.

In this section, I present my candidate for an independently intelligible relation which can explain conscious unity: the phenomenological structure that has come to be called 'amodal perception'. In amodal perception we are aware of something yet also aware of its being somehow concealed from us. While a lot of work has been done on amodal perception (E.g. Michotte 1965, Clarke 1965, Noë 2005, Matey 2013), this work has tended to focus on the relation between the subject and the concealed-yet-perceived object. I believe amodal perception also involves an interesting relation among experienced objects, and by extension among experiences; borrowing a term from some translations of Husserl, I call this relation 'perceptual adumbration'.

Subsection 3.1: Introducing Amodal Perception

It is a pervasive feature of everyday perceptual experience that we experience objects as having features beyond the immediately perceptible, and thus are perceptually aware of those features despite in some sense being unable to perceive them. The term 'amodal perception' reflects the idea that we somehow perceive these features without sensory stimulation and thus not in any sensory modality, though I use the label without any commitment to this claim of non-modality.

The standard examples involve visual occlusion. Consider seeing three-dimensional objects with fronts and backs. At present I am looking at a coffee cup; in a narrow sense, I see only the front of it; the other side is concealed from me. In a broader sense, however, I see the cup itself, a whole with a

front and back. I am in some indirect sense aware of the back of the cup, but simultaneously aware that I am not aware of it in the same sense that I am aware of the front; I perceive it amodally. Consider also the relation between the cup and the table it stands on. I perceive this table as having a broad, brown, unbroken surface – but part of this surface is behind the cup, where I cannot directly see it. Yet I am aware of the surface I do directly see as continuous with the surface I do not, perceiving some sections through vision and other sections ‘amodally’.¹²

Other visual examples might involve conditions of poor visibility, when darkness, distance, or fog prevents us from seeing something clearly. We rarely experience the obscured object as being somehow itself fuzzy or lacking in detail – rather, we experience it as having plenty of detail, which we cannot make out. Insofar as we are aware of this detail as not visible, we could be said to perceive it amodally.¹³ There can also be non-visual examples; for instance, Nanay (2010, p.241) discusses the tactile experience of feeling the handle of a cup as the handle of something with other, unfelt parts. Similarly, we might think of the concealed aspects of a sound or smell as the greater intensity or complexity which any given percept might yield if we moved closer, took our hats off, or sniffed harder.

Different authors have pursued different questions about amodal perception. Some have asked how we can properly account for its phenomenology, and the role this phenomenology plays in our impression of the object’s independence (Husserl 1982, Merleau-Ponty 1962, Kelly 2004). Others have focused on which representational faculty amodal perception involves – perception, cognition, imagination, or something else (Noë 2005, Nanay 2010, Briscoe 2011). In particular, this latter question has been connected to the broader idea, associated with Strawson’s (1974) defence of Kant,

¹² As Nanay (2010, pp.241) points out, similar perceptual phenomena can occur with overlapping shadows, or with partial illumination that reveals only part of an object; in neither case is there strictly occlusion.

¹³ Similar examples are discussed by Merleau-Ponty 1962, pp.302-311, and subsequently by Kelly 2004, who argue that amodal perception is essential to colour and shape constancy, that is, the way that we can see an object as having a constant colour or shape which we only imperfectly sense under given viewing conditions.

that all perception is in some sense infused with imagination. Whatever the proper analysis of amodal perception, the fact that we are familiar with this kind of experience of something as not-fully-experienced provides combinationism with a crucial explanatory resource.

Subsection 3.2: From Amodal Perception to Perceptual Adumbration

Combinationists are looking for a relation to bind together experiences, and amodal perception is usually analysed as a relation between subjects and objects. So my discussion will need to extract, from the amodal perception of objects by subjects, some relevant relation among experiences. The first step, undertaken in this section, is to identify a relation between *objects*; in sections 4 and 5 I will define two relations between experiences based on this.

I will call this relation among objects ‘adumbration’. When we perceive something as adumbrating something else, we are aware of it as intimately connected to, and revealing, that other thing, of which we are aware of not being aware. The word ‘adumbration’, an obscure term for ‘shadowing’ or ‘sketchily indicating’, was suggested to me by certain passages where Husserl uses the term ‘Abschattungen’¹⁴, which several translators translate using the word ‘adumbrations’.¹⁵ Husserl writes that:

What is perceived is given in adumbrations in such a way that *the particular givenness refers to something that is not-given*, as what is not given belonging to the same object. (2001, p.41, my emphasis)

In Husserl’s usage the adumbration is a thing that we experience, rather than a relation among experienced things: as Kelly puts it, “The adumbration of the object that is presented in perception is

¹⁴ Jeff Hilderly has my gratitude for calling these passages to my attention.

¹⁵ This word is consistently used by Kersten (Husserl 1982), Steinbok (Husserl 2001), and Kelly (2004), and sometimes by Findlay (Husserl 1970).

the visible side interpreted as a side of the transcendent object that goes beyond it."(2003, p.119, my emphasis). Thus I use the term slightly differently than Husserl, though hoping to pick out a different element of the same phenomenon. I take no stand on how exactly to interpret Husserl; moreover he endorses certain theses on which I remain neutral (cf. subsection 3.5). So Husserl is an inspiration and source of language, but my meanings do not exactly match his.

As well as having different relations, amodal perception and perceptual adumbration in my sense have different definitional commitments. Amodal perception is sometimes defined in ways orthogonal to my purposes, as the conjunction of perception with an absence of actual sensory stimulation (Nanay 2010, pp.241-242), or even narrowly in terms of occlusion. But this makes it impossible to think about amodal perception in hallucinations, dreams, and so on, though intuitively we can easily experience dream-things as having blocked or concealed features. By contrast, my focus is on the particular sort of phenomenology in which we are not only aware of something but also aware that it is concealed or in some sense not given to us. This sort of experience can occur whether or not we are actually receiving any given sort of sensory input.¹⁶

The principle by which I will flesh out my idea of perceptual adumbration is that when we are amodally aware of something, this is in virtue of something we are modally aware of – what is unseen is nevertheless experienced as present in virtue of its connection with what is seen. So we can move from ‘I am amodally aware of object x in virtue of being modally aware of object y ’ to ‘object y , of which I am modally aware, adumbrates object x , of which I am not modally aware.’ This will appear more clearly through an examination of the three examples given above.

¹⁶ Compare Nieder et al. 2002, which reviews evidence for modal completion phenomena in a range of animals. The criterion employed is that they respond to illusory contours in the same way they respond to real contours; but the question of whether there is any phenomenological difference is unasked.

First, when I am amodally aware of the back of the cup, this is clearly in virtue of being aware of the front of the cup, which is facing me. Indeed, the back of the cup is in a sense seen ‘in’ the front, because I see the front *as* the front *of* something which also has a back. My awareness of the back lies in my perceiving the front as one aspect of something with other aspects. Let us say that the front is, for me, the ‘revealed aspect’, and the back the ‘concealed aspect’. The front adumbrates the back (for me), and while I am aware of the former as ‘given’, I am aware of the latter only in a weaker sense, as ‘not given’.

Similarly, when I perceive the table as having parts which are occluded by the cup, this is because my perception of the surface which is not occluded reveals it *as* just a portion of a single, unbroken surface. Here the revealed and concealed aspects are two portions of a single surface, rather than two surfaces of a single object.

In the case where we amodally perceive the unseen detail of something, or the ways it would show itself under better viewing conditions, the revealed and concealed aspects are not spatially separate, but the broad outlines and fine details of a single object. The rough aspect that is given adumbrates the detailed aspect, by presenting itself as a rough and imperfect view of an object that can be seen in better ways.

In each case there is a revealed aspect and a concealed aspect, experienced as intimately connected in a single object. One aspect is given and the other is not, but in a broader sense we are aware of both. I leave unanalysed the exact meanings of ‘revealed’, ‘concealed’, ‘aspect’, and ‘connected’: they mean whatever they must to accurately describe this sort of experience. Perceptual adumbration is defined as the relation between the revealed aspect and the concealed aspect when one is given to a subject and the other, though not given, is experienced as present in virtue of the first being given. In the language of ‘seeing-as’, one might say: the revealed aspect is seen ‘as’ merely one

aspect of an object with more aspects, and that object is thereby seen ‘as’ having certain concealed aspects.

Subsection 3.3: Presence, Informativeness, Salience, and the Possibility of Pure Openness

Let us sharpen up the notion of adumbration. First, we can contrast it with deferred perception, which Dretske (1994) analyses as “com[ing] to know that k is F by seeing and hearing, not k itself, but h .”(p.263). In deferred perception, there may be no link between the thing directly perceived (h) and the thing learnt about (k), save that one carries information about the other. I might, by perceiving h here and now, learn about k which only existed a hundred years ago, or on another planet. There need not even be any causal interaction between h and k , since they might be joint effects of a common cause. By contrast, perceptual adumbration presents the concealed aspect as in some sense ‘there’, in my immediate environment, in virtue of the connection it is experienced as having to the revealed, as another aspect of the same thing. I not only learn about k , but experience k as somehow present.

I am not sure exactly how to analyse this ‘presence’; it does not seem to be simply a matter of the object’s spatial closeness, perceived importance, or ability to produce intense sensations, for these can all be very low (as when I survey a distant scene from a hilltop) without removing the impression that things have concealed aspects, and that these are present to me. Hence I leave it unanalysed.

However, perceptual adumbration does resemble deferred perception in that one thing tells us about another. Though the concealed aspect is not given to me, I do learn about it, *via* the revealed aspect. I may not be able to see the rear side of my coffee cup, but I can tell a fair bit about its overall size, shape, and location, just from the aspect which is revealed. I can also estimate its likely colour,

patterning, etc. And as with deferred perception, background information is important – an aspect may be more informative to one who knows what to look for, or how to interpret what they see.

Different cases may give more or less information. For example, a brief glimpse through a window at night may reveal a slight movement, experienced as that of some vast creature without yielding any definite information as to its nature or shape – by contrast, when the same vast creature is seen in daylight, it will still have concealed aspects (e.g. its opposite side) but their nature will be much more closely specified by what is then revealed. I will call this the ‘informativeness of the revealed’; some revealed aspects are more informative than others.

This information may be veridical or not: sometimes what is given in my experience misleads me regarding what is not given, as when I perceive a piece of flat stage scenery as having a filled-out rear side. Moreover, what the revealed aspect tells us might not be susceptible of explicit propositional formulation: it might merely allow us to recognise the concealed if it comes to be revealed – to say either ‘yes, that’s what was previously adumbrated’, or ‘no, that surprises me’. For example, when we look through the fog at a building whose fine detail is concealed, we may not be able to put into words what the revealed contours tell us about that fine detail. But we may still have a constrained enough impression of it that when we get closer, we could be either surprised or not at its details, and feel that our previous adumbration had misled us or not.

Next, note that perceptual adumbration can be more or less motivationally salient, in two ways. On the one hand, what is concealed may be important – it may matter for our plans what is on the rear side of that object. The concealed aspect may have this importance in virtue of what the revealed aspect tells us about it, or simply because of how little we know about it (i.e. it may arouse our curiosity). Call this the salience of the concealed. On the other hand, the concealment of the concealed may itself matter – it may be important for our purposes that the concealed is not revealed to us. It may motivate

actions to reveal the concealed (e.g. rotating the object, moving closer), actions to make something presently revealed become concealed (e.g. turning away, positioning things to block it out), or actions to keep the revealed in view, or the concealed out of view. Call this the salience of concealment.

Generally, the salience of concealment depends on two things – the salience of the concealed, and whether the revealed is sufficiently informative. If I can know everything I need to about the concealed just from what the revealed tells me, I need not attend to the fact that it is concealed, or take steps to reveal it. In such cases I may not even notice the fact of concealment, and may later not remember which aspects were revealed and which concealed. For instance, if the contents of a cup matters to me, but the interior is concealed due to my angle of view, I may still know enough just by seeing the logo on the front, and take the cup without noticing whether I had actually seen the contents or not.

So concealed aspects or their concealment may be salient, and revealed aspects may be informative about them. It may be that these two features exhaust the phenomenon of perceptual adumbration: our experience of the concealed aspect is just our partial knowledge of it *via* the revealed aspect, and the motivational salience of this knowledge. Were we to reduce informativeness and salience to zero, so that we neither knew about nor cared about the concealed aspect in any specifiable way, it would vanish from our experience.

Alternatively, there might be something in perceptual adumbration irreducible to informativeness and salience, which remains when those are set to zero. This would be a sort of ‘pure openness’, a sense that what is presently given is continuous with something more, even though this ‘something more’ is left entirely unspecified. We might call this as a phenomenological “etc.” or “...”, a sense that what is revealed does not exhaust what is present.¹⁷ Informativeness and salience might

¹⁷ Note that this is primarily an object-representing, rather than propositional, sort of awareness: being aware of ‘things’ as continuous with ‘something else’ need not be the same as being unsure whether P, for any particular P.

simply add specificity to this pure openness, without constituting or explaining it. Merleau-Ponty seems to suggest something along these lines when he writes:

Suppose we construct, by the use of optics and geometry, that bit of the world which can at any moment throw its image on our retina. Everything outside its perimeter, since it does not reflect upon any sensitive area, no more affects our vision than does light falling on our closed eyes. We ought, then, to perceive a segment of the world precisely delimited, surrounded by a zone of blackness... The fact is that experience offers nothing like this... The region surrounding the visual field is not easy to describe, but what is certain is that it is neither black nor grey. There occurs here *an indeterminate vision, a vision of something or other*. (Merleau-Ponty, 1962, p.4, emphasis added)

Neither denying nor affirming pure openness seems clearly correct, so I leave both options open.

Subsection 3.4: How should Perceptual Adumbration be Analysed?

So perceptual adumbration, by definition, involves two or more aspects, some revealed and others concealed, which are experienced as connected to each other in some intimate fashion. This connection ensures that when the revealed aspect is perceptually present to us, the concealed aspect is also present (or as Husserl says, ‘co-present’), even though not ‘given’. The revealed aspect may be more or less informative about the concealed aspect, and the concealed aspect, or its concealment, may be more or less salient. Beyond this, there are several questions philosophers might ask about how to analyse perceptual adumbration. In this subsection I will note some, though I remain neutral on each point.

First, I have spoken of ‘aspects’, revealed and concealed, without saying much about their ontology; this is a deliberate attempt to avoid taking a stance on the ontology of perceptual objects. On a ‘direct realist’ account of perception, we should identify ‘aspects’ with something in the external

world – a set of properties, parts, or powers of a real object. Conversely, on an ‘indirect realist’ account of perception, ‘aspects’ might be mental things.

A second question is whether perceptual adumbration is properly counted as ‘perceptual’ at all, or as a form of rapid and automatic ‘inference’, or an exercise of sensory imagination. This will depend largely on what kind of content perception can have (Cf. Siegel 2006, Hellie forthcoming-b), and on how one draws the distinction between perception and imagination (Cf. Kind 2001, pp.89-95); if perception were defined in a way that excludes perceptual adumbration then all my talk of ‘seeing’ and ‘perceiving’ what is concealed should be interpreted as a loose way of speaking of ‘representing’, ‘intending’, or ‘being aware of’.

A third question is whether the content of perceptual adumbration is conceptual or not. I do not think that this affects the uses I make of the concept later in this chapter, except insofar as it relates to a fourth question: who is capable of it? Does perceptual adumbration require the suite of special cognitive capacities which adult human beings tend to have, or is it something which other, ‘simpler’, minds may have?

On the one hand, it seems overwhelmingly plausible that many non-human animals, and pre-linguistic infants, experience perceptual adumbration. First, such creatures are capable of exploratory behaviours, approaching, circling, or manipulating objects so as to reveal more of their aspects to perception; second, they display evidence of modal completion, which may be neurologically akin to amodal completion (Nieder et al. 2002); third, they can sometimes pass tests of ‘object permanence’, i.e. give indications of understanding that objects continue to exist when unperceived (Bower 1974, Baillargeon 1991, Miller 2009, Doré 1984). While there is controversy over exactly how much the available data show, it does not appear that object permanence is restricted to highly conceptual language-using creatures like adult humans.

We cannot be so confident about simpler minds like new-born babies or shrimp (though Prete 2004 collects some supportive evidence from predatory invertebrates). An observed failure to grasp object permanence might result from the absence of perceptual adumbration, but might equally result from a lack of memory, reliability, etc. in the deployment of perceptual adumbration. Perceptual adumbration with very low informativeness might not produce any specific pattern of behavioural capacities, leaving no empirical way to establish its absence in a given creature. In particular, if one accepts the possibility of pure openness, one might attribute perceptual adumbration to absolutely all subjects, however simple. On this view, perceptual adumbration does not require any particular sort of sophistication: it simply means that one's experiences are 'open-ended', bearing a sense of their own failure to exhaust all that there is.

A fifth question: is adumbration a universal feature of perception, for those subjects who experience it at all? That is, do subjects like us experience adumbration in absolutely all our perceptions, or only in some? Certainly, the perception of occlusion is widespread in vision, but it might not be strictly omnipresent, and it might not be similarly pervasive for the other senses. While I do not see any clear way to decide this issue, Husserl and Merleau-Ponty seem to regard adumbration as universal, Husserl saying that:

every perception... points to... multifarious continua of possible new perceptions... in which the same object would show itself from ever new sides. (2001, p.41, emphasis added)

Sixthly, can there be non-perceptual adumbration, i.e. adumbration among objects not given to sense? There are some promising examples, but they are less compelling, and may not involve a genuinely similar phenomenon. In particular, the 'presence' that distinguishes adumbration from deferred perception must mean something different, since here nothing is perceptually present.

One interesting example is the flow of time. Supposing that it makes sense to say that we are aware of moments of time, it is arguable that they are experienced not as entirely new and separate from the last, but as elaborating on and continuous with it. One way to think of this would be as each moment adumbrating the next – each moment is experienced as merely the revealed aspect of something which includes the next moment as concealed.¹⁸ Thus each moment comes to us not as something so far unperceived, but as something already roughly grasped but not given. I find this idea attractive, but developing it would require wading into the topic of time-consciousness, which is both deeply perplexing and largely outside my present focus.

So I have left numerous questions open, including:

1. Is there such a thing as pure openness, i.e. adumbration without specific informativeness or salience?
2. What *is* an aspect?
3. Is perceptual adumbration genuinely *perceptual*?
4. Is the content of perceptual adumbration propositional and/or conceptual?
5. Is adumbration universal across subjects, or restricted to higher animals?
6. Is adumbration universal across a given subject's perceptions?
7. Can adumbration occur among non-perceptual objects, like texts or moments of time?¹⁹

But all I need insist upon is that perceptual adumbration is a genuine and widespread feature of our experience. In the next two sections, I will show how a combinationist might employ our grasp of this familiar feature to explain the unity of consciousness.

¹⁸ Since that moment itself adumbrates the next, it is unclear what the relevant whole would be, if not the entire stream of time itself.

¹⁹ Note that positive answers to questions 1, 5, and 6 are likely to be mutually supporting; the possibility of pure openness goes naturally with a more expansive view of the scope and extent of adumbration.

Section 4: Adumbration as an Explanation of Conscious Unity, Given Transparency

If unified experiences are phenomenally interdependent, then component subjects in a unified composite must be aware of their own experiences, but also of how those experiences form an intimately-connected whole, other parts of which they do not experience directly. In perceptual adumbration we are aware of things as aspects of a continuous whole, other parts of which are not given to us. As such, perceptual adumbration is a good candidate for the relation combinationists need to explain conscious unity.

However, as it stands, perceptual adumbration is not a relation between subjects but between objects perceived by one subject. So more definitional work is needed, and this work will be much briefer and more direct if the transparency thesis is true, so that facts about an experience's contents are the only phenomenologically manifest facts about it - that is, all that the subject is aware of is the non-experiential world presented by the experience. Thus in this section I suppose transparency to be true – that is, I suppose that all manifest features of experience, including phenomenal unity, can be accounted for in terms of content. In the next section I relax that assumption and consider how to explain conscious unity apart from content.

In this section, therefore, I am assuming a certain answer to the first of the four questions in subsection 2.4. In subsection 4.1 I define a relation among experiences, and by extension among subjects, and in subsection 4.2 I argue that this relation provides an intelligible explanation of conscious unity, if transparency is true. The following subsections flesh out this proposal by drawing on more specific notions laid out in section 3, and I close by showing how different variants of this proposal can accommodate different ways of answering the remaining three questions of subsection 2.4.

Subsection 4.1: Adumbration and the OA-Relation

Suppose that two experiences represent objects which adumbrate each other. That is, each experience has an object, which is experienced as merely the revealed aspect of an object whose other aspects include that which is revealed to the other experience. This establishes a sort of ‘sideways representation’ of each experience by the other, in that each is aware of the other’s content (its only phenomenologically manifest feature) as something not given. This need not involve any representation of the other experience *as* an experience, but merely as the actual revealedness of an aspect of the object. Call the relation between these two experiences ‘mutual object-adumbration’.

Mutual object-adumbration is still not suitable for our needs. For it leaves open that two experiences might have the relevant content without interacting, just by coincidence. This will be true, for instance, whenever two people look at opposite sides of the same object: the object of one subject’s experience is side 1, which adumbrates side 2, which is the object of the other subject’s experience. This seems much too easy a relation to yield conscious unity.

Let us say that two experiences stand in ‘the OA-relation’ if their contents mutually, veridically, adumbrate each other, each doing so *in virtue of* the other doing so. That is, each experience presents something as revealed, and what it presents adumbrates what the other presents as revealed, and does so *because* the other experience presents that thing as revealed (this ‘because’ might be causal, constitutive, or any other sort of ‘because’).

Crucially, we can define an extended sense of the OA-relation between two subjects, which obtains when at least one experience of one stands in the OA-Relation to at least one experience of the other. Thus relations among experiences imply relations among subjects, though the mutual

dependence of the experiences is important: we do not have the OA-relation when two people look at opposite sides of the same object, for each is aware of the far side as concealed because of their own visual and cognitive systems, not because that side is seen by the other person.

Note that the OA-relation satisfies the two constraints arising from my earlier responses to the boundary problem and the problem of incompatible characters: it makes the phenomenal character of each experience reflect that of the other, but does not confer both experiences on the subject of either. It is compatible with both phenomenal holism and containment-boundedness. The same goes for the EA-relation, which will be defined in section 5.

Subsection 4.2: The OA-Relation and Conscious Unity

Having defined the OA-Relation, how might we construct a combinationist analysis of conscious unity in terms of it? The first step is:

Unification-by-Object-Adumbration (UOA): A subject has a unified experience of X-and-Y-together whenever they experience X, and experience Y, and these experiences are OA-related.

This says that conscious unity among a subject's experiences is nothing but their standing in the OA-Relation. That is, if we imagine having two experiences, each containing in itself an implicit awareness of the other's content as something beyond its own but intimately connected – each indicating the other as an aspect of the same whole as itself – what we are imagining is simply conscious unity. This aims to satisfy the independent grasp requirement in an 'inward-looking' way, because our basis for accepting UOA is 'introspection', or more broadly our acquaintance with what it is like for us. UOA says that the relation of conscious unity present within our experience is nothing

over and above the OA-relation, something of which we have an independent grasp via the notion of perceptual adumbration.

It might seem contradictory for two of a single subject's experiences to be OA-related. How can my awareness of one object involve 'awareness of not being aware' of another, if in fact I *am* aware of both? But while there is a negative component to the definition of perceptual adumbration (awareness of something's *not* being given), we must be careful how we characterise this. We might gloss it as saying:

- A. "This is not given at all, to anyone."
- B. "This is not given to me."
- C. "This is not given to me 'through', 'in', or 'by' this particular experience."²⁰

Of these, I think C. is strongly preferable, because it does seem that we can experience adumbration while simultaneously experiencing the concealed aspect as given through another experience, and such cases do not give us any sense of contradiction or impossibility. For instance, I may see a three-dimensional object as having a concealed rear side, while also seeing that rear side directly in a mirror placed behind it. Hence I conclude that the negative moment of perceptual adumbration is best captured as "this aspect is not given through *this* experience (which adumbrates it)". Given this, a subject whose experiences are OA-related need not have a sense that they both are and are not being given each aspect, but merely that each aspect is given in one experience, and adumbrated but not given in the others.

Does UOA satisfy chapter 2's criteria for intelligible explanation? I think it is clearly compatible with the sharing-inheritance-connections schema, in which it functions as a 'connection principle'. And it is general in application: it does not apply only to some specific type of composite,

²⁰ Alternatively, one might deny that there is any such distinctive negative moment, insisting instead that the content "X is concealed" is entirely first-order, and should not be framed as a second-order content which mentions mental acts like perception. This demands some explanation of what 'concealed' is supposed to mean, if it is not cashed out in terms of possible but non-actual mental acts, and it is unclear whether that can be done.

but to any subject with OA-related experiences. But there are two more difficult questions: first, is the OA-relation ‘conceptually continuous with’ conscious unity - is it the right sort of thing to explain it? and second, does it follow *a priori* from some subject having two OA-related experiences, that they experience those two together, as unified? These questions can be answered only by phenomenological reflection, but I think UOA is as plausible as substantive phenomenological claims ever are. Each experience of such a subject carries a sense of the others without actually giving them: it ‘connotes’ or ‘indicates’ the others, while internally announcing its own failure to capture their full richness.

Note that since this account assumes transparency, it assumes that the unity of my consciousness is reflected in some kind of represented relation among the contents of my consciousness. Conscious unity is not something connecting ‘experiences’, but the web of experienced connections among the sights and sounds and smells that experience presents to us. This brings the account close to Masrou’s ‘Connectivity View’, on which experiences are unified whenever they are linked by a primitive ‘attachment’ relation to an experience of a relation among their objects. The difference is just that UOA satisfies the independent grasp requirement, by saying how the unity of my many experiences together is reflected in the content of each experience individually. Consequently this account, unlike Masrou’s, can explain the unity of a composite subject’s experience in terms of the experiences of its parts. From UOA and the principle that experiences, including their adumbrational character, are inherited by wholes from their parts, we can derive the following:

Unified Composites from OA-Related Parts (UCOP): A composite subject enjoys conscious unity whenever its component subjects are OA-Related.

This gives us a unified composite subject just from having each component subject’s experienced contents OA-related to those of the others: each is aware of what is given to the others, but as not-given – as concealed aspects of the unified whole which their own experience partially reveals. The whole system is conscious of a single experienced world, precisely because each part is conscious of one

aspect of that world *as* unified with (‘continuous with’, ‘intimately connected with’, ‘revealed alongside’) other aspects which are not given to that part.

Note that UCOP says nothing about what happens in the whole when component subjects are related in ways that fall short of the OA-Relation – when they are isolated and insensitive to each other, or when one’s content adumbrates another’s but not vice versa. It leaves open that such cases give rise to no unified state of the whole, or to exotic and unfamiliar sorts of consciousness.

Both UOA and UCOP speak indifferently of ‘conscious unity’ and ‘adumbration’. In section 2 I distinguished several dimensions of unity, and in section 3 I distinguished several dimensions of adumbration. In the next three subsections I will relate the latter to the former.

Subsection 4.3: Representational Unity and the Informativeness of the Revealed

Representational unity comes in many forms, depending on what relation is represented. These forms can be rich and complex or rudimentary. Combinationists might explain this by appealing to the informativeness of the revealed. Each content adumbrates the other, and insofar as they do so informatively, they represent the other as being related to themselves in some specific manner:

Complex Contents from Informative Adumbration (CCIA): A subject’s experiences jointly represent complex contents whenever, and in proportion as, they informatively adumbrate each other’s contents.

For instance, to have an experience representing P-or-Q might involve having one experience representing P, and another representing Q, each of which adumbrates the other’s content as ‘an alternative’. The experience representing P need not be so informative as to indicate *what* the other represents; it might only demonstratively refer to it as ‘that alternative to P’, just as my visual experience might not indicate what the rear side of the coffee cup looks like, but only that it is ‘the

opposite side of what I directly see'. An experience can be informative by indicating only how what it adumbrates relates to what is revealed.

Similarly, perhaps conjunction is a matter of adumbrating another experience's content as 'also true', 'true along with' one's own, without specifying what it is. And if conjunctive closure is a mark of conscious unity, this might be explained by saying that adumbrating something as 'also true' is the basic and minimal way to adumbrate it informatively.

Which forms of informative adumbration, among which simpler contents, correspond to which representational relations, is obviously a huge topic. It is not clear even what format to think of these representations in – pictorial, propositional, linguistic, or something else. So I cannot give even the beginnings of a full compositional explanation for complex contents. The key thing is that adumbration allows a subject to incorporate reference to another's content into their own, without fully grasping or deploying that content themselves. We might use the metaphor of citation: when one text cites another, its content is enriched without what does that enriching being actually in the text. Many authors citing each other may lack the skills or even the concepts to understand what the others do, but can still jointly create an intellectual edifice richer than any could create alone.

Presumably, rich representational unity, and the informative adumbration which underlies it, correlates with physically-detectable information transfers. Thus in highly-integrated systems like the human brain, where huge quantities of information are rapidly shared and integrated, each part would not only be conscious of the other parts' contents as concealed aspects, but would have a fairly rich impression of those contents. This would not rival the awareness each part has of its own workings, but might be far more than they would have in a less integrated system, in which the components were relatively isolated from or insensitive to each other.

Subsection 4.4: Dispositional Unity, Informativeness, and the Salience of the Concealed

One major form of dispositional unity is that two simple contents can, if experienced together (and especially if attended) generate a new content which reflects the contents of both. This could be put by saying that the content of each experience is available to the other. Combinationists can explain this as simply a disposition to adumbrate the other in the kind of informative way discussed above, in virtue of a generalised causal sensitivity to each other.

But another form of dispositional unity is agential: my beliefs interact with my desires to give me reasons, and my reasons interact with each other and my motor systems to generate actions. These sorts of interactions do not generally characterise beliefs, desires, reasons, and motor systems spread across different unified minds. The combinationist explanation of this must go beyond informativeness and appeal to the motivational salience of the concealed and of its concealment. I suggest that unified experiences in beings like us represent the other contents they adumbrate as extremely salient, but do not represent their concealment itself as at all salient, and so generally ignore it. Correspondingly, each component subject in a unified composite perceives things going on in the others as extremely relevant to its actions, without caring or even recognising that they are concealed from it.

Agential Integration from Salient Concealed and Non-Salient Concealment (SCNC): A subject's experiences are disposed to guide its behaviour in an integrated way whenever, and in proportion as, they adumbrate each other's contents as highly salient, while finding no salience in their concealment.

On this picture, what motivates each part of me is its sketchy second-hand awareness of contents in another part. For instance, suppose I decide that praying is good, *because* it is an expression of piety and piety is good; a subject-first combinationist might say that this reasoning takes place in one part of my brain, P1. Another part of my brain, P2 (closely connected to my cerebellum and muscles) knows how to pray – it has a motor schema coded as 'praying'. My decision motivates me to pray; how this

happens is that P2, receiving information from P1 in virtue of which it constantly adumbrates “the stuff going on over there” in greater or lesser detail, comes to adumbrate P1’s contents specifically as “praying is good”. It may well not adumbrate in any further detail – it may have no idea *why* praying is good, nor care, taking that conclusion as sufficient reason to implement the ‘praying’ schema. We might compare P2, in different respects, to an underling who receives and acts upon instructions without thinking to ask why they were sent, or to a human being who acts on a gut feeling without thinking to analyse it.

Similarly, P1 is set up so that its response to judging “praying is good” is simply to convey that information outwards, making it accessible to other parts like P2. Since this then leads to praying, P1 will likely feel that it has successfully prayed, and done so on the basis of its own reasons, at least if it is capable of receiving and interpreting the relevant feedback. We might compare P1 to a manager who decides ‘to do X’, sends a memo to that effect, and gives no thought to how X will actually be implemented, merely checking that it has been done and then thinking “good, I have done X”. We might also compare P1 to a whole human being who forms an intention, receives feedback from its being carried out by their muscles, and does not enquire as to the causal chain connecting the two.

SCNC implies that component subjects in a unified mind both do and do not ‘perceive each other’. They do perceive each other in the sense that each is aware of the experiences going on in the others, but they do not perceive each other in the sense that they do not recognise these experiences as belonging to *other* subjects. Fuller exploration of this topic, however, must wait until chapter 7, when I will directly confront the issue of self-awareness and other-awareness.

Subsection 4.5: Phenomenal Unity and Pure Openness

CCIA and SCNC together provide combinationists with an explanation of representational and dispositional unity. What about phenomenal unity? Since I am here assuming transparency, I will treat phenomenal unity as the most basic form of representational unity. As noted in subsection 2.4, there is a question about how independent this can be of dispositional unity, i.e. how far it makes sense to think of two experiences being ‘merely phenomenally unified’, yet substantially unable to interact. To the extent that they cannot be merely phenomenally unified, the explanation for phenomenal unity will not go beyond the principles already laid out. Being phenomenally unified will just mean being OA-related with some greater or lesser degree of informativeness and salience. But if inert bare phenomenal unity is possible, as suggested by some accounts of the split-brain phenomenon, how do combinationists explain this?

In subsection 3.4 I described the possibility of ‘pure openness’, adumbration without specific informativeness or motivational salience about what it adumbrates. This inarticulate sense of there being something more than the given seems to me the best candidate for a compositional explanation of pure phenomenal unity, the inarticulate sense of things being ‘there together’ in the mind. That is, combinationists might say that experiences can be OA-related *uninformatively*, i.e. without any definite suggestion of the content or specific features of the other content.

If pure openness really is entirely uninformative, it cannot be described as veridical or non-veridical in the normal way. But the fact that pure openness has no truth-evaluable content does not imply that it has no content at all; it may still have reference. That is, when two contents are each experienced as “just one aspect of something”, the ‘something’ may behave rather like a demonstrative (“just one aspect of *this thing*”, said without any idea of the character of that thing) and refer to the other content, despite containing no descriptive information about the latter. Does it still make sense to imagine that it refers in virtue of the other? Since it does not refer by description, it cannot be in virtue

of qualitative facts about the other, but it might still be in virtue of some causal or ontological relation between the two. In the split-brain case, for instance, experiences might refer to each other in virtue of their belonging to the same organism. So we should count entirely uninformative adumbration as ‘trivially veridical’, as long as it involves successful reference by each experience to the other’s content.

Of course, one or both of pure openness and pure phenomenal unity might be non-existent, a mere theorist’s hypostatisation. The point is that a combinationist has at least two tenable views here – either there is no pure phenomenal unity, in which case they need not explain it, or there is, in which case they can claim that there is such a thing as pure openness to explain it.

Subsection 4.6: Transitivity and Reducibility

So combinationists can give either a positive or a negative answer to the second question from subsection 2.4. What about the other two questions in that subsection? The third was whether phenomenal unification is transitive, and it remains unclear whether the relevant, minimal, form of the OA-relation is transitive (obviously highly informative forms will not be, just as sophisticated forms of representational unity are not transitive). However, it might be made so in two ways. Firstly, we might simply allow for experiences to count as ‘indirectly unified’ if they are ‘indirectly OA-related’, connected by a chain of OA-relations. Being indirectly OA-related is definitionally transitive, and so might be a better candidate for consisting phenomenal unity than being directly OA-related. But insofar as this achieves transitivity by definitional stipulation, it might not satisfy someone who felt that the transitivity of phenomenal unity reflected a substantive fact about it.

A second way to make the OA-relation transitive would appeal to certain theses about reference. Reference often works by a kind of ‘chain’ – I can refer to something named long ago because a succession of acts of reference, each in some sense deferring to a previous act of reference, connects me to that first naming. So we might plausibly suppose that whenever an experience refers in a basic adumbrational way to another experience, it also thereby refers to all the experiences which the second experience thus refers to, thus making the basic OA-relation transitive. That is, when a single experience represents itself or its object as merely the revealed aspect of something of which there is ‘more’, this ‘more’ refers inarticulately to many other experiences, including any experiences which are themselves thus adumbrated by an experience referred to.

Finally, the fourth question of subsection 2.4 was whether phenomenal unity could be given a physicalistic explanation, and this will depend on whether the basic OA-relation can be explained physicalistically. If physicalistic explanations are available for consciousness in general, then there should be no special problem about explaining the OA-relation: physicalists can presumably give some account of perceptual adumbration in terms of their favoured theory of perception, and then connect this with their favoured account of what an experience is, and what content is, to derive an account of what it is for two experienced contents to mutually adumbrate each other. If the OA-relation is physicalistically inexplicable it will be because experience in general, or perception in general, is physicalistically inexplicable.

Section 5: Adumbration as an Explanation of Phenomenal Unity, Denying Transparency

In section 2 I noted that combinationists might choose to present an account of conscious unity in terms of experiential contents, or in terms of the non-representational features of experiences. In section 4 I explained how a combinationist might pursue the first of these options, appealing to mutual veridical

object-adumbration between experiences in virtue of some real relation, or the ‘OA-relation’. But someone who regards conscious unity as a non-representational phenomenon might object that this is missing the primary thing to be explained. While the OA-relation might explain representational unity, the unity of the experienced world, this is distinct from phenomenal unity itself, the unity of the phenomenal field which represents that world but has other phenomenological features.

I believe combinationists can address this challenge by claiming that experiences *themselves*, not just their objects, can adumbrate each other. In the subsection 5.1 I argue that if we deny transparency, it is reasonable to allow for such direct adumbration between experiences; in subsection 5.2 I show how this grounds an explanatory account parallel to that in section 4, and in subsection 5.3 I relate this to the specific architecture of the human mind.

Subsection 5.1: Experiences Adumbrating Experiences

When an experienced *object* adumbrates another, we are aware of it as merely the revealed aspect of something with concealed aspects, of which the adumbrated thing is one. For an experience itself to adumbrate another experience, then, we would have to be aware of it as the revealed aspect of something with concealed aspects, of which another experience is one. This raises two questions: first, can we be aware of an experience as the *revealed* aspect of something; and second, can an experience be adumbrated as the *concealed* aspect of something?

The first question asks about our direct awareness of experiences, and if we endorsed the transparency thesis we might have reason to answer it negatively, on the grounds that we never have direct awareness of experiences, but only of their objects. But in this section I am supposing transparency to be false, and so *ex hypothesi* we can be directly aware of our experiences. There is

nevertheless a real question about what sort of awareness this can be, and whether it is properly described as a form of ‘representation’. A particularly serious issue for combinationists is whether it depends on mental-state concepts, or is available to creatures lacking such concepts. If creatures without mental-state concepts cannot be aware of experiences, even in a thin sense, but nevertheless enjoy unified consciousness, then combinationism’s account of unity runs aground, unable to apply to simple minds and unable to explain our complex minds as composed of simpler minds. Such a position has the odd consequence that such creatures cannot in any sense be aware of the unity of their consciousness, but it is not absurd. Thus I believe combinationists must either endorse transparency and reduce phenomenal unity to some sort of representational unity, or else suppose that awareness of experiences is involved in all consciousness, even if in a thin way perhaps not best described as representation. And if we can be aware of experiences, I think it is hard to deny that we can be aware of them as mere aspects of something bigger.

The second question is whether we can be aware of experiences as concealed. I think we can, though Husserl seems to say we cannot:

A mental process... is not adumbrated. If I look at it, I have something absolute; it has no sides that could be presented sometimes in one mode and sometimes in another. (1982, p.96)²¹

However, I think there are familiar cases which are easily interpreted as involving adumbrated experiences, which make it plausible to deny Husserl’s claim. My first example is the ‘tip-of-the-tongue’ experience, in which we cannot recall something (often a word or name), but feel that it is ‘there’ in our minds. James describes this as:

²¹ This disagreement may be partly verbal. As I am using the term ‘adumbration’, it is not definitional that if something is adumbrated for me, it is not given to me; it may be given in some other experience (as argued in subsection 4.2). Husserl might simply mean that each of my experiences is given to me by some experience (presumably itself). However, since I will suggest shortly that component subjects in a unified composite subject are aware of each other’s experiences as concealed aspects of what their own experiences reveal, I deny even this weakened version of Husserl’s claim.

...a gap that is intensely active... A sort of wraith of the name is in it, beckoning us in a given direction, making us at moments tingle with the sense of our closeness, and then letting us sink back without the longed-for term. If wrong names are proposed to us, this singularly definite gap acts immediately so as to negate them... the gap of one word does not feel like the gap of another, all empty of content as both might seem... (1950, p.251)

We might think of this ‘active gap’ as the revealed aspect of a psychic whole whose concealed aspect is a full and explicit memory. This revealed aspect may be more or less informative and often makes the concealment of the name itself highly salient. It is not clear if we should think of what is here concealed as an experience: it is clearly something mental, but it might be an unconscious mental process. Alternatively, it might be phenomenally conscious but not access-conscious (an instance of phenomenal overflow) – or access-conscious to an insufficient degree, and less so than the ‘active gap’ (Cf. Block 2011, Hellie forthcoming-b, pp.6-7).

Similar uncertainty attends my second example: we might think of unattended experiences as partly adumbrated, if we supposed that they are phenomenally rich but cannot be accessed in all this richness while unattended. The revealed aspect of this experience is the kind of peripheral awareness we have of its broad outlines, while attending to something else, which is accompanied by a sense that there is more detail there which we could appreciate or access more fully if we shifted our attention.

My third example is our awareness of the experiences of others, which I think is better handled as a case of perceptual adumbration than as a case of either inference or direct perception. For instance, when two people conversing see and hear each other’s facial expressions and voices, they may perceive these expressive actions as the revealed aspect of a mental state which is not directly given, but is nevertheless experienced as ‘present’ through its expression. We do not generally feel as though we have to infer someone’s anger from their words and expressions, at least in the intellectual way that we might infer it from, say, reflecting on a pattern in their recent actions. Rather, we ‘see the anger in their

face’ – and yet we also tend to think that the anger itself is private: we are aware that it is fully given only through this other experience that we do not have.²²

Note that in these cases an external perceptual object – a face, a voice, a gesture – adumbrates an experience. Thus these are not instances of experiences adumbrating experiences, and moreover they do not bear on the issue of epistemic boundedness, since they confer knowledge of the other’s experiences only via knowledge of non-experiential things. The point is simply that there is no obvious absurdity in the idea that an experience might be adumbrated.

Subsection 5.2: Experiential Adumbration and the EA-relation

For the opponent of transparency who has rejected section 4’s explanation of conscious unity, but who has accepted the arguments of subsection 5.1, I propose an explanation of phenomenal unity based on the ‘EA-relation’. Just as the OA-relation obtained between two experiences whose objects adumbrated each other, largely veridically and each in virtue of the other’s content, the EA-relation obtains between two experiences which adumbrate each other, largely veridically and each in virtue of the other likewise adumbrating it. And just as two subjects are OA-related when an experience of one is OA-related to an experience of the other, so two subjects are EA-related when an experience of one is EA-related to an experience of the other. I now propose the following principles:

Unification-by-Experiential-Adumbration (UEA): A subject has a unified experience of X-and-Y-together whenever they experience X, and experience Y, and these experiences are EA-related.

²² Note that such adumbration need not specify *who* has that other experience. I can see an angry face, and perceive it as the revealed aspect of an anger whose concealed aspect is a distinct angry experience which this visual experience does not give to me - whether the face belongs to someone else, or is my own, seen in a mirror. I might see a face as angry – as adumbrating angry experience – before I have worked out whether it is my own reflection or not (e.g. in low-light conditions, or when hanging out with my identical twin).

Unified Composites from EA-Related Parts (UCEP): A composite subject enjoys conscious unity whenever its component subjects are EA-Related.

These explain phenomenal unity in a composite subject by appealing to the EA-relation among the experiences of component subjects. As with UOA and UCOP, the second principle follows from the first given the principles of inheritance defended in chapter 3.

This account can take over unaltered the principles CCIA and SCNC from section 4, explaining representational and dispositional unity in terms of the adumbration relations among experience's contents. Rejecting transparency poses no problem for these principles, since doing so is consistent with thinking that representational and dispositional unification depend heavily on the contents of the unified experiences.

Combining UEA with CCIA and SCNC raises the question of how phenomenal unity is connected with dispositional unity, the second open question of subsection 1.4. If phenomenal unity is the categorical basis for dispositional unity, then we should say the following: in a unified mind each experience's adumbration of the other experiences is what allows it to interact with them in the various ways constitutive of dispositional unity. It is because each experience involves awareness of the others that they can connect their contents, influence each other's development, etc. If we thought the priority was reversed, we should say that the experiences adumbrate each other, i.e. are experienced as connected, merely because they are prone to interact.

The second, third and fourth open questions in subsection 2.4 were whether phenomenal unification is more basic than, and dissociable from, dispositional unity; whether it is transitive; and whether it can be explained in physicalistic terms. Just as with the OA-relation, combinationists who relies on UEA can accept both positive and negative answers to each of these questions. Phenomenal unity without dispositional unity could involve pure openness, and the EA-relation can be made

transitive either by defining it to include both direct and indirect links, or by supposing that the relevant sort of reference is transitive – that an experience which adumbratively refers, via pure openness, to another experience, thereby refers also to all those to which the latter thus refers. And just as with the OA-relation, whether physicalists can explain the EA-relation depends on whether they can explain experience and perception more generally.

Subsection 5.3: Attention and the Contingent Structure of the Human Mind

Let us assume that either section 4 or the above two subsections have provided an adequate explanation of the nature of conscious unity in general. Combinationists can draw on that framework to aid in understanding the specific architecture of minds recognisably like ours, in which the details of different processes are accessible to consciousness to different degrees, and in which there is competition for focal attention amongst those things present to consciousness.

The particular way a set of component subjects adumbrate each other will reflect the contingent causal connections among them, in the human case realised neurally. More informationally isolated subjects will adumbrate others less informatively and less saliently, while being adumbrated less informatively and saliently in turn. Other systems may adumbrate each other more informatively and more saliently; these will be, in the terms of subsection 2.1, ‘more unified’ in the dispositional sense. If we adopt subsection 2.3’s suggestion of plotting each experience’s position in the ‘conscious field’ according to its degrees of unity, then this field’s geometry will reflect the contingent pattern of informative and salient adumbration among its parts, which in turn corresponds to the causal structure supplied by the underlying machinery.

This applies in particular to the way that at each moment a few particular elements in the mind occupy a ‘focal’ or ‘central’ position. Following the suggestions in subsection 2.3, we could say that this involves one element being especially strongly unified with the rest of the mind generally; following the proposal of this section, this would then mean one element coming to both adumbrate the other elements, and be adumbrated by them, either more informatively or as more salient (most likely both). It would ‘loom larger’ in the view of all the other elements.

There are multiple ways to implement this structure: compare the following two models. On the first, each component subject keeps its own proprietary experiences the same throughout a shift in attention: what changes is the strength of the connections among them, so that first one subject, then another, has increased ‘bandwidth’ for communicating with the others. On the second, the strength of the connections stays the same, but some or all of the component subjects adjust their own internal processing, and hence their own experiences, to ‘mirror’ the experiences of the others, so that first one experience-type or content, and then another, is being instantiated by a critical mass of component subjects. It is an empirical matter which model better matches the way human brains work. Brains might use a combination of the two, or use one to realise the other: e.g. a mechanism for synchronising groups of neurones (closer to the second model) might make one particular brain region better able than others to ‘be heard above the noise’ because its outputs are synchronised (closer to the first model). (Cf. Baars 1988, Melloni et al, 2007)

Summary:

I began this chapter by identifying a challenge for combinationists: to explain conscious unity, in a way that meets the independent grasp requirement. This might be easy, if ‘minimalism about unity’ is true, and an adequate explanation of unity need not advert to any distinctively experiential facts about

component subjects. But many combinationists have reason to doubt this, if they accept either primitivism about consciousness or the phenomenal interdependence of unified experiences. They then need to explain how conscious unity in the whole is experienced by the parts. And, as section 1 explained, they need to do so in a way that does not make each part the subject of all the others' experiences.

I then advanced a combinationist proposal, with different versions reflecting different answers to independently contentious questions about the nature of conscious unity. The guiding idea was that amodal perception – the perception of the unperceived – gives us a grasp on what it is like for a subject to experience an experience as unified with another experience it does not have. This basic claim can accommodate most positions on the nature of conscious unity.²³ To have a convenient label for this proposal, define 'the A-relation' as obtaining between two experiences if either *a*) experience is transparent and they are OA-related, or *b*) experience is not transparent and they are EA-related. (Similarly, two subjects are A-related if they have experiences that are A-related to the other's.) Then section 4's UOA and UCOP, and section 5's UEA and UCEP, can be combined into:

Unification-by-Adumbration (UA): A subject has a unified experience of X-and-Y-together whenever they experience X, and experience Y, and these experiences are A-related.

Unified Composites from A-Related Parts (UCAP): A composite subject enjoys conscious unity whenever its component subjects are A-Related.

Call these two principles, together with the supplementary principles CCIA and SCNC from subsections 4.3 and 4.4, the 'adumbration proposal'. Does the adumbration proposal provide an adequate explanation of the unity of consciousness? Is it intelligible to us that simply from two experiences being A-related, they must therefore be unified? Ultimately, whether one sort of experience

²³ An exception is the view that i) we are not aware of experiences themselves, but only of their content, but ii) conscious unity is not a matter of content. Section 4's account of unity denied ii), and section 5's denied i), but if both i) and ii) are true there is nowhere for combinationists to posit perceptual adumbration.

suffices for another is a phenomenological question, and one which may not admit of empirical investigation. Consequently, it is hard to conclusively resolve this question; the best combinationists can do is to ‘pump intuitions’ by repeatedly describing the two experiences in terms intended to make them sound similar.

If the adumbrational proposal fails, that does not conclusively show that all combinationist proposals fail, but it does not bode well. By contrast, if the proposal succeeds, that is a major step towards a comprehensive combinationist explanation of consciousness. In the next three chapters, I consider whether the framework developed so far (chapter 3’s inheritance proposal together with this chapter’s adumbration proposal) can deal with certain problems and paradoxes which arise from consideration of micro-subjects, mega-subjects, and large overlapping sections of subjects.

Chapter 5: Blurring, Blending, and Mismatch - Microsubjects and the Structure of Experience

<u>Section 1: The Blurring Problem</u>	p.182
Subsection 1.1: Confused and Distinguishable Experiences	p.183
Subsection 1.2: Radical Confusion	p.186
Subsection 1.3: Why we are Confused	p.189
Subsection 1.4: An Exclusionary Account of Blurring	p.191
Subsection 1.5: An Inclusionary Account of Blurring	p.194
<u>Section 2: The Blending Problem</u>	p.197
Subsection 2.1: Small-Palette and Large-Palette Solutions	p.198
Subsection 2.2: Putative Examples of Blending	p.202
Subsection 2.3: Arguments Against Blending	p.204
Subsection 2.4: From Limited Blending to Ubiquitous Blending	p.208
Subsection 2.5: What is the Right Relation for Blending?	p.211
<u>Section 3: The Mismatch Problem</u>	p.213
Subsection 3.1: Why Combinationists should Privilege Informational Structure	p.215
Subsection 3.2: Confusion and Dissociation, Blending and Refraction	p.217
Subsection 3.3: Dancing Qualia and the Boundaries of Consciousness	p.221
Summary	p.225

This chapter addresses several issues connected by the theme of experiential *structure*. In part, this is an attempt to expand combinationism’s positive explanatory scheme: the last chapter’s ‘adumbration proposal’ explains why many experiences might be experienced together as a unified phenomenal field, but what explains the particular way this field is structured? But addressing the topic of experiential structure is also important to rebutting some specific objections to combinationism in general, and to panpsychism, or any other theory which posits *microsubjects*, in particular.

What makes it seem especially hard for combinationism to account for experiential structure is the apparent discrepancy between how the physical brain is structured and how our consciousness is structured. This difference has been recognised as a problem not just for combinationism but for mind-brain identity theory in general, for it seems to tell against identifying brain processes with consciousness or regarding one as the basis for the other. Here is a representative statement of this supposed discrepancy from Maxwell:

How is it that the occurrence of a smooth, continuous expanse of red in our visual experience can... involve particulate, discontinuous affairs such as transfers of or interactions among large

numbers of electrons, ions, or the like? Surely being smooth or continuous is a *structural* property, and being particulate or discontinuous is also a structural property... incompatible with being smooth and continuous. (1978, p.398)

For a more detailed plan, I am indebted to Lockwood (1993) who distinguishes three specific strands of the problem:

1. Our experience is relatively *coarse-grained*, while any plausible composite basis is very *fine-grained*.
2. Our experience is *qualitatively diverse*, while any plausible composite basis has only a *few qualitative ingredients*.
3. The *type* of structure found in experience “seems not to match, even in coarse-grained fashion, that of the underlying physiology.” (p.544)

These three problems are addressed sequentially in the three sections of this chapter: I call them the ‘blurring problem’ (of how fine-grained structure *disappears* from the whole’s perspective), the ‘blending problem’ (of how qualitative diversity *appears* from the whole’s perspective), and the ‘mismatch problem’ (of why the *types* of structure diverge).¹

The first two of these problems are more definite, but do not face all versions of combinationism. Comparisons of qualitative diversity, or fineness of grain, only arise for those who postulate a microexperiential level that closely mirrors the microphysical level. Thus the problems of blending and blurring particularly face panpsychist combinationists, or other combinationists who posit microsubjects, conscious subjects on the scale of a cell, molecule, or fundamental particle. This special relevance is reflected in the repeated invocation of the problem of structural discrepancy as a worry for, or even a decisive objection to, constitutive panpsychism (e.g. Chalmers 1996, p.306, Goff 2006, p.57, Alter & Nagasawa 2012).

¹ Chalmers (forthcoming-a) mentions these three problems among others facing experiential combination, referring to them as ‘the revelation argument’ (p.12), ‘the palette argument’ (p.4), and ‘the mismatch argument’ (p.4). Lewtas calls the second ‘the blending problem’ (2013, pp.54-55), while Dainton calls it ‘the derivation problem’ (2011, p.246). My choice of labels is largely aesthetic.

Non-panpsychist combinationists, by contrast, might hold that below a certain level of decomposition one finds physical parts without any experiential properties, so that only some physical structure needs to be matched in the experiential realm.² Experience appears only at a certain scale, and experiential combination explains experiential structure above this scale in terms of experience at this scale. So for non-panpsychists, the problems of blending and blurring are more or less removed: whatever account is given of how experience arises from a complex non-experiential basis will take over the job of explaining why their structures differ. That job may still be challenging, especially for mind-brain identity theorists, but the challenge does not concern experiential combination specifically.

The problem of mismatch, however, faces even non-panpsychist combinationists, because even large parts of the physical brain are arranged in ways that seem not to match the structure found in experience. Admittedly, this problem is more nebulous than the other two: it is not easy to say, in general terms, what the two kinds of structures are that are meant to be discrepant, especially when we remember that strictly we should be comparing experience not with the *brain* but with brain *activity*. Nevertheless a few specific worries can be placed under this heading, such as the lack of correlation between which brain structures are spatially near or far from each other, and which things we experience as close to each other, or even with which things we experience as distinct or identical.

Another aspect of the mismatch problem is the problem that allowing any role for microexperiences seems to allow for ‘dancing qualia’ (Chalmers 1995b, cf. Block 1992, p.77ff). If we accept the premise that where there is no difference in how a system reports its phenomenology, there is no difference in actual phenomenology, it seems that any change between functionally equivalent systems - e.g. a change in the microscopic parts that make them up - will make no difference to phenomenology. But if the phenomenology of those microscopic parts really does enter into the

² There are some stranger views that complicate this picture, such as what Strawson calls ‘micropsychism’ (2006, p.24 ff), the view that some but not all fundamental particles have experiential properties. Since I am focusing on the full-fledged problem as it faces panpsychism, I will set these aside.

whole's experience, we would expect a change in the former to mean a change in the latter. Combinationists need to explain how this 'screening out' of certain aspects of microexperiences from macroexperience is compatible with the explanation of the latter by the former (cf. Sebastien 2013).

Addressing the problems of blurring, blending, and mismatch involves developing two alternative responses in parallel, one for the 'inclusionary' and one for the 'exclusionary' approach. I distinguished these two versions of combinationism in chapter 2, connecting them with different views of phenomenal overflow and of physical combination. By emphasising the whole's ontological intimacy with and dependence on its parts, we get 'inclusionary' combinationism, on which the whole inherits *all* the (basic-) experiential properties of its parts, even when they are not unified with, or cognitively and behaviourally integrated with, each other. Conversely, by emphasising the whole's ontological autonomy from its parts, we get 'exclusionary' combinationism, on which the whole may lack many of the properties of its parts, including any (basic-) experiential properties which are not suitably connected to others. Then in chapter 3 I argued that for a certain sort of combinationist - the *a priori*, subject-first, panpsychist - only the inclusionary approach can provide an intelligible connection between part and whole.

How combinationism handles the issue of experiential structure will reflect the choice between these two approaches. Inclusionary combinationists must say that everything making up the structure of the parts' experiences – the same qualitative palette, the same types of relations, the same degree of detail – is present in the whole's. This forces them to pursue the difficult line that our experience is actually radically more fine-grained than we took it to be.

For the exclusionary approach things are more complicated. The exclusionary approach holds that experiential goings-on in the parts appear in the whole only when they meet certain conditions. Do the fantastically numerous and rudimentary experiences of microsubjects meet these conditions? In

particular, do they play the right causal roles to guide the whole's thought and behaviour? While some microexperiences clearly do not (e.g. whatever experiences a panpsychist attributes to particles in my hair), this question is hard to answer for experiential events in the parts of a live human brain. Considered together, this mass of experiences seems to be precisely the thing that directs the person's behaviour. But considered individually, no single microexperience makes an appreciable behavioural difference.

I think exclusionary combinationists should say that the part's microscopic experiences are inherited by the whole *as a composite experience* but not *individually*: thus no single experience is shared between microscopic part and macroscopic whole, but the whole's experiences are composed of the experiences of its microscopic parts.³ So exclusionary combinationists, but not inclusionary ones, accept a radical difference between the whole's consciousness, with its characteristic structure, and the many consciousnesses of its parts, with their characteristic structure. Of these two approaches, it might seem at first that inclusionary combinationism is wildly implausible, and exclusionary combinationism worryingly obscure. I will argue, however, that both are more viable than they seem.

Section 1: The Blurring Problem

The blurring problem concerns the *absence* from our experience of the sheer degree of detail that physics tells us is present in our brains. Consider someone smelling a simple odour, or hearing an unchanging pure tone. From the subject's point of view, these events may appear entirely simple and structureless, but we know that they arise from the simultaneous activity of a great number of different

³ Technically, there is room for a combinationist to be 'semi-exclusionary', affirming conditional experience inheritance without basic-experience inheritance, but allowing individual microexperiences to be inherited by the whole as long as they are among a group of microexperiences that collectively meet the conditions for inheritance. They would then be 'exclusionary' in the sense discussed in chapters 2 and 3, but combining this with what this chapter calls an 'inclusionary' response to the problems of structure, with all the costs and benefits thereof.

neurones, each with billions of billions of parts. So we may wonder, “how do all these microstructural discontinuities and inhomogeneities come to be glossed over?” (Lockwood 1993, p.544) For convenience, I will put this by saying that human experience appears ‘coarse-grained’, while its neural basis is ‘fine-grained’.

Exclusionary combinationists will approach this problem by granting that the whole’s experience really is coarse-grained, and explaining why the fine-grained collection of experiences belonging to the parts gives rise to this distinct coarse-grained experience. Inclusionary combinationists, by contrast, claims that the experience of the whole is in fact fine-grained in the relevant sense, since it contains all the experiences of the parts; they then explain why we are inclined to wrongly think that our experience is coarse-grained. Both approaches, however, should base their explanations on the notion of *radical confusion*, though employing it in different roles: either to explain why the details of microexperiences fail to appear in consciousness (the exclusionary approach), or to explain why details in consciousness fail to be noticed as such (the inclusionary approach).

In subsection 1.1 I elaborate this notion of ‘confusion’, inspired by similar doctrines held by Spinoza and Leibniz. In subsection 1.2 I define the notion of *radical* confusion, and in subsection 1.3 I explain why we should expect the experiences of our small parts to be radically confused. In subsections 1.4 and 1.5 I develop specifically exclusionary and inclusionary responses to the blurring problem.

Subsection 1.1: Confused and Distinguishable Experiences

I think the combinationist should claim that our experience seems coarse-grained because the details of our parts’ experiences are ‘confused’ one with another. What does this mean? I take the inspiration and label for this idea from the Early Modern rationalists, some of whom faced their own versions of the

blurring problem. In particular, Spinoza and Leibniz are both committed to the claim that every single event occurring in the human body has a corresponding mental event which is present in the human mind.⁴ How is this fantastic level of mental detail to be reconciled with our apparent ignorance of the processes occurring in our bodies? For both authors, the solution appears to rest upon the idea of *confusion*: bodily sensations are always confused, and thus while the mind perceives them in some sense, it is in another sense unaware of them.⁵ Consider a famous passage from Leibniz:

The perceptions of our senses even when they are clear must necessarily contain certain confused elements... [for] while our senses respond to everything, our soul cannot pay attention to every particular... It is almost like the confused murmuring which is heard by those who approach the shore of a sea. It comes from the continual beatings of innumerable waves. (2012, p.96)

And here is one from Spinoza:

The human body, being limited, is only capable of distinctly forming a certain number of images within itself at the same time... if this number is exceeded, the images will begin to be confused, and if the number... is largely exceeded, they will all be completely confused with one another... When the images become quite confused in the body, the mind also imagines all bodies confusedly without any distinction, and will comprehend them, as it were, under one attribute. (1994, p.140).

Both passages seem to present the same idea: the finite capacities of the human mind ensure that many of its ideas will be ‘confused’ in the sense that it will be unable to distinguish them, i.e. think or attend to them separately. Michael Della Rocca helpfully offers the following definition: “For Spinoza, an idea is confused when it represents... two separate things and yet the mind is unable to distinguish

⁴ In Leibniz’s case, the problem is even more radical, since each mind represents not only its own body but the whole universe.

⁵ Leibniz appears to recognise, and respond to, the challenge of blurring much more explicitly than Spinoza does. Developing a Spinozistic response to the blurring problem is thus more of an exegetical task; see Wilson 1999 for a discussion of some attempts and their shortcomings.

these things by having an idea that is just of one of the objects and an idea that is just of the other of the objects.”(2008, p.113)

Since there is scholarly dispute about what exactly Leibniz and Spinoza mean by confusion, I do not propose an interpretation, but simply note an inspiration. I define confusion thus: two mental elements are confused with each other, relative to a subject and a mental operation, when that subject can perform that mental operation on both at once, but not on either separately. They are distinguishable insofar as they are not confused. This definition is deliberately very broad: shortly I will zoom in on the particular form of confusion that I think can help combinationists.

What are the ‘mental elements’ and ‘mental operations’ this definition speaks of? I intend these phrases to cover any kind of mental thing which can be the object of any kind of mental process: the notion of confusion can be neutral among different accounts of how the mind is organised. Prominent examples of mental operations might include ‘thinking’ or ‘entertaining’, ‘introspecting’ or ‘being aware of’, ‘attending’, ‘imagining’, or ‘recognising’ in the sense of categorising under concepts or of judging distinct from or identical with something else.⁶ Mental elements might be ‘experiences’ (however individuated, cf. Chapter 3, section 3), or ‘ideas’, ‘contents’, or ‘phenomenal qualities’ understood as the things awareness of which constitutes the having of an experience.⁷

Note that confusion is different from indiscriminability, the relation between two items which ‘appear the same’ to a subject. Confusion is primarily a relation between particular tokens, whereas

⁶ For many mental operations, the most natural thing to identify as their ‘object’ would be something extra-mental: a physical object thought about, an abstract proposition entertained, etc. I am supposing that even for such operations we can make sense of there being at the same time an intra-mental object, though the operation might be subtly different. For instance, when I focus my attention on some external thing, I surely do so by some element of my mind - a thought, a representation, a ‘mental file’ – being ‘accessed’, ‘loaded’, or ‘activated’, if not actually attended.

⁷ In the primary instance these elements will be tokens, but we can easily define a secondary sense in which two types are confused for a subject when any tokens of those types onto which a given subject could direct a given operation would be confused. Note that type-confusion requires more than that every token of either type would be confused with *something* - two types count as distinguishable from each other as long as each one’s tokens can be thought without that particular other type’s tokens being thought.

indiscriminability involves two tokens seeming qualitatively equivalent to a subject even while they are recognised by that subject as numerically distinct.⁸ Note also that since confusion is subject-relative, the same element might belong to both part and whole but be confused for one but not for the other. Indeed, for small enough parts confusion might disappear simply because the parts each experience only one thing, and can thus trivially be said to be able to ‘distinguish’ it from what they are not experiencing.

Subsection 1.2: Radical Confusion

Let us draw three distinctions among types of confusion. First, recall that confusion must be relative to some particular sort of mental operation. A particularly interesting sort of confusion would be that which applies to *all* the mental operations a subject is capable of. Call this ‘strong confusion’, and call the contrasting case, where elements are confused only relative to some operations, ‘weak confusion’. For instance, we might be unable to call to mind the flavour of coffee without at the same time calling to mind the bitterness of its taste, yet nevertheless able to attend (and perhaps apply concepts, like ‘bitter’) to them separately. Then the experiences of flavour and bitterness would be confused relative to some mental operations (like ‘calling to mind’), but not relative to others (like ‘attending’), and so would be weakly, not strongly, confused.

Second, distinguish symmetric and asymmetric confusion. There seems to be nothing impossible in the idea that someone could think of two things together, and think of the first without the second, but could not think of the second without the first. Then there would be a sort of confusion

⁸ Indiscriminability might, however, be analysed as a particular form of confusion, relative to certain mental operations of the form ‘perceive as having quality X’ (or perhaps ‘perceptually categorise as having quality X’): when two colour patches, say, are indiscriminable with respect to colour, the subject can only perceive the one as, say, scarlet if they perceive the other as scarlet. Specifying the exact operations involved will face all the same challenges and complexities as specifying the exact meaning of ‘appear the same’; see Raffman 2012, pp.310-311 for an unpacking of some different meanings of this phrase.

involved regarding the one but not regarding the other. Let us say that these two elements are ‘asymmetrically confused’. For example, perhaps we can never experience certain bodily sensations (e.g. pain, itching, discomfort, or nausea) without also experiencing displeasure, and cannot even attend to the distinctive sensory element of the sensation without attending also to that displeasure. Nevertheless we can experience and think about displeasure independently of the sensory element, and to that extent might come to suppose that there are two distinct elements present here which are asymmetrically confused.

Third, confusion may depend on circumstances. Someone who is tired, distracted, drunk, or having to respond quickly may be unable to distinguish things which they would be able to distinguish given better conditions: that is, their experiences may qualify as confused only relative to those circumstances. Confusion may also be relative to a subject’s conceptual repertoire – it might be that they cannot distinguish two ideas using their present concepts, but would be able to if they refined their concepts, or learnt new ones. Indeed, a common activity of philosophers is to claim to have identified a confusion of this sort in our everyday concepts, which requires the introduction of technical concepts to remove.

For example, consider someone who does not distinguish between something being customary, and its being the right thing to do. They would be unable to think that a given action was customary without thinking that it was the right thing to do. This might well be strong confusion, if they could not even focus attention on the one idea without also attending to the other. Yet they could come to distinguish the different ideas by learning appropriate concepts – after which they could retrospectively recognise their previous, shallow, confusion. Let us call confusion which can be removed either by adjusting the subject’s bodily surroundings or condition, or by improving their conceptual repertoire, or in some similarly mild way, ‘shallow confusion’. By contrast, call confusion which persists even into

ideal conditions, ‘robust confusion’ (the boundary between the two being as well-defined as the notion of ‘ideal conditions’ allows).

Having defined confusion, and divided it into symmetrical and asymmetrical, strong and weak, and robust and shallow versions, let us define *radical confusion* as confusion which is symmetrical, strong, and robust. That is, mental elements are radically confused with each other when the subject cannot distinguish any of them, by any means, under any circumstances.

It is important to see that radical confusion is likely to be undetectable to the subject; someone having radically confused experiences (say) could not tell that they are. This is because noticing confusion requires some sign of there being multiple elements, and if we are completely unable to distinguish them, the normal signs will be unavailable.

For instance, we have an easy way to identify shallow confusion – we remove it, then contrast the resulting distinction with the earlier confusion. With robust confusion, that is impossible, but we might notice the confusion if it was only weak, for we would then be able to distinguish the elements in one fashion while noting our inability to do so in another fashion. For example, if we could not imagine one sensation without another arising alongside it, but could still attend to the two separately (and go on to name and conceptualise them independently), the possibility of two attentive acts would be a sign of two different mental elements. Finally, if confusion is asymmetrical, we can distinguish the confused pair from at least one element, and thereby infer the existence of a contrasting element which we cannot distinguish from the pair.

Lacking all three of the above means of recognising confusion, it seems we could not tell that we were suffering from radical confusion. An apparent counter-example is ‘cacophonous’ noises. For instance, if we enter a bar and are overwhelmed by the combined noise of many voices, chair movements, music, and so on, we seem to perceive that there are multiple sounds present, but cannot

focus on any one of them individually (cf. discussions of the ‘problem of the speckled hen’, Ayer 1940, Tye 2009). However, I think this is better analysed as a case of shallow confusion exacerbated by the brevity and equal salience of the elements. The component noises *could* be separately attended with time and effort, but are so short in duration that we cannot focus on them before they are gone, and so similar in salience that we cannot select one to devote the necessary effort to. In such a situation, if we do decide to arbitrarily seize upon one component and focus on it, we usually succeed if it persists for more than a few moments. So this is not a case of radical confusion.

Is radical confusion even possible? If I am right that it would be unnoticed by the subject, we cannot establish its possibility by adducing positive examples. But I think we should regard it as possible on weaker grounds: we can and do find cases of robust but weak confusion, strong but shallow confusion, and strong, robust, but asymmetrical confusion. Without special reason to think that these features preclude each other, we should suppose that they might co-occur.

Subsection 1.3: Why we are Confused

With the notion of radical confusion on the table, we can sketch two combinationist responses to the blurring problem. The inclusionary response claims that all the experiences of the parts are had by the whole, but that they are, for that whole, radically confused and as a result are *misinterpreted* as comparatively coarse-grained experience. The exclusionary response claims that the experiences of the parts are not individually had by the whole because they are confused relative to the operation ‘becoming conscious of’, and it is a necessary condition of having two experiences that they be distinguishable relative to this relation.

In both cases, the experiences of our tiny parts are said to be radically confused with one another, when considered relative to the whole. That is not to say that each one is confused with every single other one, taken pairwise, but that each element is radically confused with a great many other elements.⁹ This claim of confusion underlies a combinationist response to the blurring problem, but itself demands explanation: why are the experiences of our smallest parts radically confused for us?

The simple answer is that the human brain is not constructed so as to be able to individually register and distinguish all the trillions of events in its neurones, nor to direct attention onto them, report them verbally, encode them in memory, or otherwise access them. This lack of sensitivity to minute internal fluctuations is not anomalous: any plausible physical mechanisms would display it as well. To actually be sensitive to every internal event, if possible at all, would require something like a trillion electron-microscopes each pointing at each other's constituent parts. Why would evolution produce such a fantastic structure? Clearly, it has not – our sensitivity to internal events is low enough that small ones cannot reliably be detected or discriminated from each other. This inability is why the experiences of the parts are absent from, or at least not discernible in, the consciousness of the whole.

Mental operations typically involve the deployment of finite cognitive resources, and the fainter the elements to be distinguished, the greater the resources needed. To reliably extract information about something, a system must be receiving some sort of 'signal' from it, and must be able to discern that signal from background noise. In a sense all the tiny parts of the brain send 'signals' to the rest of the brain, through the chemical, electromagnetic, and even gravitational effects they have

⁹ This might mean that there are a number of clusters of experiences, all members of each of which are pairwise confused with each other, but not with the members of other clusters. Alternatively, it might involve continuous chains of confusion, with the end-points distinguishable but each pair of steps confused. The latter version would be ruled out if confusion were transitive, and as defined it appears so: if A cannot be thought without B being thought, and B cannot be thought without C being thought, then A cannot be thought without C being thought. However, this transitivity disappears if we allow for the 'cannot' to assert only very low probability: that is, if A is thought then B will *almost* certainly be thought, with a probability close to 1, but might not be. The probability of C being thought will be slightly less, and so on until we fall below the (likely vague) threshold for 'cannot'. (Cf. discussions of the non-transitivity of phenomenal indiscriminability, such as Goodman 1951, Hellie 2005, and Raffman 2012)

on their surroundings. But these signals will be very weak and stand out from background noise very little. Consequently, for small enough elements, the resources required to distinguish them are greater than the brain can muster, even under ideal circumstances, and hence they will be radically confused. To use a social analogy, it is hard for everyone in a room to hear everyone else, especially if some have weak voices. In a room of a trillion people, no individual's voice would be distinctly audible, because the others would produce so much noise (both literal and statistical), even though their voice is part of the audible roar.

Subsection 1.4: An Exclusionary Account of Blurring

On the exclusionary approach, the experiences of my brain's smallest parts are not experiences of mine, though they can still be described as my 'mental elements' in the weaker sense of 'experiences going on in me.' Considered as mental elements of mine, they are radically confused for me. But what I *do* experience is the composite experiences they compose.¹⁰ The exclusionary combinationist still accepts experience-sharing for subjects of similar size (e.g. my head and my brain), but not between microsubjects and macrosubjects: the former's experiences compose those of, but are not shared with, the latter.¹¹

Why does the whole *not* share the experiences of its microscopic parts? Exclusionary combinationists already (in light of chapter 3's discussion) recognise certain necessary conditions on how an entity must be related to an experience to be called its subject, conditions including some set of relations between that experience and the other parts, other experiences, or other behaviour, of that

¹⁰ Alternatively, perhaps I experience the microexperiences collectively but not individually (i.e. experience the collection but not its members), if this is coherent and does not collapse into experiencing a whole they compose.

¹¹ Note that the sense in which microexperiences are here called 'parts' of macroexperiences cannot be Bayne's notion of 'subsumption', for on that notion the subject of a composite experience automatically has all the component experiences the latter subsumes.

entity. Whatever the exact details of those conditions, we can stipulatively describe this as a certain mental operation being performed on that experience, an operation which serves to make its target an experience of the entity performing it. Call this operation ‘bringing into consciousness’. The specifics of this operation can be filled in with whatever is one’s preferred account of what makes certain brain events my conscious experiences: maybe they need to be sufficiently informationally integrated with the system’s other events (Cf. Tononi 2012), or be attended in some degree (Cf. Prinz 2011, 2012), or play the right functional roles in guiding behaviour (Cf. Putnam 1965, Shoemaker 2003b), or be phenomenally unified with other events (Cf. Bayne 2010). We need not even think of this as a distinct specific operation; we might regard it as a disjunction or determinable of many sorts of operation, so that all the specific operations that we perform on our experiences are just different ways of bringing them into consciousness.

Given that confusion was defined in terms of deliberately broad notions of ‘mental element’ and ‘mental operation’, strongly confused elements will be confused relative to this operation as well as others; and so the informational limitations discussed in the last subsection would entail that the whole cannot bring one of its smallest parts’ experiences into consciousness without bringing many others into consciousness at the same time. The exclusionary combinationist can then affirm the following principle:

Experiential Compression (EC): For two experiences to belong to a subject, they must be distinguishable (i.e. not confused) with respect to the ‘bringing into consciousness’ operation, relative to that subject.

This principle implies that the radically confused microscopic experiences of the brain’s tiny parts will not belong to the whole, because each is radically confused with many others. The rationale for EC would be the same intuition that motivates the exclusionary approach more generally: it is constitutive

of something being my experience that I be able to attend to it, access it, or even report it. If so, it is likely a conceptual truth that two things could only be two *distinct* experiences of mine if I were able to distinguish them – ‘able’ at least in that their confusion was shallow, weak, or asymmetrical. My experiences are ‘for me’, and so should be individuated according to what I can do with them.

This is a form of ‘quasi-inheritance’, as explained in subsection 4.4 of chapter 3, rather than direct inheritance, and the explanation of why the whole quasi-inherits this experience composed of those of its parts cannot be exactly the same as the explanation of why it directly inherits the experiences of its larger parts. But we must recall that this proposal is only intended for exclusionary combinationists, not inclusionary combinationists, and so we need not try to extend explanations based on Basic-Experience Inheritance.

Rather, quasi-inheritance would be explained either by the conceptual reduction of experiential ownership to some kind of complex causal relation, perhaps together with the relation of underlying, or by the posit of an *a posteriori* natural law. In the first case, I will naturally underlie, or be underlain by, the composite experience in virtue of my ontological relationship to my parts, and thereby to their experiences. And if, due to the organisation of my brain matter, the composite experience plays the right causal role, it can be ascribed to me as my experience. In the second case, the question is simply whether CEI, construed as a law of nature, remains suitably intelligible even after the conditions for inheritance are expanded to state that the whole will not inherit experiences which are confused relative to the bringing into consciousness’ relation, but will experience the composite of them. If too many conditions and exceptions are built in, CEI will seem like an ‘emergence law’, and to that extent an abandonment of combinationism. But if the conditions for inheritance already included in CEI, prior to incorporating a response to the blurring problem, are sufficiently simple and general then the new

addition will be too, for ‘bringing into consciousness’ will then involve only those adequately simple conditions, and so too will ‘confusion with respect to the bringing into consciousness operation’.

Subsection 1.5: An Inclusionary Account of Blurring

For inclusionary combinationists, the relation of composite experiences to their confused parts is just phenomenal subsumption, and thus entails that there is radical confusion (relative to me) in experiences I actually have, namely those I inherit from my tiny parts. All of these are present to me, but so radically confused with one another that I am inclined mistakenly think them absent. The natural objection is that this sort of mistake is implausible: if our experience is really immensely fine-grained, then that richness “couldn’t help but be manifest to consciousness” (Coleman 2012, p.144). In Chalmers’ formulation of the blurring problem, this appears as the premise that “The nature of consciousness is revealed to us in introspection” (forthcoming-a, p.12).

Note that it is not enough to point out how radical confusion will naturally lead to an inability to verbally report or reflectively identify particular confused experiences, for the objector’s contention is not that we *cannot distinguish* multitudinous fine details of our experiences, but that our experiences are directly grasped *as not having* such details. They assert that many of our experiences are visibly, evidently, and manifestly simple (understanding ‘simple’ in terms of a subsumptive notion of experiential parthood, so that parts of my experience are also experienced by me). Let us explore what that the inclusionary combinationist’s denial of this manifest simplicity amounts to.

Let the term ‘smooth’ stand for the sort of phenomenal character whose possession by some of our experiences gives rise to the blurring problem. The objector claims that smoothness is structural simplicity – not that smoothness arises from a simple basis, since they may accept that the basis of

smooth experiences is a composite brain, but that smoothness is a directly-apprehended structural property of experience. It contrasts with ‘bittiness’, the directly-apprehended structural property of those experiences which present themselves as having distinct elements.

One response that is unlikely to convince objectors like Coleman and Chalmers is that macroexperience involves an ‘illusion’, in any strong sense of that term. Illusions are when something seems one way, but is not that way – but if consciousness just is how things seem, this discrepancy cannot arise; the seeming itself cannot be false to itself. Thus if we directly apprehend consciousness as having a certain structure, then consciousness must really have that structure. At least, this is what the premise that introspection ‘reveals the nature of consciousness’ would suggest.

Inclusionary combinationists can say, however, that the full richness of macroexperience *is revealed*, but is *misinterpreted*. Thus rather than treating the smoothness of experience as itself misrepresenting anything, they could claim that the objector has misinterpreted smoothness; smoothness is some other property which is easily mistaken for structural simplicity. In particular, they might claim that smoothness is the property of ‘lacking *distinguishable* elements’, which is compatible both with having no elements and with having only radically confused elements. Bittiness would also be a directly-apprehended structural property of experience, namely ‘having distinguishable elements’. The blurring problem arises because we readily mistake our direct apprehension of a lack of distinguishable details for a direct apprehension of a lack of details.

This is not an ‘illusion of simplicity’: an illusion is where our experience tells us something false – as when a straight stick placed in water looks bent. Our experience does not ‘seem simple’ in this sense – it does not feel some way that only simple experiences feel. Rather, it is like the apparent motion of the sun – a veridical impression prone to an easy misinterpretation. The sun’s motion is not an illusion: that is how stationary objects look to a rotating observer. But we very readily infer from it

something mistaken, namely that the sun orbits a stationary earth. Similarly, says the defender of combination, our experiences feel exactly how massively complex but radically confused experiences feel. Insofar as they are ‘telling us’ anything, they are telling us something true – that we cannot distinguish details within them.

This response relies on us being prone to systematically misinterpret the directly-apprehended structure of consciousness.¹² Is such systematic misconstrual plausible? Here are two reasons to think so. First, when we take lack of distinguishable elements for lack of elements, the sort of inference we make (taking absence of evidence for evidence of absence) is easy and tempting, and often quite reasonable – indeed, it might be justifiable if we had no independent reason to think that experience arises from the massively composite brain. This corresponds, in the case of the sun’s apparent motion, to the error of neglecting to account for the motion of our own point of view, which is also an easy and tempting heuristic, and often appropriate: usually when we see something move it is not because we are standing on something that is rotating relative to it.

Second, we make this error in an unsupportive context, where our normal presuppositions do not hold, and it is hard to acquire information to correct them. In the case of the sun’s apparent motion, we cannot leave our earth-bound position to look from a third point of view (and when we do, on a shuttle or satellite, the misinterpretation disappears). Moreover, we lack the usual cues that our own standpoint is moving (e.g. air resistance). Similarly, when we interpret the smoothness of our experiences, not only do we lack the usual indications that our experiences are confused, (as described in subsection 1.2) but we are also profoundly limited by the fact that if our experiences *are* all massively complex, then we have no idea what a simple experience would be like.

¹² It is not really clear how systematic this misconstrual is: how often does the average person put any interpretation on the smoothness of their experiences? One might think that insofar as there is a mistake here, it is one mainly made by philosophers in the business of phenomenological meditation.

It bears emphasising that we are not in the position of one who has experienced both massively-complex-but-radically-confused experiences and also genuinely simple experiences, who could then observe which of the two was ‘smooth’ and which had some other character. Rather, if our parts’ experiences really are radically confused relative to us, we are in the position of someone experiencing one or the other of these, and trying to determine which without any basis for comparison. In such an unsupportive context, we might easily go wrong. Nevertheless the intuitive implausibility of accepting a systematic mistake in how we think about our own experience might persist, and might be regarded as enough to make the inclusionary response to the blurring problem seem less attractive than the exclusionary response.

Section 2: The Blending Problem

To set up the blending problem, I will quote in bulk from Lockwood:

There is nothing qualitatively distinctive about a neuron in the auditory cortex, or the corresponding action potential, to mark it out from a neuron, or the firing of a neuron, in the visual cortex. So how, on this basis, is one to account, say, for the fundamental phenomenological difference between a sound and a flash?...It seems inconceivable in much the same way, and for much the same reasons, that it is inconceivable that an artist, however skilled, should conjure the simulacrum of a Turner sunset from a palette containing only black and white paints. (1993, p.546)

Consciousness seems qualitatively rich, but any structure isomorphic with the physical brain would be qualitatively sparse. Combinationists must explain how the diversity of qualities we experience arises from the qualities experienced by our miniscule parts. Note that this statement of the problem draws on an implicit notion of ‘quality’ and ‘qualitative difference’, and an assumption that while physical structure might explain the ‘structural’ features of experience, it does not fully explain its ‘qualitative’ features: thus qualitative features require a distinctively qualitative explanation. The problem is

nonetheless a pressing one for panpsychists, since they are typically motivated at least in part by anti-physicalist arguments, like the appeal to the ‘inverted spectrum’, which have precisely that conclusion, that the qualities of consciousness are not explicable by any purely structural factors.

Whether or not there is a special challenge in explaining ‘qualities’, there is room to wonder which features of experience count as ‘qualitative’. The paradigm instances are perceptual sensations of colour, flavour, sound, and so on, but different theorists might also think of any aspect of consciousness that was regarded as irreducible to physical structure as involving a ‘phenomenal quality’ - doubt or certainty, joy or sorrow, desire or resolution, might all be, or involve, some sort of distinctively cognitive, affective, or conative qualities. While my discussion is focused on sensory qualities, it is intended to be neutral on this question.

Subsection 2.1: Small-Palette and Large-Palette Solutions

Let us accept, then, that macroexperiential qualities are to be explained by microexperiential qualities, however exactly ‘qualities’ is understood. Panpsychist combinationists must then choose between what Chalmers calls ‘small palette’ and ‘large palette’ approaches: small-palette approaches claim “that all macroqualities can be generated from just a few microqualities, if we find the right underlying microqualities”, while on large-palette approaches, we “suggest instead that the full range of macroqualities are included among the microqualities... [including] colors, sounds, smells, tastes, and so on” (forthcoming-a, p.26). We can put this by saying that the small-palette approach requires, while the large-palette approach does not, some kind of actual ‘phenomenal blending’ of microqualities into diverse ‘mixtures’. The major question that this section will discuss (in ways largely matching the

argument given in Roelofs 2014a), is whether there is a coherent and intelligible notion of ‘phenomenal blending’ available.

Let us define the required sort of ‘blending’ more precisely. It involves a composite experience which, merely in virtue of two (or more) parts of it displaying certain phenomenal qualities, and standing in certain relations, displays a single phenomenal quality, type-distinct from either but reflecting both in such a way that its dependence on them is intelligible. Call the former qualities the ‘ingredients’, and the latter the ‘resultant’. Note that ingredients are meant to persist, not to ‘vanish’ into the resultant so as to no longer be instantiated: yet nor is the resultant ‘mere appearance’: the resultant and ingredients are both genuinely present. Moreover, we only have phenomenal blending when it is intelligible why *that* resultant arose from *those* ingredients.

We can distinguish two forms of phenomenal blending: in the ‘intra-subject’ case a single subject, in virtue of experiencing certain qualities in certain relations, experiences their resultant, while in the ‘inter-subject’ case neither ingredient is itself experienced by the whole, who experiences only the composite experience whose quality is the resultant. To address the blending problem, inclusionary combinationists will appeal to intra-subject, and exclusionary combinationists to inter-subject, blending.

This definition leaves unspecified what relation between ingredients is necessary for blending; I consider this question in subsection 2.5. Moreover, this definition assumes some understanding of what counts as a single phenomenal quality and what does not; we should not count “the feeling of seeing red while angry” as a blend of redness and anger, since these are not experienced as a single quality. For the purposes of this discussion, I would like to apply a deliberately restrictive definition of a single quality: something which presents us with no qualitatively distinct elements that can be individually attended or otherwise ‘picked out’. This connects the notion of blending neatly with the notion of

confusion laid out in the last section, though not quite the notion that I there called ‘strong confusion’. Two mental elements are strongly confused if their subject cannot attend to either individually; my definition of ‘a single quality’ is something which does not *present us* with multiple targets of attention (call this ‘strongish’ confusion). These might come apart if we are enabled to pick out elements of quality *X on the basis of other experiences*, but could not do so just on the basis of experiencing quality *X*. This distinction between strong and strongish confusion will be explained more fully in subsection 2.2.

The notion of phenomenal blending lets us complicate Chalmers’ scheme of small-palette and large-palette approaches, with both extreme and moderate versions of the latter. At one extreme is the view that phenomenal blending is entirely impossible. This position appears to have been held by William James, who insists that “we cannot mix feelings as such, though we may mix the objects we feel, and from *their* mixture get new feelings.” (1890, p.157). At the opposite extreme are small-palette approaches, holding that *all* the phenomenal qualities we enjoy can be explained as compounds of a small number of fundamental qualities. This position appears to have been held by Pierce (1998, pp.35-36) and by Spencer (1899, §60) who suggests that there may be an “ultimate unit of consciousness, [so] that all the unlikenesses among our feelings result from unlike modes of integration of this ultimate unit”.

In between is the position that some but not all qualities result from blending, as when Lewtas argues that “orange-experiences result from combining... red-experiences and... yellow-experiences”(2013, p.54), but that “we don’t see, and don’t see that we ever could see, how to build red-experiences out of [other] experiences.”(2013, p.46) This gives three positions, two extreme and one moderate: the ‘Jamesian’ view, the ‘Lewtasian’ view, and the ‘small palette’ view.

Jamesian positions, combined with panpsychist combinationism, would imply a bloated ontology with thousands of fundamental phenomenal properties,¹³ for all qualities experienced by a composite subject must be experienced by at least one of its parts. This is implausible, since it seems unlikely that such simple minds could share all the diversity of qualities that human minds have. Lewtasian positions accept some limited phenomenal blending, but are still fairly unparsimonious. If, for instance, there is blending within a sensory modality, but not between the basic qualities of different modalities (e.g. redness, saltiness, low-pitch, etc.), at least one part of a human mind must experience each item on this fairly extensive list. To maintain panpsychism, as Lewtas does, we must attribute “an uncomfortably large number of strictly-basic conscious properties” to fundamental particles (2013, p.62). By contrast, small-palette positions are highly parsimonious, with a small set of basic elements generating a vast diversity of observed forms.¹⁴ Thus panpsychist combinationists have good reason to hope that the small-palette approach is viable.

To elaborate on the small palette approach, consider a ‘small palette hypothesis’ that a panpsychist combinationist might endorse:

There are a small number of basic qualities, which the simplest conscious parts of the brain experience; larger more complex brain parts support experience of the resultants of blending these. For each determinate sort of brain part, up to and including the whole brain, its structure determines a subset of phenomenal qualities out of all the possible combinations that its components’ could blend into. For instance, the brain’s structure dictates that we can

¹³ How objectionable this ‘bloating’ is may depend on whether it is seen as a matter of quantitative parsimony (number of particular existents) or qualitative parsimony (number of basic types). Basic phenomenal qualities are mutually irreducible properties (and thus multiply-instantiable universals, rather than particulars), but might be thought sufficiently akin to one another that they form a single basic type. In the latter case they would offend only against quantitative parsimony, which many consider less important.

¹⁴ Moreover, only small-palette approaches are compatible with ‘Russellian’ approaches to the problem of mental causation, on which physical properties are analysed as complex dispositional roles, for which phenomenal or proto-phenomenal properties provide the intrinsic/categorical basis. For it seems that the fundamental physical causal properties are few in number.

experience redness and yellowness in the right relation to blend, *via* the visual field, but cannot experience loudness and yellowness in the same relation.

Inclusionary combinationists will regard the blending here as intra-subject, since the ingredients experienced by the parts will be experienced also by the whole, while exclusionary combinationists will regard it as inter-subject. Yet in either version it faces 3 potential objections:

1. Phenomenal qualities cannot be blended at all (thus the ‘palette’ metaphor is misleading from the beginning);
2. Even if some can, many others are knowably basic and unblended;
3. Even if all our qualities might be blended, there is no suitable set of basic qualities out of which all could be blended.

In the remainder of this section I consider how combinationists might address these objections. In subsection 2.2, I argue that we have candidate examples of phenomenal blending in our experience, and no compelling argument for denying them that status. In subsection 2.3, I argue that we need not restrict the possibility of blending to only some familiar qualities, and in subsection 2.4 I argue that we have no good reason to rule out systematic and ubiquitous blending, of the sort posited by the small-palette hypothesis. Finally, in subsection 2.5 I discuss what relations might actually bring about phenomenal blending.

Subsection 2.2: Putative Examples of Blending

I think the best way for combinationists to establish the possibility of phenomenal blending is to claim that there are actual cases where we are distinctly acquainted, on different occasions, with both the ingredients and the resultant, and can ‘just see’ that the one is a combination of the others. We are thus directly acquainted with an explanatory relation by which a limited palette of qualities could generate more.

The examples most often appealed to involve colours. Lewtas suggests that orange experiences result from blending red-experiences with yellow-experiences (forthcoming, p.54); in a similar vein Chalmers writes that “If the same entity simultaneously is aware of a degree of redness and aware of a degree of whiteness (at the same location), it is plausibly aware of pinkness (at that location).”(forthcoming-a, p.26) This accords with the historical popularity of what Mizrahi (2009) calls a “phenomenalist’ view of colour composition” (p.2), on which ‘binary’ colours like orange and pink appear different to us from ‘unitary’ colours like red and blue. Note that this is a candidate for blending because the component colour experiences are strongly confused, even though they are not strongly confused. We can in fact focus on the reddishness of orange, and then alternatively on its yellowishness (so if they are components of it they are not strongly confused), but we could not do so if we had not on other occasions experienced red and yellow separately, and thereby learnt to recognise them - orange itself does not present us with these separate things to attend to (so if they are components they are strongly confused).

Another candidate is aromas, tastes, and flavours – the flavour of a given food or drink being a blend of tastes and aromas provided by its ingredients.¹⁵

Is there blending across modalities? One example might be the two components of pain which neuropsychology has shown to be dissociable – the affective-motivational (which makes pain feel bad) and the sensory-discriminative (which lets us distinguish pain from other bad feelings, and assign it a definite cause and location).¹⁶ An affect and a sensation are of distinct sorts, but blend so seamlessly

¹⁵ Psychologists, following McBurney 1986, have distinguished three ways for sensations to combine: analysis, when “two stimuli mixed in a solution keep their individual qualities of sensation”, synthesis, when “when two stimuli that have been mixed in a solution lose their individual qualities in order to form a new (third) sensation” (Auvray & Spence 2008, pp.1019-1020), and fusion, when “sensations [are] combined to form a single percept [which] ... remains analyzable into its constituent elements even when otherwise perceived as a whole” (Prescott 2012). The sort of cases that I have in mind are both what these schemes call ‘synthesis’ and what they call ‘fusion’.

¹⁶ Schilder & Stengel 1928, Ploner et al. 1999, Grahek 2007

into familiar pain that we find it hard to imagine them in isolation, and are surprised when we hear of subjects with one but not the other.

There may well be more examples, but these are enough to make the case; if they do not, it is unlikely that more would. None of them are conclusive: a Jamesian would deny that they involve blending at all. But they provide the combinationist with a strategy for making blending intelligible. Intuitively pink is just red and white together, and it makes sense that experiencing red and white together should constitute an experience of pink rather than of, say, green, or sourness. Note, however, that even if these examples are accepted, they will be cases of intra-subject blending (case where we experience both ingredients and resultant ourselves). The possibility of intra-subject blending may suggest, but does not entail, the possibility of inter-subject blending (as the exclusionary combinationist requires). Does a case where one part of me experiences red and another part experiences white make it just as intelligible that I must experience pink? The exclusionary combinationist needs to claim that it does, because in that case we experience the composite experience that the white and the red compose. This requires them to claim that we can abstract the relevant relation among qualities from its context within a single subject's experience, and I am uncertain of whether we can; nevertheless I at least cannot see any decisive reason to think we cannot.

Subsection 2.3: Arguments Against Blending

It is sometimes suggested that blending can be conclusively ruled out *a priori*. Here is a representative passage from James:

I find in my students an almost irresistible tendency to think that we can immediately perceive that feelings do combine. "What!" they say, "is not the taste of lemonade compounded of that of lemon *plus* that of sugar?" This is taking the combining of objects for that of feelings. The

physical lemonade contains both the lemon and the sugar, but its taste does not contain their tastes, for if there are any two things which are certainly *not* present in the taste of lemonade, those are the lemon-sour on the one hand and the sugar-sweet on the other. These tastes are absent utterly. (1890, p.158)

What is James's reasoning here? If he is relying on his general compositional nihilism, on which phenomenal qualities do not combine because nothing does (Cf. Chapter 3, subsection 1.1), then that should not convince us if we are unconvinced by that general view about composition.

But perhaps James means that because the sourness of lemon is subtly changed by being mixed with the sweetness of sugar, it is not strictly present in the blend. In most contexts this would be fallacious, since part-whole relations often involve the parts affecting each other, but there may be a special reason for objecting to such mutual adjustment in the phenomenal case, namely the principle of *phenomenal essentialism*. In chapters 3 and 4 we encountered this principle as part of the incompatible characters argument against sharing; we might employ a similar argument here. If how a quality is experienced is essential to it, and it is experienced differently in different contexts, then it is numerically distinct in those different contexts (Cf. Mørch 2014, p.154 fn19). Hence though parts are often changed by being in a certain whole, phenomenal qualities cannot be, because any phenomenal change makes them a different quality.

If we grant this argument from phenomenal essentialism, and suppose that in tasting lemonade the sweetness and sourness are phenomenally altered in some subtle fashion, then the taste of lemonade cannot be a blend *of the very same qualities* as are experienced in other circumstances. But the taste of lemonade may still be a blend; its ingredients may be the subtly-different 'counterparts' of the sweetness and sourness experienced in other circumstances.¹⁷ No plausible version of phenomenal

¹⁷ Does this threaten mind-brain isomorphism? No, if these changes to phenomenal character mirror the physical changes one neurone causes electrically in another. Is it problematic that this replaces one quality with another, but

holism can deny that we often experience phenomenal qualities, in different contexts, which are at least similar enough to warrant us calling them ‘the same’. And this same near-identity can be used to make sense of what James’s students thought: that ‘the same’ qualities are present in the lemonade-blend and in isolated experiences. Thus even if this argument succeeds, it merely constrains which blending claims we can make, without ruling out blending in general.

A final reading is that James offers the following argument:

- (a) It is not true that ‘I am experiencing the sourness of lemon’
- (b) *Therefore*, the sourness of lemon is absent from my experience

Hence blending is incoherent, simply because what it is like to experience the whole (the taste of lemonade) is not the same as what it is like to experience a part (sourness).

The fallacy here is to equivocate between systemic and additive senses (Cf. chapter 2, subsection 4.2ff). In the systemic sense of ‘experiencing the sourness of lemon’ (as meaning ‘has an experience of lemony-sourness as their sole taste experience’), claim *a* is true, but does not imply claim *b*. In the additive sense (as meaning ‘has a taste experience of lemony-sourness, perhaps among others’), claim *a* is question-begging, for if the taste of lemonade is a blend then the subject *is* experiencing lemony-sourness (blended with something else). Hence there is no sound argument from *a* to *b*.¹⁸

Hence I see no compelling reason to think blending impossible. That does not mean that blending occurs: we might still insist that while the lemonade-experiences arise from the co-occurrence

does not replace a neurone with another? No; isomorphism demands merely that some fine-grained physical property or event be replaced with another.

¹⁸ Similar remarks apply to the apparent truism that nothing can display two colours at once to the same observer – nothing could both look red and look white at once. In one sense of ‘look red’ and ‘look white’, nothing can do both, but this is because to ‘look red’ in this sense definitionally precludes displaying any other visual qualities. But in another sense, looking both red and white might just be ‘looking pink’. Pink things look red, but unlike the things we tend to call ‘red-looking’ (in the systemic sense), they also look white.

of the physical processes that independently produce sweet-experiences and sour-experiences, they are not literally composed of those two. Sweet-experience, sour-experience, and lemonade-experience might be mutually irreducible. But while I cannot demonstrate that these examples do or do not involve phenomenal blending, it is enough for the defence of combinationism that there is no bar to positing that they do.

The Lewtasian position allows for phenomenal blending in some cases, but still rejects the small palette hypothesis on the grounds that many *other* qualities we experience are knowably unblended. One way to frame this would be to distinguish qualities that display a ‘phenomenologically composite’ character from those that display a ‘phenomenologically simple’ character: only the former can be blends. For instance, maybe pink is visibly a mixture of white and red, but white and red are both positively experienced as simple, and hence cannot conceivably arise through blending.

I cannot directly refute this claim, but I think it is at least as plausible that ‘phenomenologically simple’ character is simply our having no idea of, or a confused idea of, the ingredients in a blend. Often a quality initially seems simple and unanalysable – until further experience lets us discern the components within it. Dennett describes an auditory example of this phenomenon, in which the sound of a chord played on a guitar appears simple and pure to the untrained ear, but comes to seem composed of distinct notes when one is familiar enough with the notes individually to recognise them in the mixture (1991, pp.73-74). In a similar vein, wine tasters often say that with practice, one learns to discriminate the different components of a wine’s taste (the above example of pain is also pertinent). And research showing that, e.g. untrained subjects frequently construe certain odours as increasing the sweetness of a taste, while trained subjects do not (Bingham et al. 1990), reinforces the point that we are often fallible in distinguishing different sensations (cf. Chuard 2007). Perhaps in all these cases, we were mistaken in perceiving a ‘phenomenologically simple’ character to begin with; but then how sure

can we be that there is such a character in the case of, say, redness? It seems equally reasonable to think that all qualities seem phenomenologically simple until we can discern their ingredients – so that the apparent simplicity of a given quality does not warrant denying that it has ingredients.

Subsection 2.4: From Limited Blending to Ubiquitous Blending

Suppose, then, that all the phenomenal qualities which we experience are such that they might be the resultants of blending. There remains the problem that there do not appear to be any known qualities that could plausibly be ingredients for *all* our qualities in the way that, say, redness and whiteness are ingredients for pinkness. McGinn expresses this concern when he writes that:

We cannot [...] envisage a small number of experiential primitives yielding a rich variety of phenomenologies... [for] you cannot derive one sort of experience from another: you cannot get pains from experiences of colours, or emotions from thoughts, or thoughts from acts of will. There are a large number of phenomenal primitives. (McGinn 2006, p.96)

McGinn is probably right that we cannot reasonably hope to get all qualities from any small set of *known* qualities, but the combinationist need not think that the basic ingredients are known to us. Instead, the basic ingredients may be ‘alien qualities’, unimaginable but not inconceivable. It is a commonplace that there are such qualities: just as a human born anosmic cannot imagine olfactory qualities, we are all similarly limited regarding the qualities of the many sensory modalities that humans lack. We can entertain and accept the existence of such qualities, but we cannot ‘know what they are like’.

Presumably, if familiar qualities can blend, so can alien ones. But can they blend *into familiar qualities*? For instance, might the familiar phenomenal quality of redness be a blend of two alien phenomenal qualities (call them AQ1 and AQ2)? If so, maybe all our phenomenal qualities result from blending, even when we cannot identify their ingredients.

However, AQ1 and AQ2 cannot be unimaginable in quite the same way as standard examples, involving tetrachromatic vision, or bat sonar. Our inability to imagine the latter corresponds to our inability to experience them. We *lack something*, phenomenologically (and neurally) speaking. But we *do* experience AQ1 and AQ2, whenever we have experiences of red: we *lack* nothing. How, then, can they be unimaginable?

In one sense, we *can* imagine AQ1 and AQ2, just by imagining redness. But when we do so, we cannot *separate* AQ1 from AQ2. They are imaginable together, but not *distinctly* imaginable. We do not know what they feel like on their own – not because we lack necessary resources, but because we cannot deploy one resource without also deploying another. That is, they are unimaginable on account of being robustly confused relative to the mental operation of imagination. Call this ‘unimaginability by confusion’.

For a more mundane example of unimaginability by confusion, consider an arachnophobe trying to imagine how their friend, who finds large furry spiders adorable, perceives a tarantula. This imaginative task may be impossible for them, but not because they lack anything. They can imagine spiders, they can imagine finding something cute, and they can connect these imaginings together. The problem is that they cannot generate an image of a spider without also generating a feeling of intense fear and revulsion, which would constitute a failure to imagine their friend’s experience.¹⁹

Even accepting the possibility of both phenomenal blending and unimaginability by confusion, it may still seem that the different qualities we experience are too radically heterogeneous to be blends of the same ingredients. But our ability to recognise two things as akin to one another often depends on

¹⁹ The arachnophobe’s imaginative inability is ‘shallow’, since the right sequence of experiences could let them imagine a spider without feeling fear (it is also asymmetrical). By contrast, alien qualities are ‘robustly’ (and symmetrically) unimaginable-by-confusion: only a profound transformation, possibly requiring gross physiological re-organisation, would let human distinctly imagine AQ1 or AQ2. Since blending by definition involves *strong* confusion, this is a case of *radical* confusion.

our ability to recognise and attend to the features they share, and if we cannot pick out their shared features we may wrongly feel that they are entirely unlike; musical and taste training provide many examples. Hence because we cannot recognise or attend to the basic ingredients, we may get a false impression of radical heterogeneity.

Inability to pick out shared features does not always stop us registering similarities. Sometimes two things ‘seem alike’ in some way, without us being able to say how. But this kind of inarticulate resemblance is commonly encountered among experiential qualities: we frequently describe qualities of one modality using terms drawn from another (warm, harsh, sweet, soft, loud, etc.), or use sensory terminology to describe emotional or cognitive phenomenology. It is an interesting question what determines whether a particular shared feature generates such an inarticulate intuition of similarity. Perhaps we have neural mechanisms designed to identify similarities, which can be activated just enough to produce some recognition of similarity but not enough to identify what the similarity is. Whether such a mechanism is activated or not might depend on subtle details of the wiring among different brain areas, and of what exactly the similarity-detecting mechanism is sensitive to.

Thus combinationists can allow for three categories of resemblances among blended qualities: those we can articulate by identifying the common element (e.g. the negative valence in a pain and an itch, or the redness in orange and purple), those we cannot articulate but only vaguely intuit (e.g. between redness and warmth), and those we do not register as similarities at all because we cannot imagine the shared ingredients distinctly.²⁰ If so, there remains no principled objection to the small-palette hypothesis.

²⁰ For an extensive discussion of the kinships we can recognise between qualities in different modalities, see Coleman forthcoming, pp.43-47, drawing on Hartshorne 1934, pp.35ff. Cf. Pierce 1998 p.35

Subsection 2.5: What is the Right Relation for Blending?

The small-palette hypothesis mentions ‘the right relation to blend’; defending the hypothesis does not require specifying this relation, but to enrich the proposal I will briefly suggest a candidate specification. My proposal is that phenomenal unity and strongish confusion are jointly sufficient for phenomenal blending: whenever I experience two qualities together but cannot distinguish them in any way, I experience them as blended, and more broadly whenever two experiences are unified, but are strongly confused for some subject, that subject experiences their blended composite (even if they do not experience the ingredients themselves).

Call this the ‘Blending-As-Default’ extension of the small palette hypothesis. It claims that the negative part of the definition of phenomenal blending (inability to distinguish), together with phenomenal unity, is sufficient for the positive phenomenology of two qualities forming a third. It need not be adopted, but it has some plausibility, and is attractively straightforward. To see its plausibility, first note that phenomenal unity definitely seems like a prerequisite of blending; it would be hard to experience two qualities as a single quality, if they were not ‘experienced together’. And confusion was part of the definition of blending. Given these two necessary conditions, it is not clear that anything more is required.

Consider colours. If redness and whiteness are experienced at two different points in the visual field, there is no experience of pinkness. To blend they must stand in the relation ‘experienced-in-the-same-location-as’. But what is this relation but the absence of any experienced spatial differentiation? If they were experienced at different points, they would thereby be made distinguishable, not confused, for it seems constitutive of experiencing things as occupying different points in visual space that we can pick out the one point, and what occupies it, separately from the other. But then to be strongly confused, the qualities must be experienced as co-located; the

Blending-As-Default hypothesis holds that if they are both experienced as in the visual field at all, then being strongly confused is sufficient for being experienced as at the same location, and thereby being blended.

Blending also seems to occur in those aspects of experience which do not distinguish multiple objects: sense-modalities like smell which encode little spatial information, and affective phenomena like mood. It seems characteristic of odours and moods to merge and interpenetrate rather than being compartmentalised, which suggests that blending occurs whenever two qualities are experienced together, but not separable. What is hard is not to blend two qualities but to simultaneously instantiate them *without* them blending: this requires the mental infrastructure to direct some mental operation onto one while making sure not to direct it onto the other.

Even given this suggestion as to the right relation, I have still said nothing positive about what the basic qualities might be, or which familiar qualities contain which basic qualities in what proportions. But I do not think it is incumbent on panpsychist combinationists to do so, any more than it is incumbent on philosophical defenders of atomism to say what types of atom there are and what features they have. They can leave that task for empirical science, in particular for psychophysics, neuroscience, neuroethology, and so on. These disciplines can look for patterns of resemblance and difference among the qualities experienced by humans, and the correlated patterns of resemblance and difference among their brain states; as brain technology becomes more sophisticated, they can also explore the phenomenal consequences of deliberate brain modification. It is harder to make progress in this regard than it is with the basic physical properties, but that is a natural consequence of the subjectivity of phenomenal qualities.

A related issue is whether the Small-Palette hypothesis is objectionably ‘mysterious’, in appealing to basic qualities which are unimaginable to us. Certainly, without any further detail this

hypothesis is not so much an explanatory *achievement* as a postulate that things are in principle explainable – rather like the atomic theory of matter is not in itself an explanatory achievement without any detail as to the varieties of atom, their relations to each other, etc. But it is not objectionable for a theory of the mind to postulate many unimaginable things. *Inconceivable* postulates would be objectionable, since they would undermine the theory’s logical coherence. But if we accept the irreducibility of the subjective and private to the objective and public, then we should already be committed to a vast range of unimaginable experiences. The nature of subjectivity is precisely that we can theorise about all of it from outside, but directly know only a tiny fragment.

So panpsychist combinationists should claim that by comparing the qualities we distinctly experience, we can identify a relation of intelligible composition holding among certain of them. If this relation exists, it is a live possibility that it obtains between the qualitatively sparse world of microexperience and the qualitatively diverse world of macroexperience.

Section 3: The Mismatch Problem

Lockwood expresses the mismatch problem by saying that, “the structure we do encounter at the phenomenal level seems not to match, even in coarse-grained fashion, that of the underlying physiology, as revealed by scientific investigation” (1993, p.544). Chalmers’ formulation is that “macrophenomenal [macroexperiential] structure is distinct from macrophysical structure” (forthcoming-a, p.13), even though macroexperiential structure is supposedly constituted by microexperiential structure, which is isomorphic to microphysical structure, which is what constitutes macrophysical structure.

This presupposes that we know what microphysical, macrophysical, and macroexperiential structure are. It is actually not easy to say with any comprehensiveness what kinds of structure are in question, but we can canvas some examples: macroexperiential structure “involve[es] the complex spatial structure of visual and auditory fields, [and] a division into many different modalities” (Chalmers forthcoming-a, p.5), while macrophysical structure is a mathematically-described structure in which “three-dimensional spatial arrangement, and changes therein, seem central” (Lockwood 1993, p.544). Then there is the structure of properties, such as “the scalar structure of mass [and] the three-dimensional structure of color space” (Chalmers forthcoming, p.13), and the attentional structure of focus and periphery.

We can express the problem by asking: ‘Why is macroexperiential structure not isomorphic to microphysical structure?’ But this question will have different force on different approaches to composition. On the exclusionary approach to macro-micro relations, we need not expect isomorphism between levels: wholes of various sorts are dependent on, but still importantly different from, their micro-level bases. There is still an interesting question of why macroexperiential wholes have the particular sort of structure they have, but we have no particular reason for expecting them not to: thus the question might be answered by a shrug and a suggestion that more research needs to be done. By contrast, an inclusionary approach to micro-macro relations, closer to that of classical mereology, posits a whole as somehow reflecting everything about, perhaps even being identical to, its parts. An inclusionary combinationist thus faces a more urgent threat: if they cannot explain the structural mismatch, they cannot maintain that macroexperience is constituted by microexperience. So any form of combinationism is stronger if it can explain the failure of isomorphism between macroexperiential and microphysical structure, but only for inclusionary forms is such an explanation urgently necessary.

Subsection 3.1: Why Combinationists should Privilege Informational Structure

Chalmers considers and rejects two approaches to the mismatch problem. One is to say that the macroexperiential structure which appears to conflict with macrophysical structure is in fact not present in experience itself but only in its objects: that is, experience represents things as having that structure, rather than itself having that structure. For example, one might maintain that experience itself is not divided into the visual and auditory sections, but only *represents* external things as having both visual and auditory qualities. Versions of this proposal appear in Clark 1989, and Stoljar 2001, but face a number of problems; first, we might think that an experience's representational content is part of its structure, and thus itself in need of explanation; second, even if it is not, it might still seem that experience itself displays the relevant structure (as claimed by Alter & Nagasawa 2012, p.91²¹); third, any attempt to give a general reduction of experiential structure to represented structure faces challenges resulting from illusion and hallucination: external objects cannot supply structure to, say, dream experience.

A second approach that Chalmers considers and rejects is to claim that we can find straightforward exemplifications of macroexperiential structure in the brain, for example in the 'retinotopic' areas of striate cortex (see Holmes 1944, Engel 1997). However, only some of the brain is organised like this, and seeing how even the rest of the visual system is organised serves to bring home the force of the mismatch problem more vividly. Most visual areas handle, not a portion of the visual field, but a certain aspect of its content, such as colour, movement, or shape (see Zihl et al. 1983, Heeger et al. 1999, Théoret et al. 2002, Anzai et al. 2007). If we pick out one part of the visual field (e.g. that occupied by a rotating red triangle), and ask which brain area is responsible for it, we cannot

²¹ This objection may be less compelling to inclusionary panpsychist combinationists who have already accepted a response to the blurring problem based on radical confusion: for they can explain the impression of ultimate homogeneity in experience as a misinterpretation of the radical confusion of visual experiences with each other.

name any single area, but will have to describe a set of contributions – that (say) Area 1 is responsible for its rotation, Area 2 for its redness, Area 3 for its triangularity, etc.²²

A third, more promising, approach is to say that macroexperiential structure corresponds to a particular aspect of macrophysical structure, namely *informational structure*, the structure of information transfers among physical parts of the brain. This is a specific sort of causal structure, picked out by prioritising how sensitive each element is to the fine details of another's state. I follow Chalmers in thinking that “something like this has to be the best option for the panpsychist” (p. 29), because it is antecedently plausible (and has become more plausible as neuroscience has advanced) that the structure of experience depends not on the spatial location of brain regions but on the informational relations among them. So in this section I will explore whether this proposal can be made to work.

There is a further reason for exploring the proposal that macroexperiential structure corresponds to informational structure, to the exclusion of other aspects of physical structure: doing so seems necessary to avoid the *dancing qualia* problem. This problem (noted briefly in chapter 1, subsection 4.1, and in the introduction to this chapter), is based on the thought that since it is plausible that we can implement the same functional roles in different sorts of material (carbon, silicon, etc.), it should be possible also to change the underlying material out of which some subject is made without affecting their functional structure (thought experiments of this sort are described in Chalmers, 1995b). But if their functional structure is unchanged, then their tendency to register or report a change in their experiences is unchanged, and so they will not in fact report any change in their experience. It seems bizarre and implausible that someone's experience should change ‘before their eyes’, and yet they be entirely unable to detect this change; thus we should avoid accepting that such a thing would result from the functionally-neutral change in components. Thus we seem pushed towards the conclusion,

²² This organisation gives rise to what is called ‘the binding problem’ (see Treisman and Gelade, 1980, Duncan and Humphreys, 1989, 1992, Treisman and Sato 1990) – when A1 registers rotation, A2 redness, and A3 triangularity, how does the brain tell that it is the triangle which is red, or that it is the red thing which is rotating?

either that systems functionally equivalent to a conscious subject cannot be made out of different materials (the conclusion recommended for panpsychists by Sebastien 2013), or that no feature of a component subject except its functional contributions, that is to say except its contributions to the overall information processing of the whole, can make a difference to the whole's experiences.

While none of the above arguments are clearly decisive, together they provide a powerful case that the combinationist should hold macroexperiential structure to be isomorphic to informational structure, and that other aspects of physical structure (such as spatial position) are relevant to experiential structure only indirectly, insofar as they influence informational structure. But is this claim defensible? Chalmers worries:

From the perspective of physics, high-level information structure[s] are derivative aspects of a more encompassing and more basic macrophysical structure. It is not easy to see why [...] macrophenomenal properties should have this structure rather than the more encompassing and more basic structure. (forthcoming p.29)

(Note that the appeal to what is 'more encompassing and more basic' is likely to be more persuasive to inclusionary than to exclusionary combinationists.) In the next subsection I explain why informational structure is privileged, appealing primarily to the special prominence of informational structure in attention.

Subsection 3.2: Confusion and Dissociation, Blending and Refraction

Section 1's discussion of 'confusion' already privileges informational structure, since confusion is itself an informational relation: two elements are confused when information about their individual features cannot be extracted from information about them together. It will be useful to relate this notion to chapter 4's discussion of the 'phenomenal field'. My experiences seem to me to hang together as a single field, organised around the centre of my focal attention. Putting the mismatch problem in terms

of this idea, we can say that the structure of this field seems not to match many major aspects of physical structure, such as spatial structure. One half of the phenomenal field does not correspond to one spatial half of anything, and to that extent does not correspond to any spatial part of the brain.

In chapter 4, subsection 1.3, I suggested that we think of ‘distances’ in the phenomenal field as measures of ‘attentional proximity’, the tendency of one potential target of attention to bring another to attention. I believe that this analysis implies that radically confused elements, if they are unified, must be co-located in the phenomenal field of the subject for which they are confused. To see this, recall that to count as radically confused, two underlying experiences must be such that whenever their subject attends to one, they cannot help but attend to the other. But this amounts to saying that each is disposed to infallibly and immediately transfer attention to the other, and this is simply the highest degree of attentional proximity. Thus if attentional proximity corresponds inversely to distance in the phenomenal field, perfect proximity means minimal distance.

Supposing that radically confused experiences sharing a field will occupy the same position in that field, we get the result that phenomenal blending, since it requires confusion and unification, involves two experiences appearing in the phenomenal field as a single item, a single ‘point’. I think this is a welcome result: blended experiences appear to us as a single experiential element, a single potential target for attention, a single ‘item’ in consciousness.

Phenomenal blending can explain one form of structural divergence - items which are distinct at one level can function as a single item at a higher level, so that the discernible structure of macroexperience lacks certain kinds of divisions which are present at the micro-level. This suggests a fuller explanation for the special role of experiential structure: experiential structure determines the structure that we find in experience because it determines our capacities for finding this structure - that is, the structure experience has ‘for us’ is tied to the ways that we can attend to items in our experience,

and more broadly the ways that we can cognitively access our experience. Because blended experiences are not accessible through distinct attentional acts (i.e. because we cannot distinguish them), they do not strike us as distinct.

So the combinationist can say that the way we find experience to be structured is determined by informational structure because informational structure determines the different ways that we can attend to and cognitively access our experience: access requires information to flow. But for this explanation to work, we would need to explain not only how micro-level divisions are glossed over, but also how macro-level divisions can appear. How, for instance, can activity in a single brain part appear to the whole person as two distinct experiential elements, perhaps at far distant points in the phenomenal field, which they cannot identify as based in the same part? To contrast with ‘phenomenal blending’, call this ‘phenomenal refraction’. Only by making sense of both phenomenal blending and phenomenal refraction can the combinationist explain why informational structure is privileged in macroexperience.

To understand how phenomenal refraction is possible, consider a component subject which has two or more separate sets of properties which are independently sufficient for its experiences to be attended, or more broadly accessed and operated on, by the composite subject. Each set of properties allows it to broadcast information to its surroundings in a way that other component subjects can make use of. These two independent sets of properties might involve, for instance, waves of electrical activity synchronised to different frequencies, or carried along different sets of connections, or transmitted by different forms of energy (e.g. light and sound). Suppose, moreover, that the information broadcast along these two channels was more or less independent: changing the one had little effect on the other. (There are limits to how far this can be true – destroying the component subject will presumably affect

both simultaneously – but they might nevertheless be highly dissociated in many contexts.) Call this ‘attentional dissociation’.

In the whole’s phenomenal field, these two distinct ways to attend to the experiences of that part may be at a greater or lesser attentional distance from each other, depending on the extent to which attending to that experience in one way (accessing certain features of it) disposes us to attend to it in the other way (accessing other features of it). Since there is no guarantee that attending in one way will make the other particularly salient, the two ‘attentional targets’ may be at a significant attentional distance, and thus occupy two distant points in the phenomenal field. Indeed, they may not even share a phenomenal field: they may not fulfil the conditions for phenomenal unity (adumbrational or otherwise). But let us assume that they do share a field, but occupy distinct locations in it. I take this to merit the label of ‘phenomenal refraction’, since these two ‘attentional targets’ (corresponding to the two ways of attending) will seem to the whole to be two distinct experiences. Yet they together are grounded in the consciousness of the same microsubject.

Given the possibility and intelligibility of both phenomenal blending (confusion within a phenomenal field) and phenomenal refraction (dissociation within a phenomenal field), we can see how multiple layers of these phenomena might eventually make a composite subject’s phenomenal field radically diverge in structure from those of its microscopic components. The experiences of the parts of my brain will be refracted through multiple levels of complex informational dissociations, and these refractions will be blended through levels of confusion. The manifest structure of the phenomenal field I encounter in experience will tell me little about the underlying non-informational structure of my many component subjects (e.g. their spatial locations), despite being entirely explained by those microexperiences.

Subsection 3.3: Dancing Qualia and the Boundaries of Consciousness

Finally, there is the ‘dancing qualia’ problem. It seems bizarre that my phenomenology might change without me even being able to notice, and yet if the phenomenology of my microscopic parts plays a constitutive role in mine, such undetectable changes might be possible. If, for instance, there is some qualitative difference in the phenomenology associated with carbon atoms and that associated with silicon atoms, that should yield a difference in the phenomenology of functionally equivalent carbon-based and silicon-based brains. But this difference will be undetectable because any difference in reports or responses to measurement would be a functional difference.²³

A particularly vivid expression of this thought, (though not intended as a critique of combinationism), is Block’s (1992) thought-experiment, in which tiny aliens smaller than our elementary particles contrive for enigmatic reasons to simulate our elementary particles by building spaceships that interact with each other in just the way such particles do. We are to suppose that a human astronaut lands on a planet made out of such ‘ersatz matter’, and gradually replaces all the matter of their body with such ships, just by innocently eating and breathing. Block’s aim is to make Anti-Nesting seem absurd by noting that it seems to imply that this unnoticed change should deprive the astronaut of all consciousness; but an opponent of combinationism might ask how the combinationist avoids the similarly strange consequence that this unnoticed change in the astronaut’s parts should make some sort of difference to their experience.

I believe the combinationist can answer this objection by appealing once again to the link between information and attention, though the response will take different forms on exclusionary and

²³ Of course, if ‘functional’ differences could include differences such as ‘produces a certain experience’ vs. ‘does not produce that experience’, then this problem could not arise, for there any two systems that had different experiences would *ipso facto* be functionally different. But the notion of ‘functional’ at play here is one which is defined primarily in terms of inputs and outputs to the system; its internal states are defined only implicitly, in terms of the ways they mediate between inputs and outputs. Thus no reference to the intrinsic qualities of internal states can make their way into the functional description.

inclusionary approaches. The exclusionary combinationist has an easier task, for they can allow for the experiences of carbon or silicon atoms to constitute my experience without being literally present in it. They can maintain the principle that Chalmers calls ‘organisational invariance’:

Organisational Invariance (OI): “experience is invariant across systems with the same fine-grained functional organization” (Chalmers 1995b, p.310)²⁴

The task for the exclusionary combinationist is to explain why OI holds: how functionally neutral changes in phenomenology get ‘filtered out’ of the whole’s experience.

The inclusionary combinationist, on the other hand, needs to deny OI. If different microexperiences were constituting my experience, then they would be present in my phenomenology, and my phenomenology would therefore have to be different, even if my functional organisation was entirely unchanged. However, the inclusionary combinationist can maintain a weakened form of OI, and thereby ‘tame’ the undetectable phenomenal changes that they are committed to. In place of OI, they should maintain:

Organisational Invariance of the Attended (OIA): attended experience is invariant across systems with the same fine-grained functional organization.

They could then claim that the only undetectable changes in phenomenology they are committed to are *unattended* ones, and that such changes are not at all implausible. Indeed, Chalmers explicitly recognises that OIA captures the same intuition as OI:

²⁴ What is ‘fine-grained functional organisation’? Chalmers says that it is whatever level of organisation is “fine enough to determine the behavioral capacities and dispositions of a cognitive system” (p.310). One might worry about the use of ‘determine’ here: a silicon brain and a carbon brain might produce subtly different behaviours in, say, very hot environments, or when held in a centrifuge. But the notion in question has significant intuitive appeal, and the responses I discuss will not depend on spelling it out more rigorously. This problem, of how ‘functional organisation’ talk can distinguish relevant from irrelevant causal differences, is similar to the point of Lycan’s ‘New Lilliputian Argument’, discussed in chapter 6, which seeks to show that functionalism needs a substantive criterion for carving out the relevant inputs, outputs, and internal states from irrelevant ones.

[T]he argument leaves open the loophole that *unattended* qualia might be invertible. If we are not attending to the fringes of our visual field, for example, a qualia inversion might take place there without our noticing. (1995b, p.326, original emphasis).

The problem is that it seems *ad hoc* to simply declare that all the changes produced by a functionally neutral switch of carbon with silicon would be unattended, as Chalmers notes:

But to exploit this loophole would leave one in the unattractive position that qualia are organizationally invariant when they are central enough in one's attention, but dependent on other features when they are not. (Presumably an inverted green experience on the fringe will flip back to red when one attends to it?) Such an asymmetric position would be theoretically unsatisfying in the extreme. (1995b, p.326)

Before addressing this concern, consider a more aggressive argument, to show that the undetectable phenomenal change *must* occupy the focus of attention. Whatever process in my brain is at the focus of attention constitutively involves carbon atoms; we could replace those atoms with silicon while keeping that process at the centre of attention (since any difference in attention is a functional difference), so surely any resultant phenomenal change should pertain to the phenomenal character of that focally attended process.

The idea of phenomenal refraction illuminates the flaw in this argument. When we speak of 'whatever process in my brain is at the focus of attention', we confuse two levels: we try to both carve brain activity into discrete underlying chunks, and also to pick out the specific target of attention. But, I have argued, the specific target of attention is not any of the underlying parts, but a certain information stream whereby certain information about one or more parts is made available to the other parts. So we need to look at the properties of one or more processes in virtue of which information about them is made available. These properties will plausibly be ones that make a functional difference, not ones that are functionally inert (like the properties constitutive of the difference between carbon and silicon). And if the properties of experiences that allow them to be attended are dissociable from the carbon-silicon

difference, as the hypothesis supposes, then they will be phenomenally refracted and appear as distinct items in the phenomenal field.

We can extend this answer to address Chalmers' worries by noting that, given the account of phenomenal refraction, it is not at all *ad hoc* that attended experiences, even experiences moderately close to being attended, are unchanged. Changes in *their* phenomenal character would require changes in those properties of the components which allowed them to project to focal attention, which will necessarily be functionally relevant ones. The very fact that the changes are undetectable entails that they occur 'far from' the attentional centre. Moreover, we can answer the question 'will a peripheral experience 'flip back' when one attends to it?' by saying that for all intents and purposes, these experiences *cannot* be attended, without internal reorganisation so extreme as to make it perfectly plausible that their phenomenal character would be changed thereby.

This notion of the 'unattendable' requires more explanation, especially since I earlier spoke of points in the phenomenal field as essentially potential targets of attention. We should first recognise that the visual field is a misleading analogue here, because its periphery is marked by a definite boundary, fixed by the anatomy of retina and eye socket. Without this kind of boundary, the periphery of the phenomenal field might extend extremely far, to points at vast phenomenal distances from the focus. Since phenomenal distance reflects attentional proximity, saying there are 'vast distances' amounts to saying there is a miniscule degree of attentional proximity: when attending to some focal experience, what is in the far far periphery has barely any tendency to enter attention. This could be thought of as a lowering of probability, so that the odds of that experience being attended are negligible, or alternatively as an increase of difficulty, in that it is so hard to attend to the experience in question that it could only be done by major internal reorganisation, such as might require therapy or surgery. Whether we think in terms of probability or difficulty, what is important is the idea that there

is only a difference of degree between these ‘unattainable’ experiences and experiences which are merely ‘hard’ to focus attention on, due to lack of salience or (situationally) due to distractors, tiredness, etc. They still, however, stand for potential acts of attention – just acts of attention which are inordinately unlikely, for the actual structure of the brain is completely ill-suited to perform them. I will describe such experiences as ‘realistically unattainable’, though attainable in principle.

This talk of unattainable experiences will sound suspicious to anyone who denies the possibility of phenomenal overflow. Can there really be elements in my experience that I cannot focus my attention on? But the availability of both inclusionary and exclusionary versions of combinationism lets the combinationist hedge their bets. If realistically-unattainable experiences are conceptually impossible, then that constraint itself will provide the exclusionary combinationist with a rationale for saying that any feature of the underlying parts’ experiences that is realistically unattainable (so that were it to be in the whole’s phenomenal field, it would occupy this distant periphery) simply corresponds to no element of the whole’s experience. And they thereby avoid undetectable phenomenal changes, and provide a rationale for OI.

Conversely, if unattainable experience *are* possible – if something can be there in the periphery of my experience even though the mechanisms that guide my attention preclude it becoming attended – then the inclusionary combinationist can use them to ‘tame’ the dancing qualia objection, and provide a rationale for OIA.

Summary:

In this chapter I explored the apparent divergence of experiential structure from physical structure, through three specific problems. One of these, the mismatch problem, faces all combinationists who

think that our component subjects are physical entities of some sort, while the problems of blending and blurring are specific to views like constitutive panpsychism which build us out of microsubjects.

My proposed response to these latter two problems combines the observation of radical confusion among microexperiences from section 1, the small palette hypothesis from section 2, and the blending-as-default addition from subsection 2.5. Call this the ‘confused blends’ proposal. It comes in two versions, one exclusionary and one inclusionary. Both versions agree that the experiences of my microscopic parts are radically confused with each other for me - I cannot perform any mental operation on one of them without simultaneously doing the same to many others. Consequently, the only elements I can distinguish within my consciousness are composites, with radically confused microexperiences as parts. The quality of these composites intelligibly reflects that of their components, being a ‘blend’ of them in the same way that green is a blend of yellow and blue.

The inclusionary and exclusionary versions differ not on whether the whole experiences these confused blends, but over whether the whole also experiences their component microexperiences. On the inclusionary approach, the whole experiences them all: thus the phenomenal blending that they posit is among the different experiences of a single subject, and their part-whole relation among experiences can be assimilated to Bayesian subsumption. This commits inclusionary panpsychists to holding that our own consciousness is much more fine-grained than we might think: we mistakenly interpret the phenomenal character of radically confused experiences as indicating structural simplicity. On the exclusionary approach, by contrast, the composite subject experiences the composite experience, but not its minute parts. This means that the blending they posit is inter-subject, rather than the intra-subject sort we seem to be directly acquainted with, and their notion of experiential composition cannot be Bayesian subsumption.

For the mismatch problem, which faces all combinationists, I have again defended a proposal that has inclusionary and exclusionary versions, and which I call the ‘refracted aspects’ proposal.

Starting from the last chapter's idea of consciousness as a 'phenomenal field' organised by attention, this proposal denies that this field is built up by adding together 'phenomenal pixels', each corresponding to a different component subject. Rather, the fields of component subjects are superimposed so that each point in the whole's field corresponds to some way that the whole might direct attention.

If the same component subject has two independent ways of broadcasting information to the others, its experiences will occupy two distinct points in the whole's phenomenal field. I called this 'phenomenal refraction', and it underwrites two different responses to the dancing qualia problem, both based on the idea that functionally inert properties of an underlying experience will be phenomenally refracted away from functionally relevant properties. The exclusionary combinationist holds that the former properties are 'filtered out' of the whole's consciousness because the superficial experiences they underlie are so attentionally distant from the whole's other experiences that they 'fall outside' its phenomenal field: the inclusionary combinationist, by contrast, holds that they are experienced by the whole, but are 'realistically unattendable', lying in the 'far periphery' of the phenomenal field, and thus go unnoticed.

I think the 'refracted aspects' proposal satisfactorily resolves the mismatch problem, but the 'confused blends' proposal's success is less clear: in particular, it is not clear that the inclusionary version is compatible with how our experience is revealed to us. The viability of inclusionary panpsychist combinationism is thus still contestable.

Chapter 6: Group Minds and World Souls

<u>Section 1: Reasons for Rejecting Mega-Subjects</u>	p.230
<i>Subsection 1.1: From Basic-Experience Inheritance to Mega-Subjects</i>	p.230
<i>Subsection 1.2: Disavowed Experiences and Inconsistent Beliefs</i>	p.232
<i>Subsection 1.3: Basic-Having and Full-Having</i>	p.235
<u>Section 2: Could Mega-Subjects have Unified Consciousness?</u>	p.237
<i>Subsection 2.1: Phenomenal Unity without Dispositional Unity</i>	p.238
<i>Subsection 2.2: The Argument from Vagueness</i>	p.239
<i>Subsection 2.3: Does my Consciousness Adumbrate Yours? Boundedness and Pure Openness</i>	p.242
<i>Subsection 2.4: Social Mega-Subjects and Informative Adumbration</i>	p.244
<i>Subsection 2.5: What is it like to be a Unified Mega-subject?</i>	p.246
<u>Section 3: Human Bodies as Organic Mega-Subjects</u>	p.248
<i>Subsection 3.1: The Panpsychic Body</i>	p.249
<i>Subsection 3.2: What if my Foot Became Conscious?</i>	p.251
<i>Subsection 3.3: Why some Experiences are not Mine</i>	p.252
<i>Summary</i>	p.253

This chapter concerns ‘mega-subjects’, conscious beings which contain humans, or similar creatures, as proper parts. My primary aim is to defend the viability of versions of combinationism which entail the existence, and even the proliferation, of mega-subjects, by showing that this consequence is not a *reductio ad absurdum*. A secondary aim is to expand the range of cases that combinationism can be applied to, thereby illuminating both how combinationism works, and how we should think about various large composite entities.

First, let us review some varieties of mega-subject. One type is ‘social mega-subjects’, conscious beings constituted by multiple humans (or other social creatures) in social relations of some kind. These might be populations, institutions, nations, clubs, governments, even families or mobs. These are the mega-subjects to which it is most pre-theoretically plausible to ascribe mental states - we commonly speak of them as pursuing goals, acting on beliefs, and so on, and their possession of phenomenal consciousness has at least been discussed by philosophers (see esp. Block 1992, Knobe & Prinz 2008, Schwitzgebel 2014).

Second, there are mega-subjects whose parts are connected by some sort of biological or ecological integration - ecosystems, the whole biosphere, or perhaps multi-brained animals. Call these ‘organic mega-subjects’: there is less pre-theoretical support for ascribing them mental states, but they

are sometimes regarded as collectively ‘alive’ (cf. Margulis 1998, Lovelock 2000), and a strain of environmentalistic panpsychism advocates the idea of the world’s sentience as offering a way out of a perceived metaphysical impasse reflected in the ecological crisis (Matthews 2003). Things like colonial or eusocial animals (bees, termites, coral polyps, etc.) might also be thought of as something like organic mega-subjects, if their individual animal members were thought conscious, as might pairs of conjoined twins if they were thought to ‘share a body’, and that body was thought a candidate for consciousness. Note that experience-first and subject-first combinationists will take subtly different views of organic mega-subjects: insofar as the biosphere, say, is a physical substance with physical parts, experience-first combinationists will not say that it itself is a conscious subject, but that it is something like the ‘body’ of a conscious subject, namely the ‘biosphere-soul’ constituted by all the experiences of the biosphere’s parts.

Both social and organic mega-subjects display some special sort of organisation; this distinguishes them from ‘aggregative mega-subjects’, mere collections of items which include human beings but lack any special sort of organisation. This includes both relatively ‘natural’ aggregates, like ‘everything on planet earth’ (for subject-first combinationists) or ‘all the experiences felt on planet earth’ (for experience-first combinationists), and more ‘gerrymandered’ entities like ‘the fusion of my body and the CN tower’ (for subject-first combinationists), or ‘the sum of all the experiences of all the mammals in California’ (for experience-first combinationists).

Finally, there is the ‘cosmic subject’, the conscious whole that encompasses everything in existence; experience-first combinationists can use the term ‘cosmic subject’ for the being constituted by the totality of all experiences in existence. On most forms of combinationism, this will be merely the largest aggregative mega-subject, but holistic or monistic views would ascribe it a special position as the ground of and basis for all consciousness in its smaller parts. (Indeed, pantheistic views might ascribe it not just consciousness but supreme intelligence.)

Not all combinationists need believe in mega-subjects, but certain versions of combinationism arguably do imply a world rich in mega-subjects, and section 1 aims to show that this implication is acceptable. In section 2 I turn to the question of whether any mega-subjects might have unified consciousness, which turns on whether any human beings stand in the relations constitutive of unity. I argue that while human beings probably do not at present stand in robust and informative unity relations to each other, a Sorites-style argument might provide motivation for thinking that we are nevertheless linked by bare phenomenal unity. Finally, in section 3 I examine what combinationists should say about the idea of a human body that has independent experiences based in separate body parts, making it somewhat analogous to an organic mega-subject.

Section 1: Reasons for Rejecting Mega-Subjects

In this section I explain why a combinationist might be committed to the existence of mega-subjects, why this commitment seems so implausible, and how this implausibility is avoided by recalling the definition of ‘basic-experiential properties’.

Subsection 1.1: From Basic-Experience Inheritance to Mega-Subjects

Most of us accept the *existence* of the various entities listed in the introduction to this chapter (ecosystems, nations, sets of experiences, etc.) but normally we do not take them to be conscious. But if experiential properties are unconditionally inherited by wholes from their parts, then accepting the existence of these large composites will commit us to regarding them as mega-subjects. This is especially worrisome with aggregative mega-subjects, for unlike with social or cosmic mega-subjects, nobody has any pre-theoretical inclination to ascribe experiences to these beings.

Not all combinationists accept unconditional inheritance: exclusionary combinationists, who provide a different intelligible rationale for *conditional* experience inheritance, can agree with the commonsense view that all these large composite entities, though they contain and perhaps ontologically underlie (or are underlain by) many conscious experiences, are not conscious and do not *have* those experiences. An exclusionary combinationist might nevertheless think some of them are mega-subjects - a pantheist might make a case for the existence of a cosmic mind, a deep ecologist for the consciousness of 'Gaia' - but these claims would have to be advanced or opposed on their own merits, without directly impacting the plausibility of combinationism.

In chapter 3 I suggested that physicalist, experience-first, and *a posteriori* combinationists all have ways to make CEI intelligible without any unconditional experiential inheritance. But subject-first property primitivist *a priori* combinationists - a significant constituency, particularly in the literature on panpsychism - must derive CEI from BEI, the unconditional inheritance of 'basic-experiential properties'. Thus they must hold that any composite containing human beings is at least as conscious as is implied by having basic-experiential properties, i.e. standing in the relation of basic-ownership to experiences.

A combinationist committed to BEI might think to free themselves of a commitment to mega-subjects by heavily restricting composition itself. They might claim that our everyday talk of things like the universe, or the biosphere, or the population of Canada, are merely devices for talking in a plural way about many things, without any commitment to there being, for each of these phrases, a single referent having a great many disparate things as parts. (Or perhaps our everyday talk does carry a commitment to such entities, but is wrong to do so.) Then all that exists are fundamental simples and the relatively few highly-organised composites they compose (such as human beings and other organisms, as Van Inwagen 1990 argues), and only these entities have experiential or basic-experiential properties.

The problem with this solution is that it cuts against the rationale for BEI itself, which was explained by the heritability of fundamental properties *given some level-connecting view of composition*. It is hard to reconcile heavily restricted composition with the level-connecting views: for a start, it directly contradicts nihilism, and rules out composition being identity – after all, *everything* is self-identical, so it's not clear where we can find the conceptual room to say, of some things, that they do not compose any whole. Heavily restricted composition is also probably incompatible with holism, if it rules out the existence of a cosmic whole. It might be compatible with priority pluralism, but even here the conjunction is a little strange. Wholes are nothing over and above their parts – but nevertheless there are very few. Most of the things we call 'wholes' do not exist – even though their parts do exist, and they would (if they existed) be nothing over and above those parts. Such a view may be consistent but I doubt it could be compelling, given the tension between privileging wholes on the one hand and deflating them on the other.¹

Subsection 1.2: Disavowed Experiences and Inconsistent Beliefs

While we might refuse to ascribe experiences to aggregative mega-subjects simply out of intuitive repugnance at the idea, there are a couple of more specific justifications we can give for this repugnance. Both are already identified in Lycan's (1979) discussion of a doctrine much stronger than BEI, namely the unconditional upward inheritance of *all* mental properties. Lycan argues, inspired by Block's 'Homunculus-Heads' (1992), that certain ways of formulating machine-state functionalism

¹ Some of these concerns could be raised about any form of restricted composition, but more radical restrictions pose more sharply the question of what explains why some collections of things compose a whole and others do not. What is it about these particular things that gives them this special status? The natural way to explain this is to think of the rare wholes as 'more than their parts' in a distinctive and special way, for instance by ascribing them strongly emergent causal powers (cf. Merricks 2001, 2005).

would entail mental inheritance; he takes this to be a reductio, revealing the need for functionalists to add further restrictions on what can be counted as a system's 'inputs', 'outputs', and 'internal states'.

The argument works thus: suppose that some part (s) of a system (S) is instantiating a program (P), i.e. there is a one-to-one mapping from some set of its real-world inputs, outputs, and states, to the logical inputs, outputs, and states used to define that program. Since there is no restriction on what we can count as the inputs, outputs, and states of a functional system, we can equally well regard the whole, S , as instantiating the same program (P), just by defining the goings-on in S 's part s as inputs, outputs, and internal states for S . For instance, if a cell in my brain pairs some input (electrical stimulation) to some output (firing) whenever it is in some state (polarised), then we can equally say that my whole brain, or the planet earth, or anything else containing that cell, pairs a corresponding input (electrical stimulation of that cell) with a corresponding output (firing by that cell) whenever it is in a corresponding internal state (having that cell polarised). So whenever s is realising P , S will be too, and thus if mental states are defined as programs, S will share whatever mental states its part s has.

Lycan takes mental Inheritance to be absurd for two reasons. First, the whole might "have thousands of explicitly contradictory beliefs" and second, it would be conscious of all its parts' experience, "despite [its] overwhelming inclination to deny it" (p.286). These are problematic insofar as it is plausible, first, that subjects do not generally have contradictory beliefs, and, second, that experiences are usually reportable by their subjects. Call these the *problem of Inconsistent Beliefs* and the *problem of Disavowed Experiences*.

I will not defend mental inheritance, in large part because of these two problems. Rather, I will try to show that BEI escapes these problems, and is thus more acceptable than mental inheritance, or than the unconditional inheritance of full-experiential properties.² To illustrate that the problems of

² Does the problem of inconsistent beliefs disappear simply because we shift from talking of mental states to talking of experiential states? Perhaps, if experiences and beliefs are entirely separate types of mental state, but many would deny

inconsistent states and disavowed experiences are *prima facie* pressing, consider some mega-subject with multiple human parts - such as the universe, a philosophy department, or “the contents of this room”, said to indicate a room with several people in it. Suppose one human part sees a talk, finds it interesting, and comes to admire the speaker; suppose their friend sees the same talk, but is bored and ends up disdainful of the speaker. Does it follow that the universe simultaneously feels excited and bored, or that it both likes and hates the talk? It seems absurd to attribute such wildly conflicting mental states to a single entity, even if we had reconciled ourselves to the universe having mental states at all.

Similarly, we cannot expect the universe (or even the philosophy department) to issue a joint statement avowing any of these conscious mental episodes, which makes it even less plausible to attribute these mental states to them. This is most obvious with social groups, which do sometimes avow things and so make it obvious that they do not avow all the experiences of their parts. However, we can see the same concern applied to non-social mega-subjects, in the form of an objection that has occasionally been raised against cosmopsychist theories, on which the universe has a mind, of which each of our minds is merely a fragment. Objectors suppose that were this the case, the universe would be aware of each of our experiences, and therefore able to avow or report those experiences. Since each of us is merely a part of it, the most natural way for it to avow our experiences would be through producing appropriate utterances through our mouths, and it should be able to report with one human an experience of another human, which would appear at our level as a sort of ‘telepathy’. Yet such telepathy does not occur, and hence the theory must be false (Cf. Avicenna 1952, 5.3.7-8, James 1909, pp.201-203, Goff 2012).

that they are: experiences often seem to have representational content, and conscious beliefs often seem to have phenomenal character. I will assume a connection of beliefs and experiences simply to show that even then, BEI remains viable.

Subsection 1.3: Basic-Having and Full-Having

The appropriate response to these problems is to focus on chapter 3's contrast between 'basic-having' and 'full-having'. This distinction was introduced because different views on the nature of experiential properties will make them either 'systemic', in that whether some entity instantiates them depends on its total set of properties, 'additive' in that their instantiation is not sensitive to their bearer's other properties in this way, or a mixture of additive and systemic components. No systemic property can be inherited, and so the very idea of experiential inheritance might be incoherent. By defining 'basic-experiential properties' as 'whatever is additive in experiential properties' we can discuss the doctrine of basic-experiential inheritance, which will at least be consistent whatever experiential properties are like. The original notion, including both additive and systemic components if any, is then labelled 'full-having' in contrast.

Combinationists who accept BEI should maintain that our intuitions about experience ascription, avowal, and consistency are not properly applied to basic-experiential properties. This might be because full-experiential properties involve systemic requirements that basic-experiential properties do not, and our intuitions are appropriate only to full-experiential properties. But even if basic-experiential properties just *are* full-experiential properties (because the latter have no systemic components), combinationist who accept BEI can point out that precisely because experiential properties require nothing systemic, they cannot automatically fit our intuitions about ascription, avowal, and consistency, because those seem to make over systemic demands.

So first, our expectation that experiences be generally consistent applies only to unified experiences, those which are 'experienced together'. If they are not, they may either not interact at all, or interact in ways unrelated to their content, so that the mental tension usually generated by a contradiction, which prompts revision, does not arise. We expect that a single subject will have largely

consistent experiences because we do not normally describe something as the subject of a set of experiences (do not take them to fully-have those experiences) unless they are all unified with one another. But this requirement of overall unity is a systemic requirement, and so does not apply to basic-having.

Similarly, the principle that people can usually report and reflect on their experiences is true only of experiences that they fully-have, not those they basically-have. Partly this is again a matter of unity, specifically ‘access-unity’, the joint accessibility of two experience’s contents. Partly it is also a matter of causal control: to fully-have an experience, a subject must display overall behaviour (including reporting behaviour) that is guided by that experience and its content. But causal control is again a systemic requirement, missing from basic-having.

What does basic-having involve, then, if it does not involve conscious unity or behavioural control? On physicalistic or experience-first accounts of experience it might involve very little, perhaps just the ontological qualification of having an experience occur ‘in’ you. For on those accounts, experiential ownership might be primarily a matter of causal relations among physical events, or phenomenal relations among experiences. However, BEI is primarily of interest to subject-first primitivists about consciousness, and for them the most natural account of experiential ownership involves a relation of subject to experience which is irreducible to any non-experiential relation, or to any purely inter-experience relation. This core would be part of basic-having, and BEI would imply that any composite containing a human being will bear this core relation to the experiences of its parts, and would thus be, in this inert and disunified sense, ‘conscious’.

So the universe, for instance, ‘has’ many millions of experiences, including mine and yours, but only in the basic sense. These experiences are grouped into clusters by their unity relations, and each cluster is in pretty good control of one small physical part of the universe. These parts are conscious

organisms: they have experiences in the full sense, while the universe does not. When one part sees a talk and enjoys it, the universe does indeed ‘have’ that enjoyment, but only in the basic sense. This is why it doesn’t avow it, and isn’t under any pressure to be consistent about it.

The choice of experience-first or subject-first views will determine exactly what BEI entails about mega-subjects. On the subject-first view, each of us is a physical thing that is conscious in the full sense, and we are parts of various physical wholes which are conscious in the basic sense in virtue of our being so (a monist will reverse the ‘in virtue of’). On the experience-first view, we are experiential things, based in our bodies, and parts of various larger experiential things, up to and including the ‘world soul’, which are based in the larger physical things of which our bodies are parts. These larger experiential things, unlike us, are only subjects of the experiences that constitute them in the basic sense.

If basic-having involves something irreducibly experiential, then there is in a sense ‘something it is like’ to be the universe. But if the universe’s experiences are not unified, there will not be any one single unified ‘thing it is like’: being the universe would be like simultaneously but separately being each of the individual minds in it (though the next section will complicate this claim). The key point is that the defender of BEI can accept a major difference between normal human subjects and mega-subjects: their commitment is simply to the idea that this difference is entirely accounted for by unity relations among experiences, and causal structure - even if experiential ownership is *not* fully accounted for by those factors.

Section 2: Could Mega-Subjects have Unified Consciousness?

In the last section I took for granted that the experiences of mega-subjects are not unified with one another, noting that this is compatible with their basic-ownership of all those experiences. However, we

saw in chapter 4 that there are many senses of ‘the unity of consciousness’, which may be dissociable from each other. This opens up the possibility that some forms of unity may in fact run much more widely than we tend to think: in this section I explore the motivations for, and implications of, ascribing unified consciousness to mega-subjects.

Subsection 2.1: Phenomenal Unity without Dispositional Unity

Phenomenal unity, the mere experienced-together-ness of a set of experiences, has been conceptually distinguished from representational unity (representing things as connected) as well as from access-unity and other forms of what I have called ‘dispositional unity’, unity that consists in the tendency towards certain sorts of causal interactions. Some philosophers accept the further claim that phenomenal unity is irreducible to and dissociable from these other forms of unity, though others deny this. Suppose for now that we accepted the irreducibility of phenomenal unity, and its dissociability from other forms of unity: there could then be a sort of ‘bare phenomenal unity’, in which experiences were phenomenally connected, but causally isolated and insensitive to each other. This sort of unity might then not be noticed by creatures like us, who run phenomenal and dispositional unity together because we are so familiar with cases that combine the two.

If bare phenomenal unity is coherent, perhaps it obtains among the experiences of distinct individuals, connecting up all the experiences of the mega-subjects they compose. We might, in the spirit of Russellian monism, take phenomenal unity to be the categorical basis of some basic physical relation, most likely spatiotemporal distance or causal interaction (cf. Chalmers forthcoming-a, pp.21-22). We might then consider the following, grotesquely counter-intuitive claim:

Unrestricted Unity (UU): All experiences in the universe are phenomenally unified with each other, i.e. form a single phenomenal field.³

Alternatively we might restrict this unity to certain mega-subjects: perhaps phenomenal unity still requires some degree of causal integration, as much as is found in social mega-subjects but not in merely aggregative ones. Note that UU is not a specifically combinationist doctrine: any account of conscious unity needs to say something about how phenomenal unity relates to richer forms of unity, and what sort of underpinnings it requires. If phenomenal unity is a fundamental relation which can exist without other forms, then that opens up the possibility that it extends much further than we tend to think.

Subsection 2.2: The Argument from Vagueness

Is there any positive reason to posit bare phenomenal unity among the experiences of distinct people? There is at least one reason, if we accept the metaphysical fundamentality of phenomenal unity, denying its reducibility to any more basic set of relations. Consider two sets of experiences, one clearly enjoying phenomenal unity (e.g. the experiences of some human being), one intuitively not (e.g. the experiences of the entire universe). These experiences will differ greatly in their tendencies toward a number of forms of interaction, in ways reflecting the underlying causal connections between them. But all these underlying causal connections - nerve fibres, radio signals, electrical or gravitational fields, etc. - are matters of degree. We can imagine an arbitrarily gradual sequence of steps between the two initial sets, different minutely in their tendencies to interact in each type of way (Cf. Parfit 1984, pp.231-243, Unger 1979, 1990, pp.191-206; note that the sequence may be multi-dimensional).

³ This is presented as a nomological but not conceptual necessity: one of the fundamental facts about our world is that it has a single phenomenal field, but this may not be conceptually or metaphysically necessary.

There need be no point where any fundamental break occurs, from the perspective of these causal tendencies.⁴ But experiences are either phenomenally unified or not, even though the change in the relevant underlying factors can be made so gradual that it consists of many miniscule changes, none of which seem a plausible candidate for ‘the big transition’. Structurally, this is a 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, and I do not wish to wade into their details. But broadly speaking, most solutions involve the idea that the word being applied has or could have had a wide variety of fractionally different meanings, ranges of reference, or degrees of application. 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’.⁵

Solutions in this broad category work well enough for words 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 for a fundamental concept which does not admit of further analysis, matters are much harder. Being indeterminately bald just means having a number of hairs we are unsure how to classify, but what is it to be indeterminately unified? What could it be like to experience two things, and have it be indeterminate whether they are experienced together?

⁴ This fact, that “Fundamental spatiotemporal and causal relations do not seem to... have [the] character [of stopping at the limits of the human brain]” (Chalmers forthcoming-a, p.23) is part of what motivates the ‘boundary problem’ discussed in chapter 4, section 1.

⁵ On epistemicist approaches, 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, forthcoming), and on other points. But the plurality of acceptable ranges of application is common ground.

The impulse to reject vagueness for what is fundamental provides the basis of one of the most popular arguments in support of universalism about composition (Lewis 1986 p.212, Sider 1997), which takes ‘existence’ to be a fundamental concept and infers that composition cannot be restricted in accordance with any vague criterion, for vagueness about whether some things compose a whole would entail vagueness in whether that whole existed. Since most of the plausible criteria by which composition might be restricted are vague, it seems to follow that composition cannot be restricted. This argument could be rejected by endorsing *metaphysical* vagueness, which may apply even to fundamental matters (see Van Inwagen 1990 pp.213-283, Barnes & Williams 2009, Wilson 2013-c), but there is dispute over the coherence of this idea (see Lewis 1993, 1988, Evans 1978, Eklund forthcoming).

Consequently, there are four things we could say about phenomenal unity, and different combinationists will find different options attractive:

1. Treat phenomenal unity as vague on account of a multiplicity of acceptable ranges of application, thereby accepting that it is reductively analysable in terms of some underlying dimension – most likely dispositional unity.
2. Treat phenomenal unity as fundamental, but subject to metaphysical vagueness.
3. Treat phenomenal unity as fundamental and hence non-vague, and conclude that there must be some apparently arbitrary moment when two sets of experiences suddenly become unified with each other.
4. Treat phenomenal unity as fundamental and hence non-vague, and conclude that if there is phenomenal unity at any point, there must be phenomenal unity at all points along any relevantly gradual spectrum.

In the fourth case, since “the flow of interaction in the universe is inherently unbounded, and no merely abstract pattern presents a natural condition for containing it”(Rosenberg 2004 p.88), we will probably

be led into universalism about phenomenal unity by an argument analogous to that which leads many into universalism about composition.⁶

Universalism about unity is very counter-intuitive, but it is important to note that we do not really have any compelling *empirical* evidence against it. UU is either conceptually incoherent (if phenomenal unity cannot come apart from more robust forms of unity), or it is invulnerable to empirical disconfirmation (because phenomenal unity, when distinguished from other forms of unity, has no distinctive empirical marks). If it is coherent, its evaluation will turn upon its theoretical virtues and vices, and here a case can be made both for it (that it is more parsimonious than alternatives, and that any precise boundary would have to be arbitrary) and against it (that it goes against common-sense, unverifiable, etc.).

Subsection 2.3: Does my Consciousness Adumbrate Yours? Boundedness and Pure Openness

The usual reason for rejecting the possibility of phenomenal unity in mega-subjects is the idea that our individual consciousness is manifestly ‘bounded’, in a way incompatible with our experiences being unified with any others. In section 3 of chapter 4, I elaborated this through the notion of ‘epistemic boundedness’: each of us knows about the experiences of others only in virtue of knowing about non-experiential things. But, as I observed in subsection 3.5, the epistemic boundedness of human consciousness might be disputed: it might be maintained that our experiences do, independently of what they tell us about non-experiential facts, convey to us some sort of information about other experiences. Investigating this possibility will require looking back at chapter 4’s account of conscious unity in terms of ‘adumbration’, and more precisely the ‘A-relation’.

⁶ See Goff 2012 for another variation on this same argument, supporting universalism about phenomenal composition, i.e. the view that all sets of objects in the universe are conscious subjects, a particularly radical form of panpsychism.

First consider aggregative mega-subjects experiencing bare phenomenal unity. In chapter 4 I argued that if bare phenomenal unity is possible, it is experienced by the parts as ‘minimal A-relatedness’, the relation between two experiences when each reliably adumbrates the other – i.e. represents the other as something not-given but continuous with what is given in that experience – and when this adumbration secures reference independently of any specific information regarding what is adumbrated. That is, each experience involves an unspecific awareness of the existence of a broader background encompassing its specific object, but with no idea what this background is, beyond its being ‘this further thing’: moreover, each subject’s awareness of ‘this further thing’ refers successfully to the other’s experience. If all human experiences are linked by the minimal A-relation, then they are not strictly epistemically bounded, though they may be bounded with respect to *specific* information: my experience informs me simply that there is a wider experiential world than is revealed to me individually.

So we should ask two questions: is it plausible that human phenomenology is pervasively characterised by pure openness? And is it plausible that each human’s experience of pure openness refers to the experiences of others? I think we should answer both questions with a cautious ‘yes’.

First, there is nothing implausible about pure openness in human experience. Indeed, arguably it is more phenomenologically accurate to describe our awareness as constantly ‘opening out’ onto an assumed, implicit, world of things not yet given, than it is to just list the specific objects of our experience, as though what we are primarily conscious of is a collection of discrete items, one thing after another. Of course it is specific items that tend to occupy our attention: the presence of the world as the universal background against which they appear is comparatively easy to ignore or forget about. And because this background sense of openness would be hard to notice if it did exist, it is hard to say with confidence whether or not we encounter it. But certainly it is not obviously absent.

For this feeling of openness to imply unity among all the experiences of a mega-subject, it would need to connect somehow to the experiences of other people, in particular by referring to them in virtue of features of them, but independently of any specific information it carries about them. Here again it is hard to have any confidence about whether such a connection exists, because it is hard to say conclusively what does or does not suffice for one thing to refer to another. But reference is commonly assumed to flow along causal or ontological links - if I am in a position to refer to something, I will also be in a position to refer to 'whatever caused it to be here', or 'whatever it is a part of'. In particular, I am usually able to refer to something in virtue of its causal or ontological relations to *me*. So all the defender of UU needs is some sort of causal or ontological relationship that extends widely through the universe, and this is not hard to find: the very fact of sharing a world might work, as might spatial distance, causal influence considered independently of strength or directness, or shared constitutive dependence on the same set of natural laws.⁷

If it is not implausible to claim that the many experiences of different subjects are ontologically connected just from sharing a world, and have a pervasive character of opening out onto a world prior to specifying its character or content, then the defender of UU can take these facts to explain, and to either constitute or be constituted by, the thorough-going phenomenal unification of all the world's experiences, and the corresponding phenomenal unity of the cosmic subject's consciousness.

Subsection 2.4: Social Mega-Subjects and Informative Adumbration

⁷ The relation of causal interaction may not be entirely pervasive, since objects at great distances can only interact over long periods of time, given the lightspeed limit imposed by relativistic physics - periods of time that might exceed the known lifespan of the universe. However, if the most basic form of phenomenal unity depends on some other relation, like sharing of laws, then it might be strictly pervasive.

Whether or not bare phenomenal unity pervades the universe, do any mega-subjects enjoy unified consciousness in a stronger sense, involving *informative* adumbration? The natural place to look would be at social mega-subjects, which often display complex and functionally useful information processing.

I have already suggested, in subsection 6.1 of chapter 4, that we routinely have an adumbrational awareness of other people's experiences: indeed, this may well be the *only* way that experiences we do not share can *perceptually* appear to us. But this sort of adumbration does not yield conscious unification. On the version of the adumbration proposal that denies transparency, what is necessary is for *my experience* to adumbrate someone else's, but in these cases the other's experiences are adumbrated by external physical objects like a word or gesture. On the version that accepts transparency, the problem is different: the adumbration makes me aware *of an experience* rather than of the content of that experience. Since experience is naturally transparent - that is, its only phenomenologically manifest features are its representational content - to be aware of an experience primarily *as* an experience, as an event 'in someone else's head', is actually an alienated and misleading way to be aware of it, when compared to 'seeing through it' to its content in the way that its subject does. So on either version of the adumbration proposal, this sort of routine inter-personal adumbration does not break through the epistemic boundedness of our consciousness.

In order for my experiences to be substantively unified with another person's, they would have to indicate to me what it is like to be that person, as a concealed aspect directly continuous with my own experience. I would have to get insight into how the world seems to them, in a way that came to me as automatically and immediately as my own experiences do, not requiring (though perhaps being amplified by) my focusing on any particular object. Perhaps the best model we have for this sort of connection would be with people whose viewpoints we have become so intimately familiar with, and

so constantly concerned with, that we cannot think of or perceive anything without it occurring to us how they would think of or perceive that thing. In cases like this - which typically involve parents, teachers, or mentors of some other kind - we might naturally say that we have the other person 'in our heads'.

But this kind of relation to another person's thoughts fails a different part of the test for being A-related: we are not aware of their experiences in virtue of any actual relation to their present experiences, but in virtue of a sort of 'internalisation', whereby we have constructed a working simulation of them in our own heads. It could, after all, persist even after they themselves have died. For our present experiences to be really unified with theirs, in a way that goes beyond bare phenomenal unity, we would have to have the same sort of automatic and immediate awareness of how things seem to them, but have it in virtue of a present flow of information.

Perhaps there are actual cases where this happens, with people who are both closely attuned to each other's ways of thinking and also directly involved in some co-operative activity (Cf. Bhattacharya & Petsche 2005, Nagel 1971, p.409). But perhaps even in those cases the degree of sensitivity is not high enough for us to say that their experiences are unified, or perhaps whether it is high enough is semantically indeterminate because our notion of 'conscious unity' is vague. Certainly, though, we can imagine hypothetical cases where, by providing more efficient technologies of communication, we enable two people's experiences to exchange information even more fully and sensitively: a thought experiment of this sort occupies much of chapter 8. If actual cases do not meet the standard of 'unified consciousness', they differ only in degree from cases that would.

Subsection 2.5: What is it like to be a Unified Mega-subject?

If we accepted Bayne & Chalmers's subsumptive analysis of phenomenal unity, or more broadly the idea that when two experiences are phenomenally unified there is a single thing it is like to have both together, then UU implies that there some single thing it is like to be the entire universe, some single experience subsuming all others.⁸ How should we think of this universal experience? What could it be like? Unfortunately I have little positive to say: I do not think we can do much to get our heads around such a dizzyingly remote idea.

In principle we could sympathetically imagine being the universe, simply by simultaneously imagining being all the individual organisms in it, at once, together. But obviously this is not possible in practice for humans, because we cannot combine so many experiences at once, because many of them will be deeply alien to us, and because we find it hard to imagine multiple experiences without connecting them in a way that would misrepresent their actual disunity.

Perhaps the last chapter's ideas could be illuminating here: does the universe experience everything in a confused way, as a single giant quality that is a blend of the qualities of all our experiences? Or does it experience them as laid out in an phenomenal field structured by attention? But the notion of confusion is defined in terms of various sorts of 'mental operations', and while we have a good intuitive idea of what it means to speak of a human mind performing mental operations on its contents, we have no idea what it means in the case of the universe. Similarly, the notion of attentional proximity breaks because we have no idea what would count as the universe as a whole attending to one particular experience. Admittedly, we can still think of a phenomenal field as structured by causal proximity more broadly, with two experiences that are prone to interact closely and transmit a lot of information being 'closer' than more isolated ones.

⁸ Cf. Bayne & Chalmers 2003, p.40, "In nature, it may be that the most basic sort of conscious state is the total phenomenal state, or the phenomenal field, or even the phenomenal world."

It might then make sense to say that the phenomenal fields of a mega-subject's human parts would be distinctively-structured regions of the mega-subject's field. If one accepted the last chapter's idea that the distant periphery of a human phenomenal field contains experiences that are phenomenally unified but realistically unattainable, that might suggest that these far peripheries of each of our phenomenal fields would shade gradually into parts of the far peripheries of other people's. But it is unclear how much of our sense of experienced as 'organised' and 'laid out' remains meaningfully applicable, once we abandon the possibility of directing or shifting attention.

Both phenomenal blending and the phenomenal field are notions drawn from our own, distinctively human, sort of phenomenological structure, and are imperfect guides to what a very different sort of consciousness might be like. We probably just cannot imagine what it is like to be the universe, though we may draw some consolation from the thought that the universe itself probably cannot imagine that either, since it probably (supposing pantheism to be false) lacks the cognitive architecture necessary to do anything that would count as imagination. Similarly, in any strong and reflective sense of 'knowledge', it does not know what it is like to be itself, though it basically-has many conscious thoughts about what it is like to be particular parts of it - namely those which we fully-have.

Section 3: Human Bodies as Organic Mega-Subjects

We usually assume that the only experiences going on in a human being are those 'in the brain': I have tended to follow this assumption in treating brains, heads, and whole human bodies as experientially equivalent subjects, sharing the same set of experiences (or, on the experience-first view, as each being the basis for the very same subject and its experiences). But what if there were experiences occurring independently in other parts of the body? This might be a result of some strange hypothetical procedure

that bestowed consciousness or some body part, or it might be one way of thinking about the shared body of some pairs of conjoined twins, or it might simply be a consequence of the truth of panpsychism, according to all human tissues, like all physical things, contain experiences. The human body itself would then be something like an organic mega-subject, with two or more separate sets of experiences going on in it. What should combinationists who accept BEI say about this possibility?

Subsection 3.1: The Panpsychic Body

What should panpsychists say about the experiences of my non-neural body parts? Note that exclusionary combinationists will likely deny any sort of mega-subject here: even if there is such a thing as the whole body, or the aggregate or all experiences going on in it, it will not be a subject of the non-brain experiences in any interesting sense at all. Without the kind of complex cognitive capacities that are distinctive of the brain, there is nothing remotely like phenomenology. But this simply reflects that exclusionary combinationists deny BEI.

What about inclusionary combinationists? What they say will depend somewhat on whether they take a subject-first or an experience-first view. Subject-first combinationists will say that there is one entity, the whole human body, which simultaneously has both the experiences based in the brain and the experiences based elsewhere, though because not all of these experiences are causally integrated or contribute to intelligent control of the body, it has the non-neural experiences only in the basic sense. Moreover, if any of these experiences are not unified with one another, and if full-ownership requires unity among all of one's experiences, then the body does not have *any* experiences in the full sense.⁹ The brain, however, or some part thereof, does fully have experiences,

⁹ I here suppose that what is required for full-having is unity amongst all the experiences which the subject basically-has: a different way to interpret the unity requirement would be as requiring unity among the experiences

because all of the experiences in the brain are unified with one another.¹⁰ (Of course, in addition to the brain and body there will be any number of intermediate groupings, like the body's top half: I focus on the brain and whole body for simplicity.)

Experience-first combinationists, on the other hand, will focus on two distinct but purely experiential beings: one more extensive, but not sufficiently organised to be said to constitute a subject in the full sense, the other containing fewer experiences but constituting a subject in the full sense. The latter being would be based in your brain, and by extension in the whole body too, though neither brain nor body should strictly be called a conscious subject. The former, more extensive, experiential being is based specifically in the whole body, as no proper part of that body is ontologically sufficient for all of the experiences it includes at any one moment.

If we do have two subject-like beings, one more extensive (identical to or based in my body) and one more integrated (identical to or based in my brain), it is an open question which is a better candidate for being identified with 'me the person'. Just as non-panpsychists must decide how heavily to weight the brain's special role in consciousness, panpsychists must decide how heavily to weight the brain's special role in unified, causally integrated consciousness.

It might be thought obvious that only the the less extensive, more integrated, entity could be 'me', because if I were the other, I would have many thousands of additional experiences which, to all introspective appearances, I do not seem to have. But this is a cogent argument only if we assume that any experiences I have must be available for cognitive use (reflection, inference, report, and so on) and inclusionary combinationists have already given up that principle. If I am my whole panpsychic body,

which the subject would (given other facts) fully-have, and in that case the body might qualify as fully-having the experiences based in the brain.

¹⁰ Is it certain that all the experiences in the brain are unified with each other? If the last chapter's account of 'phenomenal refraction' is accepted, perhaps not. In that case the full-subject will be something other than the brain itself, perhaps a part or functional aspect thereof (e.g. loops of cortico-thalamic oscillation, whatever their ontology).

then the experiences of my arms, legs, hair and so on will be ‘realistically unattendable’ in the sense explained in chapter 5, subsection 3.4: so distant from the centre of my phenomenal field that the ‘shift of attention’ needed to focus on them would involve radical physiological re-organisation. Ironically, we would then be in something like the position of the universe: unable to reflectively know, or even imagine, the whole of what we are currently experiencing, but only a small, central, part of it.

Subsection 3.2: What if my Foot Became Conscious?

To see that this issue is relevant not only to panpsychists, suppose that by some series of unfortunate events, likely involving the intervention of technologically advanced aliens, my left foot becomes conscious. Perhaps it just feels discomfort when flexed, and comfort when extended, with greater intensities proportional to increases in the degree of flexion or extension. My foot is part of me, and it is now experiencing these sensations: BEI would entail that my body, considered as a whole, had these experiences in the basic sense. At the same time, this is not the usual, full, sense of experiencing: my usual methods of introspection will not reveal the foot-experiences, I will deny any knowledge of it, and if ‘I’ enjoy flexing my foot, while the foot itself feels discomfort, then it seems wrong to say that any single being is experiencing a conflict among their feelings.

One question is which entity to regard as ‘me’, as ‘Luke Roelofs’. A subject-first combinationist might say that I am the whole body and now begin to basically-have new experiences which are not unified with my existing ones, and which do not control me – i.e. experiences which I have only in the basic sense, not the full sense. The tricky thing about this is that if *I* am having this new foot-based experience, which is not unified with my other experiences, then none of my experiences are unified with *all* my other experiences. But in that case I cease to have even my brain-based experiences *in the full sense*. We can, of course, say that *my brain* has those experiences in

the full sense, and if the experiences in my foot become more complex and start to govern goal-directed behaviours on the part of my foot (e.g. tapping out messages in morse code with a toe), we might say that my foot has those experiences in the full sense. On this way of describing the case, one entity (my body) first has some experiences in the full sense, and then comes to only have experiences (including those directly flowing on from its old experiences) in the basic sense. Another being (my brain) experiences more continuity, being a full subject of similar experiences at all stages. If we wanted to capture the continuity that 'I' seem to experience, we might think that the brain is a better candidate for being 'me' than is the whole organism.

Experience-first combinationists would say that there is, at first, a single complex, integrated, stream of experiences based in my brain and, by extension, in my body, which constituted (perhaps was identical to) a subject (me). Over the course of this thought experiment all that happens is that another stream of experiences, constituting another subject, comes into existence, based in the same thing that constitutes me (though depending specifically on a different part thereof). The sum of these two sets of experiences (a larger set of experiences) could be said to constitute a composite subject, but one which experiences its constituents only in the basic sense.

Subsection 3.3: Why some Experiences are not Mine

Suppose panpsychist combinationists are asked "If every part of my body is having experiences, why am I not having those experiences? Why are they not part of my consciousness?" (Cf. Rosenberg 1998, Chalmers 2013a, p.5; Margaret Wilson 1999, pp.128-130, raises this problem especially for Spinoza's version of panpsychism). The question may seem especially forceful when we observe that any precise boundary between the brain and the rest of the body will seem arbitrary when we look closely - why draw the line at the brainstem, and not the midbrain, or the spinal column? But in fact this question is

misconceived. It is like asking ‘why is my neck not included in my head - why does my head exclude the neck?’ We can carve reality in many ways, but once we have singled out one section, there is no explanation for why it does not include other sections except that we did not choose to cover them with the terms we were using.

That is not to say there are no substantive questions here. One question is: why are the experiences of my non-neural body parts not reportable by me? But the answer is straightforward: they do not broadcast information to other subjects (or to each other) in the way that would allow for that. A more difficult question concerns self-reference. When we ask ‘why do certain experiences not belong *to me?*’, the meaning of the question depends on the meaning of those last two words, and combinationism raises a lot of difficulties in relation to that already puzzling topic. In the next chapter I will confront these difficulties.

Summary:

In this chapter I have explored what combinationism can tell us about conscious beings with humans as parts: this could include social groups, ecological systems, mere aggregations, or the cosmos itself. Certain versions of combinationism are committed to BEI (basic-experience inheritance), and BEI implies the existence of a certain sort of mega-subject, one which ‘has’ experiences only in the basic sense. More precisely, BEI implies that if some collection containing human beings composes a whole, that whole will be such a mega-subject. But the distinction between basic and full ‘having’ also provides the tools for explaining away the counter-intuitiveness of this implication.

I also examined the issue of conscious unity in mega-subjects. According to the adumbration proposal of chapter 4, the experiences of different human beings are generally not dispositionally or representationally unified, because these experiences do not informatively adumbrate each other.

However, the adumbrational proposal does not tell us how much dispositional unity is necessary for phenomenal unity, and if the answer is ‘very little’ or ‘none’, it does not tell us how broadly phenomenal unity extends. These questions are not specific to combinationism: the adumbrational proposal simply tells us what these questions correspond to for the component subjects involved.

In particular, if phenomenal unity is a fundamental relation which can exist without richer forms of unity, then there may be a case, based on the theoretical virtues of simplicity and non-arbitrariness, for ‘Universalism about Unity’, the doctrine that all experiences anywhere are phenomenally unified. The adumbrational proposal does not in itself affirm or deny this, but it tells us what this would mean for human consciousness: that human consciousness is suffused with a basic awareness of its own continuity with an indeterminate ‘something more’, and that this awareness successfully refers to all other experiences that exist. For richer forms of unity we would need richer forms of interaction, and current forms of social communication, though they probably fall short of conscious unity in the everyday sense, differ from it only in degree. The sort of progressive deepening and tightening of inter-personal links that might underwrite conscious unity will be discussed in more detail in chapter 8.

Finally, the analysis of mega-subjects bears on how panpsychists should think about the human body and its myriad experiences both inside and outside the brain. A panpsychist who accepts BEI will recognise the existence of both a brain-based subject, which fully has all and only the brain-based experiences, and a body-based subject that basically-has both the non-brain-based experiences and the brain-based ones. ‘The person’ as we normally think of them might be identified with one or the other: if we chose the latter, we would arrive at the surprising conclusion that each of us is in fact a sort of mega-subject ourselves, an experiential universe extending well beyond the central focus of the brain. This, like many of the ideas discussed in this chapter, is wildly counter-intuitive, but it is not something

that all combinationists must accept: I have simply tried to show where some of the more radical versions of combinationism might lead, and how we might make sense of their surprising consequences.

Chapter 7: Knowledge of Self, Other, and Part

<u>Section 1: Why Don't Our Parts Notice Each Other?</u>	<i>p.259</i>
<i>Subsection 1.1: The Patterning Principle</i>	<i>p.261</i>
<i>Subsection 1.2: The Patterning Principle and External Events</i>	<i>p.262</i>
<i>Subsection 1.3: The Patterning Principle and Internal Events</i>	<i>p.264</i>
<i>Subsection 1.4: The Patterning Principle and our Oblivious Parts</i>	<i>p.268</i>
<u>Section 2: What Does 'I' Mean?</u>	<i>p.271</i>
<i>Subsection 2.1: Secure Self-Reference and Epistemic Optimism</i>	<i>p.271</i>
<i>Subsection 2.2: An Exception for Introspective Self-Ascriptions?</i>	<i>p.273</i>
<i>Subsection 2.3: Relativised Reference</i>	<i>p.276</i>
<i>Subsection 2.4: Privileged Reference</i>	<i>p.277</i>
<i>Subsection 2.5: Indeterminate Reference</i>	<i>p.278</i>
<i>Subsection 2.6: Dual Reference</i>	<i>p.280</i>
<u>Section 3: Who am I?</u>	<i>p.284</i>
<i>Subsection 3.1: Can We Self-Identify Relative to Our Experientially Equivalent Parts?</i>	<i>p.285</i>
<i>Subsection 3.2: Can We Self-Identify Relative to Our Experientially Different Parts?</i>	<i>p.287</i>
<i>Subsection 3.3: The Importance of Self-Identification</i>	<i>p.290</i>
<i>Subsection 3.4: Self-Identification as Action-Guiding</i>	<i>p.293</i>
<i>Summary</i>	<i>p.296</i>

Our thought about, reference to, and knowledge of, ourselves is complicated if some parts of us are also thereby thinking about, referring to, and knowing, themselves. In this chapter I examine these complications, arguing that they push combinationists towards a set of claims which systematically lower the importance of knowing exactly which subject we are.¹

Not all conscious parts of a self-conscious being need be self-conscious: microsubjects of the sort discussed in chapter 5 are probably incapable of thinking about subjects of experience as such, or thinking of themselves as subjects of experience, just as a cat or a worm, even if conscious, is probably incapable of such thought.² But many parts of a human being are large and complex enough to be at least *prima facie* capable of self-conscious thought: for the subject-first combinationist this will include

¹ I am here assuming that conscious occurrent thoughts, judgements, and beliefs are at least connected with, if not identical with, conscious experiences: that is, I suppose that experiences can have content and beliefs can have phenomenology, even if they have them derivatively or indirectly.

² Perhaps "consciousness essentially involves self-consciousness" (Kriegel 2004, p.182, cf. Sartre 1956, p.xi). Then even the most basic sort of experiential state, possessed by a simple animal or by a microsubject, contains some awareness of its own subject. In that case the issues discussed in this chapter have an even wider scope than I suggest, though finding examples will be harder. I think all the claims I defend retain their validity even with this increased scope.

parts which share the human being's whole consciousness (e.g. their head), as well as such awkward entities as the brainstem-and-left-hemisphere, or the-brain-minus-one-lobe. And even for the experience-first combinationist there will be many experiential sections that comprise many or most, but not all, of the whole's experiences. In chapter 1, subsection 4.2 I called these 'large overlapping sections' of human subjects: when I speak of 'parts' in this chapter I should be assumed to have parts like this in mind.

A combinationist might sidestep all the problems in this chapter by embracing a highly restricted view of composition, according to which all that exists is the whole and parts of it small enough to be conceptually incapable of self-consciousness (e.g. Van Inwagen's view on which only organisms, cells, and fundamental particles exist). But (for reasons discussed in the introduction to the last chapter) this kind of restricted view sits very uneasily with combinationism, since it suggests that certain wholes are something over and above their parts, the opposite of the combinationist project. Moreover, even this attempt at evasion will fail if all conscious beings, even microsubjects, are self-conscious. So I will assume in this chapter that the combinationist's principle of composition is at least liberal enough to generate the problems I discuss.

Another sort of 'restrictive' solution would say that even though large sophisticated parts of me exist, and could think 'I-thoughts' were the rest of me removed, they do not think such thoughts when connected to the other parts of me - only the whole is capable of this kind of thinking, whatever the intrinsic capabilities of its parts. This implies that being connected to each other prevents my parts from thinking I-thoughts: but what sort of prevention is this? If it is simply that my parts' thoughts cannot *count as* I-thoughts, despite being intrinsically just like I-thoughts, then we have 'solved' the problems I will consider only by verbal fiat. But it is not plausible that being connected to other parts could altogether *annihilate* a part's self-referential capacities, so that nothing even remotely like an I-thought

could arise. The only satisfying and plausible sense of ‘prevents’ would thus have to be something more like ‘interfering with’ or ‘disrupting’: being connected to other parts makes each part’s thoughts diverge from I-thoughts in some important way, despite being still significantly akin to them. But then we can demand an explanation of how the ‘almost-I-thoughts’ of the parts diverge from genuine I-thoughts, and the answer will come through engaging with the sorts of considerations discussed in the following sections.

I distinguish three particular problems, which I call the Problem of Oblivious Parts, the Problem of Self-Reference, and the Problem of Self-Identification. In sections 1, 2, and 3, I discuss these problems in turn, arguing that the first two can be removed, but the third remains: combinationism limits my ability to know which of a number of subjects is me.

The first problem is simply this: it seems that the parts of me do not recognise that they are surrounded by *other* parts – by distinct and even discrete subjects with whom they interact. At least, I have no inkling of such a recognition, which suggests there is none.³ Yet in general it seems that intelligent subjects can *notice* when they are in constant close interaction and communication with discrete subjects, i.e. they notice when there is ‘someone else’ there. I could not be cohabiting and co-operating closely with other people and remain ignorant of this, but the large conscious parts of us are, according to combinationism, living and acting with other subjects, apparently without any awareness of this. Indeed, the intuitive attractiveness of Anti-Combination reflects that there seems to be no recognition within the group of communicating subjects that constitutes me that there is such a group. Call this the Problem of Oblivious Parts.

The second problem concerns the reference of first-person terms in conscious thoughts that are shared by whole and part. Usually we assume that such terms refer to ‘the subject’ of that thought, but

³ This problem was suggested to me in conversation with Benj Hellie.

if thoughts can belong to multiple subjects, which of them is the referent? This question becomes particularly pressing when some of those subjects differ from others in their possession of the feature that the thought self-ascribes: for then the thought's truth or falsity will depend on which subject it refers to. Once we abandon the assumption of a unique subject for each judgement, first-person judgements give rise to a conflict between accurate self-knowledge, and reliable self-reference. Call this the Problem of Self-Reference.

The third problem is the one which has been most often employed to argue against combinationism (Merricks 2001, pp.103, Unger 1980, pp.461-462; cf. Olson 2003, pp.329-331; cf Appendix subsections 6.4 and 6.5): who am I? Am I a human being, or a human head which thinks it is a human being? If they think all the same thoughts, it is hard to see how any of them could tell. Call this the Problem of Self-Identification.

Note that satisfactory responses to the first two problems might make the third harder. To address the Problem of Oblivious Parts we might say why our parts are in an epistemic situation which they cannot distinguish from that of the whole – but this will then make it even harder to explain how any of these subjects could identify themselves. And to address the Problem of Self-reference, we might say why true self-ascriptions made by the whole need not involve the parts in error – but this too will make it harder for either to discern itself from the other.

Section 1: Why Don't Our Parts Notice Each Other?

There are two basic ways to respond to the Problem of Oblivious Parts. One would be to claim that the parts are not really oblivious – at least some of our large, cognitively sophisticated parts are aware that

they are not alone, and hold some attitude toward the other parts which they are attached to. The other is to accept that they are oblivious, and explain why.

I think combinationists have good reason to go for the second approach: any awareness of accompaniment in my parts should be inherited by me. Since the relevant awareness would be too cognitively sophisticated for it to be plausible that it was isolated from the whole's other experiences, or uninvolved in the control of overall behaviour, it would meet the conditions in any plausible version of conditional experience inheritance. It would be bizarre for our brains to develop the machinery for social cognition and then cut it off from the rest of the brain's function, and doubly bizarre for such a structure to be present in every single one of our large, cognitively sophisticated, parts.

So I will suppose combinationists take the second approach, of explaining why our parts might fail to recognise each other, when connected in the ways that they are connected. Note that not only do our parts not recognise each other as subjects, they do not recognise each other as external things at all. Yet it is not plausible that they do not perceive each other at all, in any sense: they are in constant, vital, sensitive contact. Hence they must perceive each other *as extensions of themselves*. More precisely, they interpret the signals they receive from each other as telling them either of *their own voluntary actions*, or of *events going on 'in them'*. In this regard they are in something like the opposite position to a schizophrenic suffering from 'thought insertion', who perceives their own internally generated thoughts and experiences as produced by outside forces. Where the schizophrenic perceives what is actually 'internal' as being 'external', the conscious parts of a normal human perceive events that are actually 'external' as being 'internal'.⁴

⁴ Throughout, terms like 'internal', 'external', etc., are meant ontologically rather than spatially; 'in me' means 'grounded in some part of me', while 'outside me' means 'wholly grounded in things discrete from me'.

Subsection 1.1: The Patterning Principle

Since some of our parts employ the very mechanisms that we do, whatever it is that lets them categorise things as internal and external, as ‘self’ and ‘other’, must also be what lets us do that. If combinationists claim that our parts miscategorise each other as ‘self’, they must think that we could make a similar miscategorisation in the right circumstances. That is, they are committed to:

Parity Claim (PC): If we were connected to another subject in the way that our parts are connected to each other, we would not regard that subject as a distinct entity, but instead would ascribe their decisions, thoughts, and experiences to ourselves.

The Parity Claim is not uncontroversial or trifling: it strikes at the heart of a certain conception of self-knowledge, implying that we are not infallible about whether certain experiences and thoughts and feelings are ours. So there is a real burden on combinationists to show that PC might nevertheless be true and intelligible. To discharge this burden, I will defend a certain schematic account of how we make judgements of self and other, which I will call the ‘Patterning Principle’:

Patterning Principle (PP): Our spontaneous impressions⁵ of whether an event is external or internal, and of whether it is our voluntary action or not, are determined by the patterns of correspondence and divergence we detect between it and other things.

Obviously the Patterning Principle is still incomplete insofar as it leaves open what exactly the relevant patterns are, and how they are detected. But even in this schematic form, it allows for the Parity Claim, since both internal and external events can stand in the same patterns, and so we might self-ascribe actions or events which in fact occur externally and independently of us, if we detect the right pattern in them. This allows combinationists to say that because of the pattern of interactions among the parts of each human, each ascribes to itself all the events and processes which it perceives going on in the

⁵ ‘Spontaneous impressions’ means those by which we constantly and automatically construe situations as containing a certain number of subjects, one of them ourselves, who are responsible for various perceived events. This contrasts with judgements based on theoretical beliefs, testimony, etc.

others. Each thus feels itself to be alone and responsible for all the mental activity in the whole human, which then inherits this unanimous judgement of solitude.

I think that most of us are inclined to accept some role for patterns of correspondence and divergence in determining our impressions of internality, externality and causal responsibility, but only a limited one. At some point, we tend to think, we fall back on a direct metaphysical insight into certain events being ours, either in the sense of being our actions or of being ‘in’ our own minds. Thus I think we normally tend to endorse a ‘hybrid’ view, with some role for detecting patterns and some role for direct insight. It is this direct insight that poses problems for combinationists: shouldn’t each of our parts know automatically that some of the thought-processes that guide this human body are *theirs*, while others are not? So the strategy I will pursue on behalf of combinationism is to argue that patterning considerations can in principle entirely explain the relevant class of judgements, making direct metaphysical insight superfluous.

Subsection 1.2: The Patterning Principle and External Events

The most plausible role for patterning considerations is to determine our impressions of the causal relations among external events. What makes it seem to us that one event we perceive is caused by another? Surely the answer has to be ultimately in terms of some sort of covariation, things either changing at the same time (or in quick succession) or remaining constant together while other things change. Whether we express this in the language of Bayes’ Theorem or in that of Humean laws of association, the basic idea seems clear enough: we respond to regularity among the changes we perceive in the world. To put a name to this general idea, I shall speak of observing ‘harmony’ between two events when they are correlated with each other in the relevant ways: the two events are then

‘harmonising’. This notion is meant to be a placeholder for a more detailed account of the statistical relations that we respond to.⁶

But, given some idea of which external events are causing which, how do we identify some of these events as *our own* actions? There has been a lot of empirical work on this question, but for the most part it is accepted that we rely on considerations of patterning: the dispute, for instance, between the ‘comparator model’ (see Helmholtz 1866, Blakemore et al., 2002, Frith 2012, Carruthers 2012) and the ‘multifactorial model’ (see Synofzik et al. 2008, Moore & Haggard 2008, and David et al. 2008) is a dispute over the particular weighting and mix of factors used, over whether there is a single privileged comparison or not. What is agreed on is that some brain system has to compute, based on signals from perception and from the internal processes that produce action, which events are ‘done by me’. This idea is borne out by the possibility of ‘tricking’ subjects into self-ascribing responsibility for externally caused events by manipulating their perceptions (Nielsen 1963, Ramachandran & Rogers-Ramachandran 1996, Lynn et al. 2010, Ebert & Wegner 2010, cf. Wegner 2002), or by direct brain stimulation (Desmurget et al. 2009, cf. Fried et al. 1991).

So we might say: we identify external events as our own voluntary actions when and only when we perceive them as harmonising with our internal decisions and volitions – that is, if we see our arm move just after we’ve consciously decided to move our arm, the harmony between these two events is what gives us our strong impression that the arm rose *because* we raised it. However, this presupposes that we already self-ascribe the conscious decision, which just pushes the question back a step. It would be an obviously vicious regress to say that we regard it as our own decision because it harmonises with a prior decision to make that decision. Hence there is a strong temptation to think that this decision is

⁶ There may or may not be additional, innate or learnt, ‘models’ we are particularly prone to recognise and perceive as causal. For instance, we might automatically regard the transmission of rectilinear motion as ‘how things work’, and be more sensitive to such patterns than to, say, S-shaped motion (Cf. Cheng 1997). But these models cannot be the whole story, since we can identify and recognise causal relations even among unfamiliar or bizarre items.

self-ascribed in some special way that is infallible and does not depend on patterning; this is the notion that combinationists must try to resist.

Subsection 1.3: The Patterning Principle and Internal Events

It is regarding internal events that the Patterning Principle is contentious. Don't we categorise mental events simply by modality: what we learn of through sight and hearing and so on is external, while internal events are those we learn of by 'introspection', or 'acquaintance', or some such non-perceptual awareness. But given the Adumbration Proposal of chapter 4, this does not settle the matter: our basic, non-perceptual awareness of experiences involves both the revealed aspects - our own experiences, fully given to us - and the concealed aspects - the experiences of other subjects - of a whole phenomenal field. Whenever our experiences are unified with those of others, we will learn of both sets of experiences via 'introspection' or 'acquaintance', and some further mechanism will be needed to discriminate the two. Could that mechanism involve 'harmony'?

Certainly some internal events may fit neatly into the very same model as external events, e.g. imagination: some experiments suggest that faint sensory stimuli may be miscategorised as imaginary when they match what subjects were independently attempting to imagine, i.e. when they harmonise with a prior intention (Perky 1910, cf. Segal 1972). But many of the events we experience as being 'in our minds' are not preceded by any distinct decision to produce them – our decisions, thoughts, impulses, and so on seem to be 'ours' *on their own*, with no need for us to compare them with any other events. How can the notion of harmony even apply here? I think there are actually a few available ways for PP to cover internal events, and in the remainder of this subsection I will briefly review three.

First, we might self-ascribe events not just when they harmonise with particular other events, but also when they harmonise with our background psychology, reflecting the fact that our mental events are typically “caused by a combination of our background beliefs, desires, and interests” (Campbell 1999, p.617). We might focus in particular on the mass of dispositional intentions, desires, and goals that could be called the ‘the background will’ (as contrasted with ‘the occurrent will’, our present conscious feelings of desire and intention). The background will is the underlying structure of what we want or intend to do in different situations, given different contingencies, when confronted with different stimuli. It is probably impossible to ever fully articulate this structure – we cannot write out a list of what we would want or do in every possible situation – but there does seem to be a significant determinacy in it. Even a libertarian must recognise that there is a complex and subtle set of persistent facts about us in virtue of which we are inclined to will some things, and in virtue of which we perceive things as unwanted or desirable, frustrating or welcome.

Note that we are not considering the obviously regressive idea that a subject judges some thoughts to be theirs because of its fitting well with the background psychology that they have already *judged* to be theirs: the background psychology need not be self-ascribed, indeed need not be the object of any kind of thought or awareness, in order to govern and guide what self-ascriptions are in fact made. This evaluation of harmony could be, and probably is, largely unconscious and inaccessible to rational reflection: often the only way we can find out what it is we really want is to expose ourselves to actual or hypothetical cases and see how we react.

Nevertheless there is a significant problem with this proposal: we can regard a thought as ‘ours’ despite its content being wildly out of character, just as we can think that someone *else* is thinking exactly what we would think in the circumstances (cf. Gerrans 2010, pp.233 ff). Being ‘in character’

for us seems neither necessary nor sufficient for being perceived as ‘ours’. Thus the defender of PP should supplement the appeal to background psychology with other factors.

A second option is to appeal a sort of *stability*: the thoughts and feelings which we ascribe to ourselves flow into one another, affecting and being affected by other things we self-ascribe, and by our own will (background or occurrent), while external things remain fixed as our attention, plans, or ideas flow over them. For example, contrast a perception of a red square with an idea of a red square that occurs to me unbidden: the latter will shift, recede, disappear or transform according as I attend to it, ignore it, suppress it, connect it with something else, consider something related or something unrelated, and so on. And those changes which thus affect it are themselves similarly mutable: while we do not choose each step, we can intensify or inhibit things voluntarily if we try to. By contrast, the perceived square either stays constant during all those fluctuations, or else changes in a manner uncorrelated with them. We might say that while neither is positively voluntary, the perception *resists* my will, whereas the idea does not. This is just a version of the long-standing idea that we identify things as external by their resistance to, or at least independence of, our will (e.g. Descartes 1985, V2 pp.26-27, 55, Locke 1836, p.484, Berkeley 2008, p.41).

Again we might worry about vicious regresses: doesn’t perceiving things as responsive to ‘our will’ require having first self-ascribed our will? But as with the first suggestion, this misunderstands the proposal. PP says that our spontaneous impressions of some X being ‘in us’ or ‘outside us’ are based on detecting harmony between it and some Ys, but does not require that the Ys be *themselves* the objects of any judgement or impression. We might not even be conscious of the Ys, yet still have our conscious awareness of X affected by its harmony with them.

But aren’t there inner mental states which are stubbornly resistant to our efforts to ignore or induce them, and which display stability over time? Here we must distinguish two cases. Things like

the pain of a toothache or illness, which are pretty entirely indifferent to our will, are best viewed as a special sort of external perception, where the perceived object is inside the body and invisible to other senses (though sometimes visible to medical examination). This stable pain is a perception of ‘something bad happening in my tooth’, just as a sudden flash of pain might be a perception of the tooth being struck or broken.

On the other hand, obsessive or haunting ideas and feelings which refuse to go away are not, it seems to me, genuinely indifferent to our will. Rather, they recur and return constantly, in spite of being partially inhibited, or at least modulated, by our efforts to dispel them. Consequently they are hard to pin down or analyse exactly – their aspect is affected by our efforts to attend to and analyse them, just like ordinary thoughts and unlike the pain of a toothache. So it seems to me that neither case of persistent involuntary pain – physical or emotional – conflicts with the application of the patterning principle given here.

A final idea is that the ‘efferent copy’ mechanism employed in the monitoring of motor actions could also be present with thoughts, feelings, and other mental events. Feinberg (1978) and Campbell (1999) in particular suggest this, advocating regarding thought ‘as a motor process’. The idea is that the brain processes that produce actions or thoughts do not just produce those actions or thoughts, but also produce ‘efferent copies’, signals reporting that such-and-such an action or thought has been produced. These copies are processed by a ‘comparator’, some brain system that also takes in feedback from the actual execution of these actions and thoughts, and from other internal and external events, and monitors the correspondence or lack of correspondence among them. Since the comparator could operate below the level of conscious awareness (cf. Campbell 1999, pp.617-618), it would easily account for cases where something seems internal to us despite not cohering with other consciously accessible elements of our minds. And it seems well-placed to account for the disorders of thought and

of self-consciousness found in schizophrenic patients. But it might seem extravagant for every single conscious event to be part of such a monitoring process, with every whim, twinge, flicker of doubt or snatch of memory generating a copy to be submitted to a comparator mechanism. If so, we might posit a comparator-type monitoring mechanism for some but not all mental events, relying on the first two proposals to explain our impressions of the others.

I conclude that the sort of ‘harmony’ which leads us to regard events as occurring ‘in our own mind’ can be something less than conformity to a definite and coherent prior volition: it can be any combination of conformity to background psychology, sensitivity to a wilful stream, or the verdict of a dedicated comparator mechanism.

Subsection 1.4: The Patterning Principle and our Oblivious Parts

If we adopt PP, along with the above suggestions about how to understand ‘harmony’, we must conclude that whether some thought or action seems to be ‘mine’ depends not on whether it really is, but on its correlation with various other events and dispositional properties. If two subjects interact harmoniously, so that what each does harmonises with the psychology of the other, then they will each self-ascribe everything the other does that they are aware of; if they are set up so as to interact in this way all the time (call this ‘harmonious connection’) then they will go their whole lives without ever realising they are not alone.

I think it is plausible that the parts of a human being are harmoniously connected. The electrical behaviour of a brain part, and the physiological structure that underlies that behaviour, is very sensitive to that of surrounding parts, and has developed in constant interaction with them for years. Moreover, they are alike in their biological requirements (pH, temperature, salinity, etc.), their ways of coding

information, the timescale on which they act, and so on. It is hard to imagine better conditions for harmonious connection.

In chapter 4 I argued that in a unified whole, each part's experience (or their objects) adumbrates the other parts' experiences (or their objects). This adumbration can be more or less informative as to the concealed aspect, and its concealment can be more or less salient. Moreover, I suggested that in human minds, each part would adumbrate the others very informatively, while the fact of concealment would have very low salience. PP helps to explain why this is: concealment is not salient to the parts because there is harmonious interaction between the revealed and concealed aspects, i.e. between their own experiences and the others'.

This means that in one sense the parts are all aware of roughly the same thing – the same unified experiential whole. But they differ in which parts are revealed and which concealed, as well as in how informatively the concealed parts are adumbrated by the revealed parts. We might compare the composite mind to a social institution whose members identify with it very strongly. Labour is divided among different members (e.g. one undertakes research, one decides on policies in light of research, and a third implements policies), and the whole can be described as performing all the members' tasks as long as there is the right sort of communication between them (e.g. the researcher informs the decider what policies the data support, the decider picks one and informs the executor, and the executor implements the policies which the decider picks). Moreover, we may imagine the members constantly updating each other with summaries of their progress. Each is thereby aware of the overall activities of the group as a whole, and so in a sense they are all aware of the same things. But each particular activity is known in much more detail to one person than to the others, since that person is doing it themselves and the others are just 'receiving the memos'.

Of course, in a human institution the parts recognise their distinctness: the decider distinguishes sending instructions to do X from actually doing X, the executor distinguishes doing X because of receiving instructions from doing X for their own reasons, etc. But we can imagine them growing more and more slavish and automatic in their relations to the organisation, losing sight of its actual structure or the possibility of working outside that structure, so that when the decider sees good reasons for policy X, they ‘simply implement X’, as oblivious to the mechanism they rely upon (instructions sent to the executor) as we are oblivious to the mechanism by which we ‘just raise our arm’. And when the executor receives instructions to do X, they feel simply that ‘X is a good idea’, as oblivious to the origins of that perception (instructions sent from the decider) as we are oblivious to the origins of our hunches, impulses, and perceptual judgements. The thinking parts of one human being are likely to be far more slavish and automatic in their relations to each other than human members of an institution ever are, and hence *unable* to draw the distinctions which our imagined humans did not *bother* to draw.

Note that I have left open exactly what mechanisms implement and determine ‘harmony’ in the actual human brain, or in any other possible brain. That is an empirical question: PP just says that something about their interactions is responsible. However we filled out the details, PP is opposed to the ‘hybrid’ account on which we have direct insight into the fact that certain events occur in our own minds, or are our own decisions, allowing us to self-ascribe certain events simply because they really are ours. I do not think that the hybrid account can be conclusively refuted; however, all combinationists need is that PP be a live option. For then they have a response available to the Problem of Oblivious Parts, and that problem is neutralised. They might, however, claim in addition that PP is independently preferable to the hybrid account, either because it is simpler (in appealing to a single broad type of factor), or because it is more naturalistic, in eschewing appeal to direct metaphysical insight into causal facts. They might also claim that the patterning principle is better able to explain the impression of ‘thought insertion’ reported by schizophrenics. I will not attempt to judge which account

is independently preferable: I simply conclude that the Problem of Oblivious Parts can be solved by adopting PP.

Section 2: What Does ‘I’ Mean?

Our usual understanding of first-person thoughts and utterances combines two plausible ideas, which seem to conflict when we accept combinationism. In this section I consider possible replacement principles, arguing that there are several coherent options available to combinationists, especially if we already accept the Patterning Principle defended in the last section.

Subsection 2.1: Secure Self-reference and Epistemic Optimism

Consider an example: a human being thinks to themselves truly “I weigh 60kg”, and since this thought is going on inside their head, it seems *prima facie* that if the head is a conscious subject it also thinks this thought. But what does it thereby think? If it thinks that *it* weighs 60kg, it will be wrong, since it weighs far less than that – and this error will not be due to faulty evidence, bad reasoning, or any kind of rational failing on the head’s part. By contrast, if it thinks that *the person* weighs 60kg, it will be correct, but will be in the strange position of referring to someone else by thinking ‘I’ – and doing so apparently without any conscious attempt at insincerity, play-acting, quotation, etc. A similar dilemma will appear with any shared thought which self-ascribes an unshared feature - whether that feature is physical, mental, or something else.⁷ This forces us to choose between two intuitively attractive ideas:

⁷ These thoughts may not need to be linguistically structured, with singular pronouns and so on; if not, my talk of ‘first-person terms’ should be read as picking out whatever element or aspect of the thought serves to refer to the subject in the way that first-person singular pronouns do.

Security of Self-Reference: A subject consciously thinking a first-person thought always refers to themselves by the first-person term in that thought – there is no risk of it referring to something else, or failing to refer.

Epistemic Optimism: A subject who commits no ‘epistemic wrongdoing’ (e.g. never believes something for which they have insufficient evidence, etc.) will not inevitably and systematically believe falsehood.

I will call these theses ‘Security’ and ‘Optimism’ for short. Security is the more straightforward, since it seems basically definitional of first-person terms that they refer to their users. Optimism is less straightforward because it does not seem like a conceptual truth. Indeed, it is implausible if construed as a necessity claim - it certainly seems that subjects *could* be trapped in a deceptive environment where they are systematically wrong, as we imagine whenever we entertain some sceptical hypothesis. But we usually assume that we are *not* in such an environment - that the actual world is not a sceptical scenario. It is an interesting and vexed question what justifies this assumption, but we do at least seem to make it, and consequently there is something unsettling about a theory which implies that a great many subjects are in precisely such a situation.⁸

It is hard to combine Security and Optimism when a self-ascription of an unshared feature is shared (such as in the judgement ‘I weigh 60kg’). For the whole they are compatible, since the ‘I’ in the thought refers to the whole, its subject, and the thought is true. But what does the head (or the frontal lobe, or the top half, or any component subject that shares the thought) refer to by the ‘I’ in this thought? If it refers to the whole human being, then Security is violated, for someone has referred to someone else with ‘I’. On the other hand, if it refers to itself then it thinks something false (that a head weighs 60kg), and not due to any violation of epistemic rationality, any failure to properly respond to evidence – for it performs all the same epistemic acts, based on all the same evidence, as the whole

⁸ We could also base epistemic optimism on a sort of charity: we should avoid construing the content of a subject’s thoughts in such a way that they would be systematically and unavoidably wrong about things.

human being does. This blameless error will not even be a random fluke event, but a systematic sort of error which the head, considered as a subject, can never escape or even detect. Thus Optimism is violated.

Experience-first combinationists will need different examples (since they do not regard my head as a component subject) but the same problem will arise. For instance, suppose that one well-integrated subset of my experiences contains all and only my left-hemisphere-based experiences, and constitutes a component subject: suppose also that it shares with me some such thought as ‘I am in full and direct motor control of two arms’, or ‘I receive sensory stimulation from all the nerves in this body’, or even just ‘I am a maximal conscious being, not contained within any other’. Here again, Security and Optimism seem to be at odds.

In the remainder of this section I review some available accounts of the reference of first-personal terms: some accept Security and give up Optimism, other accept Optimism and give up Security, and others still try to preserve both together. But first I note the possibility of a significant and interesting restriction on the scope of the Problem of Self-Reference.

Subsection 2.2: An Exception for Introspective Self-Ascriptions?

It may be impossible for an *introspective* self-ascription - a self-ascription of a conscious state, based on knowledge ‘from the inside’ - to be shared by subjects who do not share the property ascribed.. This will be the case if we adopt what I will call the ‘containment’ model of introspection, on which an introspective thought about an experience (or that thought’s neural realisation) contains the experience itself as a proper part (or contains its neural realisation as a proper part). Shoemaker (1994) defends such a model, opposing it to any sort of ‘distinct existences’ model (as defended by Armstrong 1963)

on which an experience, and an introspective thought about it, are metaphysically separate events whose connection is entirely contingent. On the containment model introspection does not involve the arising of second-order mental states that are merely caused by first-order ones, but rather builds additional cognitive structure around first-order mental states in such a way that the complex formed of first-order state and additional cognitive structure is a second-order mental state.⁹

If we accept the containment model, then no introspective self-ascription could be shared with component subjects that do not also share the experience self-ascribed. For any component subject that had the introspective thought would contain as a proper part a component subject that had the self-ascribed experience, and would inherit this experience. The only way it might not inherit the experience would be if that proper part was isolated from all its other parts, in which case it is hard to see how it could possibly give rise to the introspective thought in the first place.

For instance, suppose that when I visually experience redness, processes occurring in my occipital lobe-and-brainstem are intrinsically sufficient for the occurrence of this experience, whereas introspection requires the activity of my frontal lobe. On the ‘distinct existences’ model, the introspective self-ascription “I am visually experiencing redness” might be something entirely going on in the frontal lobe, which does not visually experience redness though it receives signals of such experiences from the occipital lobe (and, in accordance with the Patterning Principle, self-ascribes the experiences those signals tell it of). Then a conflict between Security and Optimism would arise when

⁹ One of the theses Shoemaker takes this model to support and explain is the ‘Immunity to Error through Misidentification’ of introspective self-ascriptions (cf. Shoemaker 1968, Wittgenstein 1958, pp.67-68): while they may be false, they can never be false just on account of misidentifying who their predicate applies to – it is never the case that the subject knows that *someone* satisfies the predicate, but goes wrong in taking themselves to be that subject. (Some authors have argued that other judgements, such as bodily judgements like ‘my arms are folded’ can also be immune to error through misidentification, e.g. Evans 1982, pp.218-225, cf. Prosser & Recanati 2012.) Wherever a combinationist gives up ‘Epistemic Optimism’ they must also give up Immunity to Error through Misidentification, for the falsity of the self-ascription will be based specifically on mistaking something true of the whole one belongs to for something true of oneself. Conversely, if combinationists can save Epistemic Optimism for some set of self-ascriptions (as the Containment Model seems to for introspective ones), they will also save Immunity to Error through Misidentification, if that is independently plausible.

we considered the frontal lobe as a subject thinking “I am visually experiencing redness”. But on the containment model, the introspective self-ascription contains the ascribed experience as a part, and so cannot be ascribed to any subject containing the frontal lobe but excluding the occipital lobe. Rather, the self-ascription would be an event spread over both lobes, ascribable only to a subject containing both (note that defenders of the two models might agree entirely about the causal story linking occipital-lobe events to frontal-lobe events: they simply point to larger or smaller sets of those events when asked ‘what is the introspective thought?’).

However, even if combinationists adopt the containment model, a residue of the Problem of Self-Reference might remain for introspective self-ascriptions. For consider the frontal lobe itself, or more realistically a component subject that contains it but excludes the occipital lobe (e.g. my brain’s occipital-lobe complement’). What is it like to be this subject, when an introspective self-ascription is being thought? Having denied that it experiences redness, and thus denied that it is thinking “I am experiencing redness”, it becomes unclear just what it *is* thinking. Perhaps it is not thinking anything - perhaps it is too gerrymandered a being to have any conscious states at all. But if it is consciously thinking, perhaps it is thinking an ‘incomplete’ thought, whose content we might render as “I am experiencing *that*”, where ‘that’ is accompanied by a phenomenology adumbrating the occipital lobe’s experiences, and thus refers to phenomenal redness. And in relation to this incomplete thought, there would be a conflict between Security and Optimism, for if the ‘that’ refers to phenomenal redness, and the ‘I’ refers to the frontal lobe, then the thought will be false even as the whole brain’s complete thought is true.

Admittedly this notion of an incomplete introspective thought is odd, for according to PP the frontal lobe cannot distinguish self-ascribing redness-experience in the normal way, which involves the experience itself, and self-ascribing it in this indirect demonstrative way, which does not. Of course the

frontal lobe does have some phenomenological grip on redness, by its informative adumbration of the occipital lobe's experience - but what is this like? Perhaps the best thing for combinationists to say is that occipital lobe complements do not think any recognisable thought at all when they are involved, together with the occipital lobe, in an introspective self-ascription.

Subsection 2.3: Relativised Reference

One response to the conflict between Security and Optimism would be to simply give up Optimism while retaining Security. The natural way to achieve this would be a doctrine of *relativised reference*, with a single though having different contents, and a single term different referents, relative to each subject thinking it. We cannot evaluate the thought's referential properties absolutely, but only relative to one of the subjects that thinks it.

Heller 2000 has already employed the idea of multiple contents for first-person terms, though in a diachronic case involving personal fission. He defends the coherence of multiple contents for one token thought or word using the example diagram to the right, showing “a

This	sentence	contains	exactly	six	words.
sentence					
contains					
exactly					
three					
words.					

crossword puzzle with words in each square instead of letters”(p.375). He writes that “the across-sentence is true. The down-sentence is false. The single token of ‘this’ has two contents, referring to both the across-sentence and the down-sentence.”(p.376).¹⁰

¹⁰ Alternatively, Heller notes, this could also be seen as an extension of context-sensitivity (which indexicals such as ‘I’ are already taken to have), but with multiple contexts simultaneously present.

The cost of this strategy is that the thought may be false relative to many subjects. If the thought “I weigh 60kg” means that the head weighs 60kg, as thought by the head, and that the brain weighs 60kg, as thought by the brain, and so on, then it will be false on most relativisations. Similarly, if the thought “I am experiencing redness”, as thought by the frontal lobe, means that the frontal lobe is experiencing redness, then it will be false for the frontal lobe. And this amounts to abandoning Epistemic Optimism: many parts will regularly be wrong about their own properties, despite drawing only justified conclusions from their evidence.¹¹

Subsection 2.4: Privileged Reference

Alternatively, we might maintain Optimism while abandoning Security, by adopting a doctrine of *privileged reference*. Whenever any part of me thinks a first-personal thought, it refers not to itself but to the whole human being – and thus the thought is true, barring some epistemic failure or misfortune. The semantics of ‘I’ are not what we had thought (a straightforward ‘the subject thinking/speaking this thought/utterance’), but rather include a criterion by which to privilege one particular subject over all others. This criterion might be mereological (e.g. “the maximal subject thinking this thought”), or organisational (e.g. “the most integrated subject thinking this thought”), or ontological (e.g. “the most fundamental subject thinking this thought”).

Something like this view has already been defended. For instance, Noonan 2010 affirms that human animals (essentially biological beings) can think first-personal thoughts, but that these thoughts refer not to the animal but to the person (an essentially psychological being), which coincides with the

¹¹ One might go further and relativise *justification* to subjects too, hoping thereby to explain the errors of the parts by saying that the inferences they (and the whole) performed were justified relative to the whole, but not relative to them. But then the problem reappears as the threat that it may be impossible to design a system which will draw only inferences justified relative to all its parts.

animal. And Heller 2000 says that “When a part of a person has an ‘I’ thought, it is a thought about the person, not about the part. (That is why not one of my ‘I’ thoughts is a thought about my brain or a part of my brain.)”(p.377)

The cost of this view is that it makes some subjects are apparently unable to refer to themselves in a first-personal mode, which we might find implausible. It seems that the point of first-personal thoughts is that they let subjects think about themselves, without having to pick themselves out by some fallible description.¹²

One option is to say that which subject is privileged depends on what concept is employed in the first-personal thought, so that we can in a sense choose what to refer to with ‘I’ by choosing what concept to employ in our ‘concept of self’. Thus Parfit at one point declares that, if persons have biological persistence conditions, then the first-person pronouns used in his 1984, from page 291 onwards, all refer to ‘series-persons’ rather than persons (a ‘series-person’ is a series of psychologically connected persons). Heller expresses the principle thus: “A cat’s ‘I’ thought would only refer to a cat if the cat’s concept of self involved the concept of a cat, just as [a normal human person’s] concept of self involves her concept of a person.”(p.378) Similarly when a human head thinks first-personal thoughts, it might do so using a ‘concept of self’ which involves the concept of a whole human, not a head.¹³ But the head is not precluded from ever thinking of itself first-personally, since if it (and, simultaneously, the whole human) were to involve the concept of a head in its self-concept, its first-person pronouns

¹² Moreover, privileged reference makes even the self-referential ability of human individuals contingent on certain somewhat abstruse questions about ontology. If ‘I’ refers to the mereologically largest subject, then whether ‘I’ refers to a human being or to the universe depends on whether the universe exists, and whether it qualifies as a thinker of thoughts that occur in it. Similarly, if ‘I’ refers to the most organised, or most fundamental, subject, then whether it refers to a human being turns on the relative fundamentality or organisation of human beings, as opposed to such things as bodies, organs, and aggregates of particles.

¹³ One wrinkle here is that people taking this line usually suggest that the concept we do in fact employ is that of a person, and arguably human heads are people, since they lack no mental capacity that human beings have. To single out the whole human being, we would need to build in some criterion which distinguished that person from others, such as maximality.

would then refer to itself. This might or might not be a sufficient mitigation of the loss of Security:¹⁴ moreover, it has the strange result that if someone says “I am a human head”, and really has incorporated the concept ‘human head’ into their sense of self, we should agree (while bearing in mind that there is a whole human being here who is not a head but refers to one with their ‘I-thoughts’).

Subsection 2.5: Indeterminate Reference

Some readers might find relativised or privileged reference satisfying, despite their substantially abandoning Security or Optimism. Others might find both unacceptable, or might at least hope for an approach which can preserve both theses. I will outline two candidates for such an approach. The first turns on the idea of *indeterminate reference*. This is the idea (often employed in discussions of vagueness, e.g. Lewis 1986 p.212, Lewis 1993, pp.28ff) that a referring term may have a set of referents, with no fact as to which one of those is *the* referent. For the thought involving that term to be true, it must be ‘super-true’, true for all candidate referents, while for it to be false it must be ‘super-false’, false for all candidate referents. If it is neither super-true nor super-false, it is indeterminate in truth-value.

Now, one use of indeterminate reference would be as an adjunct to the privileged reference approach, if we thought that the criterion for privilege did not pick out a single unique referent (e.g. we might think, with Unger, that there cannot be a single unique maximal human being). But we could also employ indeterminate reference as a rival approach, with all the subjects thinking a thought being

¹⁴ We might find it implausible that *all* first-person thoughts are concept-dependent in this way: is it not possible to self-refer without resting on a prior idea of what one is? It seems that we could meaningfully think to ourselves “what am I?” without presupposing an answer (Cf. Perry 1979, p.171 ff), or without presupposing an answer beyond ‘a rational thinking being’. This same worry will recur in sharper form in subsection 2.6: if the solution I advance there is adequate it could also be appropriated by a defender of the privileged reference approach.

indeterminately referred to by its first-personal terms. Thus it could be true ('super-true') that I am inside a building, false ('super-false') that I am an elephant, but indeterminate whether I am a head.

This lets us say that while not all subjects can hope for true first-personal thoughts if they do everything right, they can at least hope for their thoughts to not be *determinately false*, since there will always be at least one subject for which they are true, barring epistemic failing or misfortune. And it lets us say that while subjects cannot always securely refer to themselves by means of 'I', they can at least *indeterminately* refer to themselves, along with many other subjects. The cost of this strategy is that neither thesis is retained in its determinate form. Self-reference is not determinately secure, and some subjects are still denied any hope of systematically being *right* about themselves. Someone really attracted to Epistemic Optimism might find this an underwhelming sort of vindication: thinking something whose truth-value is indeterminate still seems like a significant failure, if the goal was to think something true.

Subsection 2.6: Dual Reference

We can combine the best elements of privileged reference and relativised reference through what I call a 'dual reference' approach, on which a subject thinking first-personal thoughts refers *both* to itself, in a relativised way, and *also* to the privileged subject, the thing which our first-person ascriptions are actually true of (absent epistemic failing or misfortune), and moreover affirms a certain relationship (possibly identity) between the two subjects it refers to.

For an example of how this might work, suppose there are three senses of 'I', a broad one ('I_B') and two narrow ones, one 'relativised' and one 'privileged' ('I_R' and 'I_P'). I_R refers to whichever subject of a thought we are evaluating the thought relative to, while I_P refers to whichever subject of a

thought best fits a certain criterion. But ordinary uses of ‘I’ are actually uses of I_B , which is more complex:

“ $\phi(I_B)$ ” means “ $\phi(I_p)$ and $((I_R = I_p)$ or $(I_R$ have a certain relation to $I_p))$ ”

So for instance, when the thought “I weigh 60kg” occurs in someone’s brain, each of the overlapping subjects thinking that thought is thinking that they *either* are, or are somehow related to, the whole human, which weighs 60kg. As long as each of those subjects really is related to the whole in the manner claimed, they will all think something true, even while referring to themselves.

This is still a very schematic proposal: it leaves open what exactly makes one subject the ‘privileged’ one – merely being the biggest, or being the most well-organised, or the most ontologically fundamental, or some combination of factors. It also leaves open what exactly the relation asserted between the two subjects is – merely proper parthood, or something stronger.

But even at this level of abstraction, this approach faces significant difficulties. For a start, it seems an awfully complex concept to be wrapped up unrecognised in the simple word ‘I’. Perhaps this will be tolerable if we think that self-reference is a very complex cognitive achievement, accomplished only by a handful of higher animals, but it will seem strange if we think that ‘all consciousness is self-consciousness’.¹⁵

Second, there seems to be a big difference between the formula’s two disjuncts – being identical to something, and being merely related to it. Generally when we think disjunctions whose disjuncts are importantly different we wonder which disjunct is true, and seek to find out. We do not

¹⁵ Would it mitigate this worry to say that the formula given is not the *meaning* of ‘I’, but simply a formulation of the rules that determine its meaning on each particular occasion (rather as Kripkeans think that ‘whoever taught philosophy to Alexander the Great’ is no part of the *meaning* of ‘Aristotle’, but rather is just part of what in fact fixes the reference of the name, Kripke 1980, p.74ff)? That depends on questions about the notion of meaning that are beyond my current topic.

rest content with the disjunction. For example, if I ask “is my mother dead?” the answer “she is either dead, or sick” will not satisfy me, even if it is all I can presently know. Why should it be that when we ascribe some feature to ‘ourselves’, we leave open whether we actually have that feature, or are merely related to something that does, and feel no dissatisfaction with that uncertainty?¹⁶

Finally, the dual-reference proposal may just push the problem back a step. If we still have I_R available to us, it seems we should be able to think thoughts using it, like “ I_R weigh 60kg”. If we can’t think this, why not? If we can, why is this not the thought we do think in everyday situations when we think that we weigh 60kg? Isn’t that what we take ourselves to be thinking? Indeed, isn’t that what we set out to think?

I think these difficulties are resolved if we formulate the dual reference approach using the concept of ‘harmonious connection’ developed in the last section. Then $\varphi(I_B)$ would mean:

“ $\varphi(I_p)$ and ($(I_R = I_p)$ or (I_R am a proper part of I_p harmoniously connected to the other parts))”

Note that this still leaves open what makes one subject the privileged one. All that is added is a specification of what the crucial relationship is: harmonious connection, i.e. being connected such that what each does is so co-ordinated with the psychology of the other that each perceives the other’s doings as their own.¹⁷

I believe that this version of the dual-reference approach addresses the above objections, as long as the Patterning Principle is true. For then being something, and being a harmoniously connected part of it, are equivalent in how they seem and what evidence we can have for them. We draw on the

¹⁶ It will not help to say that the relation is proper parthood, and then point out that the disjunction ‘either identical to or a proper part of’ is defined, by formal mereologists, in terms of the non-disjunctive concept ‘part of’. Even if we took the definitional order used by formal mereologists as a guide to our actual concepts, there remains an important difference between being a whole human being and being merely a lobe thereof.

¹⁷ Moreover, this adds a further open question, of how many other parts each subject must be harmoniously connected to, and how harmoniously connected they must be. I here suppose that the answers to these questions will be found in more general semantic inquiries, such as into the workings of vague relations.

same factors to assert either – harmonious interactions over a period of time. Since these factors generate our sense of self, it is not surprising that they be involved in the meaning of our self-referential terms. And given how we actually make judgements of identity and parthood, we will naturally neglect the difference between the two disjuncts of the formula for I_B .

What about the difficulty over the availability of the narrow senses of ‘I’? Even if subjects do typically think in terms of I_B , governed by the above formula, can they nevertheless formulate thoughts involving I_R , or I_p ? I think the answer is that they *can*, but only when they undertake a special, and usually unmotivated, effort to refine and specify their meaning. We do not generally have any reason to think in such terms, nor any inclination to do so: when we try to think directly and immediately of ‘ourselves’, we do not ascend to the heights of precise formulations, but descend into the intuitive vividness of our natural, spontaneous self-apprehension, which is a matter of perceiving harmony.

Nevertheless these thoughts are possible, for those who endeavour to formulate them. Thoughts involving I_p will be equivalent to thinking about the largest, most integrated, or most fundamental being involved in that thought – so they will not refer to any of the other subjects, which are mere parts of that thing. But this does not threaten Security, because this is not the primary sense of ‘I’, and that primary sense allows all the other subjects to refer to themselves.

Conversely, thoughts involving I_R will have different contents for each subject, and will be true for each only if they ascribe a feature which is shared by all those subjects. Those subjects for whom the thought is false will go wrong through misidentifying themselves with the whole for which it is true. But this does not threaten Optimism, because neither they nor the whole are justified in endorsing these thoughts involving I_R , in the first place. As I will explain in section 3, these subjects and the wholes they compose should all be agnostic about their own identity. They can be sure of which harmoniously-connected system they are part of, but not whether they are the whole system or some

part of it. Here it is important that these errors are not ones that occur in the ordinary course of life, for in the ordinary course of life we talk and think in terms of I_B , not I_R . These are errors that result from a special decision to use pronouns in an unusual way, combined with a decision to keep using the sorts of evidence that would normally be relevant. Thus they are traceable to an epistemic mistake, and do not threaten Epistemic Optimism.

I conclude that *if* PP holds, then the dual-reference approach to the meaning of first-person pronouns is defensible, and that approach resolves the Problem of Self-Reference by preserving both Security and Optimism. Even if that approach is rejected, combinationists have options: they can preserve Security with the relativised reference approach, they can preserve Optimism with the privileged reference approach, or they can preserve a weakened form of both with the indeterminate reference approach. Of these, I think the privileged reference approach is probably the best, though in sacrificing Security it is inferior to the combination of the dual-reference approach and PP. Since I argued in the last section that the PP may well be true, I conclude that combinationists are well-placed to address the two problems of Oblivious Parts and Self-Reference. But the problem of Self-Identification will be less tractable.

Section 3: Who am I?

When many harmoniously-connected subjects overlap, how is any of them to determine which it is? Equivalently, how do *I* know if I am a human being, a human brain, or something else? I will refer to this epistemic task as ‘self-identifying’: to self-identify relative to some subjects is to know which of those subjects one is.¹⁸ In the first two subsections of this section I argue that if combinationism is true,

¹⁸ Of course, if the ‘I’ in ‘who am I?’ is I_B or I_p , then the question can be answered by finding out which of the subjects asking the question is the ‘privileged’ one, by whatever criterion of privilege. But the Problem of Self-Identification would be posed by asking the question using I_R .

we cannot self-identify relative to many of the overlapping, harmoniously connected subjects present in a human being, though we can be sure that we are among the most cognitively sophisticated. In the final two subsections I then argue that this result need not be a *reductio ad absurdum*, though it may be, depending on what position one takes on independent questions of ethics and ontology.

Subsection 3.1: Can We Self-Identify Relative to Our Experientially Equivalent Parts?

I will consider two classes of subjects relative to which we might wonder about self-identification, which I call ‘experientially equivalent’ and ‘experientially different’ subjects. It is fairly easy to show that experientially equivalent subjects cannot self-identify relative to each other; in the next subsection I argue that if PP is true, many experientially different subjects will *also* be incapable of self-identification relative to each other.

By ‘experientially equivalent subjects’ I mean subjects who differ only in parts which make no contribution to experience, such as a whole human being and their head. These are clearly different things, but they will also, on plausible assumptions about the basis of experience, have exactly identical mental lives. (Of course, if subjects are constituted by their experiences, there may be no such thing as experientially-equivalent subjects: perhaps any two subjects that share the same experiences are identical. So the experience-first combinationist faces the problems of the next subsection, not those of this one.)

To see why experientially equivalent subjects would be unable to self-identify, observe that the only way to self-identify seems to be making connections between first-personal and third-personal knowledge. We learn empirically about the existence of various subjects, and learn that various things are true of them. Simultaneously, we witness our own experiences and derive from them various

judgements true of ourselves. By comparing the set of judgements we know to be true of ourselves with those we know to be true of various subjects, we can deduce which of those subjects we are. So for instance, if I hear someone say “the guy who gave the talk this morning spoke too fast”, and I have memories of giving a talk this morning (and evidence that only one person did so), I can infer that the guy being referred to, who spoke too fast, is myself.

Given this, it will be impossible for two subjects to self-identify relative to each other if exactly the same things are true of both, such as two qualitatively identical people located in qualitatively identical surroundings on either side of a symmetrical universe. Self-identification will also be impossible if exactly the same experiential judgements are true of both, even they differ in other ways, such as a real human walking around, and a qualitatively identical human brain held in a fluid-filled container being fed electrical stimulation precisely mirroring those received by the former human’s brain.

So the mere possibility that subjects be unable to self-identify cannot be considered a *reductio*. What matters more is the actuality of it: combinationism likely implies that real people (and their parts) are all in fact unable to discriminate themselves from a large number of other subjects. All experiential facts will be equivalent among experientially equivalent subjects, so how could they self-identify relative to each other? Each will affirm of itself the same set of experiential judgements, and if it has accurate information about the set of subjects, will know that the same set of judgements is true for each. Hence there is nothing by which to discriminate among them.

It might be thought that a whole human being could discriminate itself from its head by the fact that it, for instance, had direct first-personal knowledge of the position of its legs. But, given the plausible idea that certain brain events are necessary and sufficient for leg-experiences (as suggested by numbness resulting from spinal cord damage, on the one hand, and dreams or phantom limb symptoms

on the other), the head will have experientially-identical knowledge of the position of the legs harmoniously connected to it.

This conclusion can be reinforced by observing that if one of these subjects were able to self-identify, its self-identifying judgement (e.g. “I am a head”) will also be made, falsely, by the other (e.g. the human will think it is a head). So any success in self-identification for one would mean a failure of self-identification for the other, though it seems arbitrary which should succeed and which fail. This should encourage us to accept the impossibility of self-identification for both.

Subsection 3.2: Can We Self-Identify Relative to Our Experientially Different Parts?

What about subjects who differ significantly in their experiencing parts – such as the brain and its occipital-lobe-complement? Unlike experientially equivalent subjects, they are not prevented from self-identifying relative to each other by the lack of an experiential difference. However, self-identification requires not just a difference in experiences, but a consequent difference in judgements about those experiences, and it seems that in many cases an introspective judgement made by one subject will have to be shared by others (e.g. wholes containing it as a highly-integrated part) even if it is true of one but not the other.

Consider section 2’s example where my occipital lobe experiences redness, and my frontal lobe subserves introspective judgements. Whenever the whole brain judges that it is experiencing redness, the occipital lobe complement will make the same judgement, even if it is not. If introspective thoughts contain their objects as proper parts, perhaps the occipital lobe complement will not *have* the introspective thought in full, but it will have an incomplete thought in which it adumbrates redness as ‘that’ and self-ascribes an experience of ‘that’. Even if, due to the semantics of the first-personal terms

employed, this judgement made by the frontal lobe is *true*, it will not allow the occipital lobe complement and the whole brain to self-identify, since each makes introspective judgements that are, if not identical, indiscernible from their own subject's perspective. To get different judgements requires not just different experiences, but also independent mechanisms for forming judgements based on those experiences.

The Patterning Principle explains why this difficulty arises. Harmoniously connected subjects make introspective judgements not based strictly on their actual experiences, but on all those experiences which they learn of *via* harmonious interactions, including those of other subjects. So even when experiences differ, the judgements may be based on the same set of experiences. This problem could be seen as the descendant of chapter 4's 'boundary problem', since it involves the failure of 'positive epistemic boundedness' - knowing which experiences are and are not one's own. But isn't boundedness phenomenologically evident? In a sense: our experience gives us a clear sense of *some* boundaries, namely those of the set of experiences belonging to whatever harmoniously-connected whole we are part of. And for that whole, this phenomenological impression of boundaries will be accurate - it will encompass just those experiences the whole has. But even for the whole this may not qualify as positive epistemic boundedness, for if the whole cannot *know* that it is the whole, it cannot *know* that these experiences are all its own.

Recall the social analogy discussed in subsection 1.4: our parts are like especially slavish and deferential employees of an institution, who do not bother to distinguish what they do from what the group does. Consequently if, so to speak, one member of the group sends around a query asking "is anyone doing job X?", and some other member replies saying "yes, I am", each member will self-ascribe both the question and the answer: each member will feel as though they introspected upon their activities, looking for a particular one, and found it. Even if one component subject is not involved

in X, they cannot recognise this fact without ‘asking the question’, and as soon as they do so, that other subject will ‘pipe up’, interfering with the first subject’s attempts to ascertain which experiences they are and are not having.

Something that the social analogy does not capture is the variety of sizes among these parts – the brain’s parts are also made of smaller parts, and so on. The smaller a component subject is, the less of the mind’s processing is fully revealed to it (known ‘directly’) and the less informatively the processing of other parts is adumbrated. At one extreme, to the whole everything is revealed (setting aside adumbration of external things), while at the other extreme, a single neurone may experience only a very small element of the phenomenal field as revealed, with the rest of it being adumbrated in very uninformative terms, merely as the background which its own experiences open out onto.

For these reasons, I believe that even experientially quite different subjects will be unable to self-identify relative to each other. How far does this go? Fortunately, there is a limit: the whole brain will not be precluded from discriminating itself from a single neurone. Even if two experientially different subjects cannot form different negative judgements (one thinking ‘I am experiencing x’, the other thinking ‘I am not experiencing x’), they may still form larger or smaller sets of positive judgements.

Self-ascribing an experience plausibly requires understanding what experience it is, as well as understanding what it is for a subject to have an experience; one cannot affirm what one does not understand. So whatever it adumbrates, no subject incapable of understanding that subjects have experiences will self-ascribe any experience. It follows that each component subject will only self-ascribe experiences insofar as it is conceptually capable of doing so. This means that if it seems to me introspectively that I am experiencing redness, I cannot be sure that this feeling is strictly and literally in me (and thereby cannot use, e.g., neuroscientific data about the neural correlates of colour

experience to self-identify), but I *can* be sure that I am a subject with enough cognitive sophistication to understand what redness-experiences are and self-ascribe them. Thus I can at least know that I am not a neurone, given the plausible premise that neurones are not capable of that. But if many overlapping sections of the brain are sophisticated enough to understand and self-ascribe experiences, I cannot know which of them is me, and will have to adopt a stance of agnosticism towards my own identity.

Subsection 3.3: The Importance of Self-Identification

It seems, then, that accepting experiential combination means accepting that we cannot self-identify relative to a fairly large class of subjects: we cannot know exactly who we are. This might be a *reductio ad absurdum* of experiential combination, or merely a surprising but bearable result. Which it is will turn out to depend on what view we take on certain independent questions, which I discuss in these final two subsections.¹⁹

There would be something unsettling in the impossibility of self-identification if it meant that we were each necessarily ignorant of an important fact, namely ‘who we are’. However, it is hard to say exactly what sort of fact this is for us to be ignorant of. What is it for one particular subject, out of all those which exist, to be *me*? We might think that it is a fairly trivial fact. When I declare “I am Luke Roelofs”, the truth conditions for that utterance are that it was made by Luke Roelofs ; when I ask “who am I?”, the question thus appears to be equivalent to asking “who is presently speaking?” The answer to this question can be supplied by the objective statement that “Luke Roelofs is presently speaking”, which we can discern empirically without any special need for self-identification. Experiential combination would complicate this picture but not substantially change it: the answer to “who is

¹⁹ From here onwards when I speak of the importance of self-identification, I mean the sort of self-identification I have argued would be impossible given combinationism – identification relative to cognitively sophisticated subjects.

presently speaking?” would refer to a multitude of subjects (and possibly to subtly different senses of the word ‘speaking’), but could still be completely objective and non-indexical, with no special need for self-identification.

Various other facts can be conveyed by self-identifying judgements. If you already know that Luke Roelofs is from London, and I tell you, while we are sitting in Toronto, that ‘I am Luke Roelofs’, that will enable you to know the objective fact that someone from London is presently in Toronto. But this kind of objective information is only contingently conveyed by self-locating judgements: being incapable of self-location will not stop us from learning all the objective facts about how many subjects there are, how related, with what properties and what histories.

Some people may feel, though, that the fact that *I* am a certain subject is a fact beyond just these ordinary objective facts. To bring this out, imagine a complete objective description of the world – an exhaustive litany of everything that exists and everything that happens, physical and mental, at each point in space and time. This description includes no indexical terms like ‘I’ or ‘now’, or ‘this’, and privileges no particular subject over any other. It is, so to speak, ‘centreless’. This description does not say who *you* are. It says, of course, that various thoughts and utterances in the world contain indexical terms, and it specifies which worldly subject these worldly thoughts refer to. But that will not allow you to know which subject you are, without supplying some already-indexical knowledge (cf. Perry 1979, Lewis 1979, Nagel 1986, Seager 1990, Hellie 2013, forthcoming-b, p.14-15). Yet surely there is a fact of the matter as to which subject you are! It is a fact you have had to live with all your life, after all (though everyone else has had to live with an equivalent fact that they are themselves).²⁰

²⁰ There is a parallel difficulty concerning the present moment in time. An exhaustive history of the universe will not specify which time is the present, though it will of course specify that whenever anyone at any time speaks of ‘the present’, they refer to the time at which they speak.

And it seems a very important fact, in many ways (cf. Hellie 2013 pp.309-310). So what does this objective description leave out?

The importance of what is left out becomes obvious when we ask what a subject who knew *only* these objective facts could actually do with them. Suppose that among the objective facts is ‘Luke Roelofs’ trousers are on fire’. If I am Luke Roelofs, this fact gives me a good reason to remove my trousers; if I am not it gives me no such reason. If I do not know whether I am Luke Roelofs, then this objective knowledge, however complete, is practically useless to me, because I must act based on where *I* am.

Recognising the practical importance of self-identification still lets us ask what it is that has this importance. Perry considers the idea that there are propositions of ‘limited accessibility’: propositions that only a certain person can express. When I say ‘I am Luke Roelofs’, what I express with that sentence is not the same as what other people express with ‘I am Luke Roelofs’ (which would be false, not true), or what other people express with ‘Luke Roelofs is Luke Roelofs’ (for I could know the latter while doubting the former), and so must be a proposition which nobody except me could express: “Others can see, perhaps by analogy with their own case, that there is a proposition that I express, but it is in a sense inaccessible to them.”(1979, p.179)²¹

But these ‘inaccessible propositions’ are not the only way to make sense of what is left out of the objective world-description. On Perry’s preferred account, we must recognise belief states that do not correspond to any single object of belief: anyone can believe the proposition ‘Luke Roelofs is Luke Roelofs’ (or something contingent like ‘Luke Roelofs lives in Toronto’), and anyone can be in the

²¹ Hellie advances an even more radical account: “only my stream of consciousness is genuine... the unique stream of consciousness connects to the physical world by being the temporally extended perspective of *this creature*, the one who looks like the photo on my driver’s license” (2013, p.12). The supposed equal existence of other streams of consciousness is really just the rational mandate to *simulate* certain experiences in relation to certain other creatures. I think that combinationists are probably committed to rejecting this proposal, which Hellie calls ‘Inegalitarianism’, since combinationism assumes the existence of more than one stream of consciousness.

belief state that gives rise to the sentence ‘I am Luke Roelofs’ (or ‘I live in Toronto’). What is unique about me is that I can believe that proposition *by* being in that state, while other people would believe different propositions in that state, and would have to be in different states to believe the same proposition.

So combinationists could plausibly maintain that there is nothing left out. There are no irreducibly indexical facts for us to be ignorant of, merely indexical ways of expressing objective facts. There is nothing in the meaning of “I am Luke Roelofs” – nothing involved in this being true – beyond the fact that Luke Roelofs typed it. But even granting this, it is undeniable that self-identification has immense *practical* importance. How do combinationists think we can act in the world if we are, as the arguments above suggest, incapable of complete self-identification?

Subsection 3.4: Self-Identification as Action-Guiding

So if self-identification is important, its importance comes from its practical role. But what sort of role is this? How much does knowing what to do require self-identification? We can divide this into two questions: if I knew what my ultimate goals were, would I need to self-identify to know how to pursue them? And do I need to self-identify to know what my ultimate goals should be?

On the question of pursuing given goals, while self-identification in general is vital, combinationists can maintain that the kind of self-identification their theory rules out is not. This is because the alternatives among which we might be unable to decide – that we are a whole human being or a harmoniously-connected part of one – do not relevantly differ in what powers they make available. If a given hand and arm reliably do what I wish to have done, the further fact of whether I include them is irrelevant to their usefulness in serving my goals. If I know the objective fact that Luke Roelofs’

trousers are on fire, this will rationalise the same actions whether I am Luke Roelofs or Luke Roelofs' head, because Luke Roelofs' head has access to the same 'levers' as the whole.

This is not to say, obviously, that a human head has the same causal powers as a whole human being. But whatever a human being can do intentionally and deliberately with its body, the head can also bring about intentionally and deliberately, though often less directly. For instance, although they differ in weight, the head has control over enough external weight (the rest of the body, which it can manipulate using the fibres that connect it to the neck) that it can indirectly exert the same gravitational forces on things (e.g. can trip the same weight sensors) as the whole. What matters is not what 'I_R' can do, but what 'I_B' can do.

This applies even to experientially different parts, like a single cerebral hemisphere. The same objective fact that gave me reason to remove my trousers if I was Luke Roelofs ('Luke Roelofs' trousers are on fire') will, if I am in fact a single one of Luke Roelofs' cerebral hemispheres, give me reason to do something closely related: to initiate the movements in one half of Luke Roelofs' body that partly constitute trouser-removal, while influencing the other hemisphere to initiate the corresponding movements in the other half of that body.²² More precisely, what I have called 'influencing the other hemisphere' is a subtle mixture of monitoring its activities, checking that it is initiating those movements, and sending signals that will stimulate it to do so. The important point is that from the first hemisphere's perspective, this will feel just like initiating both sets of movements at once. Thus the actions called for by the fact 'Luke Roelofs' trousers are on fire' if I am Luke Roelofs or am one of his cerebral hemispheres are indistinguishable actions for me: the two levels of uncertainty cancel out, so that although I am not sure which action to perform, I am equally uncertain which action would be *this* one, but can be sure that *this* action is the right one to perform, whichever one that is. So

²² In reality, the important role of the cerebellum probably makes it problematic to divide the roles of the hemispheres in this way, but the example is merely illustrative.

when it comes to knowing which action to perform to promote my goals, the uncertainty produced by a failure of self-identification ‘comes out in the wash’.²³

What about knowledge of ultimate ends? Views which make self-interest especially important will make self-identification crucial. This includes views on which morality demands promoting only one’s own good, or especially one’s own good, or views on which ‘prudence’ (a non-moral form of practical rationality) demands such things. Whether someone is me would then be crucial to how much reason I have to promote their good, and it will be a major disability not to know who I am (see e.g. Hare 2007). The same applies to views on which, while my reasons for promoting my own good are the same as my reasons to promote others’, the permissible means of doing so do differ – e.g. it is permissible to inflict things on yourself which you must not inflict on others.

On the other hand, at least three coherent views of practical rationality make self-identification unnecessary. The first is the view that all subjects have, ultimately, the same reasons. Call this Ethical Neutralism (following Broad 1953). For example, a utilitarian must say that all subjects, ultimately, have reason to do whatever best promotes total utility, whether theirs or someone else’s. On such a view, knowing which subject’s utility is one’s own utility is irrelevant because your reason to promote utility is unaffected by whose it is.

Second, one might think that some category of wholes (all wholes, maximal wholes, highly integrated wholes, etc.) have absolute moral and rational priority over their parts. Call this Ethical Holism. On this view, each subject has reason to promote the wellbeing of the largest, best-integrated, or otherwise privileged subject that they are a part of – self-interest, for any subject other than the

²³ Similar remarks apply to the role of self-knowledge in guiding reasoning, which Burge (2000) argues is indispensable: only knowing which thoughts and reasons are ours can we know which thoughts are to be amended by which reasons in a direct and immediate fashion. Harmoniously connected subjects are precisely those who can directly and immediately act on each other’s thoughts in the light of each other’s reasons, so what is really necessary for reasoning is to know which thoughts and reasons belong *either* to us or to a harmoniously connected subject.

privileged whole, is a mistake and an illusion. Cerebral hemispheres have reason to work for the good of the whole human, regardless of their own fate – and perhaps human beings have reason to work for the good of their civilisation, their biosphere, or even their universe, regardless of their own fate.

Third, even a view which sharply distinguishes self-interest and altruism may then adjust this to allow the interests of beloved relatives and partners to acquire the status of self-interest. If, as Aristotle says, ‘a friend is another self’ (2007, p.165), perhaps caring for our friends is more like caring for ourselves than like caring for others. Call this ‘Self-Referential Altruism’ (cf. Broad 1953). And if this is granted, it might be that our own parts are similarly ‘morally bound’ to us, and to each other, so that their interests count as ‘self-interest’ for us and for each other.

So is the impossibility of self-identification a crippling problem? If one already holds to Ethical Neutralism, Ethical Holism, or Self-Referential Altruism, then it is not a problem; if one is undecided as to ethical matters, but attracted to experiential combination, this might serve as motivation to adopt one of these three views (compare Parfit’s discussion of the relation between ethics and his reductionism about persons, 1970, pp.26-27, 1984, pp.321-347). Each of these views could be worked out in different ways, but there are at least ways to work them out which secure that unimportance. On the other hand, if one is strongly committed to the sort of view that makes the strict self-other distinction ethically or prudentially important, one might take the impossibility of self-identification to be a *reductio ad absurdum* of any view which implies it.

Summary:

In this chapter I have discussed three problems facing experiential combination that in various ways involve self-knowledge: the Problem of Oblivious Parts, the Problem of Self-Reference, and the Problem of Self-Identification. I have argued that the first two of these problems can be resolved fully,

but not the Problem of self-identification. Combinationism renders self-identification impossible relative to the set of our experientially equivalent parts, and probably also relative to the set of our cognitively sophisticated parts. Rather than showing how self-identification is still possible, combinationists have to bite the bullet and claim that self-identification is not important: knowing which set of harmoniously-connected overlapping parts we belong to is all we need.

For convenience, we can group together the theses I have recommended for combinationists, and call them ‘the anti-individualist proposal’. This proposal comprises:

- First, the patterning principle, according to which our spontaneous impressions of which events are internal or external depend not on whether they really are, but on pattern-detection mechanisms;
- Second, the dual-reference approach to the semantics of first-person terms, on which the truth conditions for first-person ascriptions deal primarily with whether the self-ascribed property belongs to the maximal, most integrated, most fundamental, or otherwise privileged one of all the subjects of that ascription;
- Third, an acceptance of the impossibility of self-identification relative to the set of our large, conceptually sophisticated parts;
- Fourth, a Perry-style account of indexical thoughts and utterances, on which they do not express ‘propositions of limited availability’ but only distinctively action-guiding ways of believing mundane objective facts;
- Finally, some view like ethical neutralism, ethical holism, or self-referential altruism, on which I and my cognitively sophisticated parts have the same moral reasons for action.

Together these theses amount to a major deflation of the importance of being any specific individual.

No special proposition or fact corresponds to ‘me being me’, beyond some subject being self-identical - which is fortunate because, according to the patterning principle, my brain and its parts are not set up to detect any such special facts. Consequently I cannot know exactly which subject I am, but knowing that would not affect the truth or falsity of my self-ascriptions, or the rightness or wrongness of my actions. Individually, each of these claims might seem wild, but together they each make the others more plausible. They add up to a coherent view on which, metaphysically, epistemically, semantically, and ethically, it does not matter who I am: what matters is who I am connected to.

Chapter 8: Integration and Fusion

<u>Section 1: Recap</u>	<i>p.298</i>
<i>Subsection 1.1: The Core Proposal</i>	<i>p.300</i>
<i>Subsection 1.2: Objections to the Core Proposal</i>	<i>p.303</i>
<i>Subsection 1.3: Extensions to the Core Proposal</i>	<i>p.307</i>
<u>Section 2: Introducing Mind-Fusion</u>	<i>p.310</i>
<i>Subsection 2.1: Mars Needs Fused Humans</i>	<i>p.311</i>
<i>Subsection 2.2: Four Possible Observable Outcomes</i>	<i>p.313</i>
<i>Subsection 2.3: Is it Fusion?</i>	<i>p.315</i>
<i>Subsection 2.4: Rival Descriptions of Fusion Cases</i>	<i>p.317</i>
<i>Subsection 2.5: The Mereological Approach to Fission and Fusion</i>	<i>p.319</i>
<u>Section 3: Fusion from the Perspective of the Parts</u>	<i>p.321</i>
<i>Subsection 3.1: The Early Stages</i>	<i>p.322</i>
<i>Subsection 3.2: Is this Electrical Telepathy?</i>	<i>p.324</i>
<i>Subsection 3.3: Achieving Conscious Unity</i>	<i>p.326</i>
<i>Subsection 3.4: Responsibility and Soul-Searching</i>	<i>p.329</i>
<i>Subsection 3.5: Do the Participants Survive?</i>	<i>p.333</i>
<u>Section 4: Fusion from the Perspective of the Whole</u>	<i>p.335</i>
<i>Subsection 4.1: The Initial Pair</i>	<i>p.336</i>
<i>Subsection 4.2: Unity and the Coalescence of the Phenomenal Field</i>	<i>p.339</i>
<i>Subsection 4.3: Attention, Introspection and Self-Consciousness</i>	<i>p.340</i>
<i>Subsection 4.4: Blending and Blurring, Redux</i>	<i>p.342</i>
<i>Concluding Remarks</i>	<i>p.344</i>

In this chapter I first summarise the combinationist proposals I have made in previous chapters, reviewing their strengths, weakness, versions, and alternatives. I then illustrate my version of combinationism by applying it to a thought-experiment involving the fusion of two persons into one – or, in combinationist terms, two persons becoming conscious parts of a composite person. Over the course of this thought experiment, subjects like us become subjects like our parts, and I believe it is one of the advantages of combinationism that can make sense of this as a gradual change of status rather than an abrupt transition from only two subjects existing to only one subject existing.

Section 1: Recapitulation

My guiding thought in this work was that there was a widespread idea that minds can't compose other minds, and that this idea was constraining our philosophical thought without having been conclusively

demonstrated. In chapter 1 I defined a number of versions of this idea: Experiential Simplicity, Anti-Nesting, and Anti-Combination. I then tried to show that these theses, and in particular Anti-Combination, the claim that a whole's consciousness cannot be fully explained by that of its parts, mattered to several debates in the metaphysics of mind, and in chapter 2 I looked at what it would take to reject Anti-Combination (and *a fortiori* Experiential Simplicity and Anti-Nesting), and to endorse its negation, a view I called 'experiential combinationism'. In the following five chapters I examined the challenges facing combinationism, and the resources available to combinationists for responding to them.

I deliberately left open several questions about the nature of consciousness, composition, and explanation, yielding a range of types of combinationism. Combinationists can be physicalists or primitivists about consciousness, panpsychists or non-panpsychists, and universalists, nihilists, or something in between about composition. They can regard subjects of experience as constituted by sets of experiences, or as ontologically prior material substances, and they can affirm or deny the possibility of 'phenomenal overflow', conscious experience without cognitive access. They may regard wholes as grounded in, grounding, or identical with their many parts, and they may regard the explanation of wholes by parts, in the experiential case and elsewhere, as an *a priori* metaphysical necessity or as an *a posteriori* law of nature. All I have presupposed is that consciousness is an aspect of the natural world, and that within the natural world facts about wholes are fully explained by facts about their parts and the relations among them.

I cautiously believe that combinationism is true, but my aim in this work has been simply to show that it is defensible. Partly this has involved responding to arguments against combinationism, but since one of these arguments turned on the apparent lack of entailments from mental parts to mental

wholes, I have also had to suggest some positive theses. Here I review the various claims that I have defended as elements of an overarching combinationist theory.

Subsection 1.1: The Core Proposal

I will refer to the set of theses which I think provide the best prospects for consistent, intelligible, mental combination as my ‘core proposal’. It combines the inheritance proposal of chapter 3 with the adumbration proposal of chapter 4 and the refracted aspects proposal of chapter 5. Its three main components are TSE, CEI, and UA:

Token-Sharing of Experiences (TSE): Particular experiences can belong simultaneously to two different entities.

Conditional Experience Inheritance (CEI): A whole has an experiential property whenever one of its parts does and that part is appropriately related to its other parts, simply in virtue of the part having that experiential property and being appropriately related to the other parts.¹

Unification-by-Adumbration (UA): A subject has a unified experience of X-and-Y-together whenever they experience X, and experience Y, and these experiences are A-related.

From these, we can derive UCAP:

Unified Composites from A-Related Parts (UCAP): A composite subject enjoys conscious unity whenever its component subjects are A-Related.

These theses together provide an explanation of composite subjects with unified consciousness, based in the presence of experience in their parts, and the causal, representational, and phenomenal relations among the experiences of those parts.

¹ Note that on one view, the appropriate relations are just conscious unity, and so UA and CEI to some extent collapse into one another.

Many of these principles contain gaps whose filling-in depends on background theory: this makes the core proposal compatible with any of the versions of combinationism noted above. CEI speaks of being ‘appropriately related’, meaning ‘related in those ways which are conceptually required to ascribe ownership’. Many authors have argued that it would be a conceptual mistake to ascribe disunified experiences to the same subject, or to ascribe experiences to a subject that do not govern it causally in the right ways: a combinationist’s prior views about these requirements will determine their conception of what it is to be ‘appropriately related’.

The ‘A-relation’ was likewise defined conditionally, but always involves ‘perceptual adumbration’, the relation between what is given in an experience and what is indicated as not-given but present through its connection with what is given. If consciousness is transparent, so that its content is its only phenomenologically manifest feature, then the A-relation is the OA-relation, the relation between experiences when their contents bring each other to veridically adumbrate each other. If consciousness is not transparent, then the A-relation is the EA-relation, the relation between experiences when they bring *each other* to veridically adumbrate each other.

Whether the A-relation, and the unity it is meant to explain, is between experiences or between their contents, it provides a way to make phenomenological sense of unity between the experiences of distinct subjects: component subjects in a unified composite are aware of their experiences (or their contents) as revealed aspects of a more extensive whole, whose other aspects are revealed only through the other experiences theirs is unified with. By taking different views on the ‘minimal A-relation’, involving entirely uninformative adumbration, combinationists can reconcile the adumbration proposal with either view on whether or not phenomenal unity is dissociable from other forms of unity, whether or not it is transitive, and whether or not it is metaphysically fundamental.

In chapters 3 and 4 I also defended several supplementary theses:

Weak Exclusivity (WE): A single experience cannot belong to multiple discrete subjects.

Weak Privacy: A single experience cannot be directly known by multiple discrete subjects.

Complex Contents from Informative Adumbration (CCIA): A subject's experiences jointly represent complex contents whenever, and in proportion as, they informatively adumbrate each other's contents.

Agential Integration from Salient Concealed and Non-Salient Concealment (SCNC): A subject's experiences are disposed to guide its behaviour in an integrated way whenever, and in proportion as, they adumbrate each other's contents as highly salient, while finding no salience in their concealment.

The first two make the intuitive ideas of exclusivity and privacy compatible with the token-sharing of experience, while the latter two employ the notion of adumbration to explain richer forms of experiential structure than just phenomenal unity.

Finally, in chapter 5 I added the 'refracted aspects' proposal, a claim regarding the structural organisation of consciousness at each mereological level. On this proposal, a given subject's set of unified experiences is experienced by them as organised according to the different ways that subject can direct their attention, so that experiences of different components that are coupled for attentional purposes will appear together, while experiences of the same component that are attentionally dissociated will appear apart.

Despite their technicality, the inheritance proposal and the adumbration proposal are meant to capture two fairly simple ideas. The inheritance proposal is meant to capture the idea that if wholes are nothing over and above their parts, then wholes having experiences is nothing over and above their parts having experiences (and conversely if parts are nothing under and below their wholes). This is made plausible in particular by what I called 'the easy case', where a highly-integrated composite has one conscious part which controls its overall behaviour: here it seems natural to ascribe consciousness

to the composite. The adumbration proposal is meant to capture the idea that the complex structure of our consciousness consists in certain relations among experiences. This second idea can develop in many directions, corresponding to the sorts of conscious structure which one might wish to explain, and I have not attempted to vindicate every aspect of it. Rather, I have focused on those features which critics of combinationism have held to be especially resistant to compositional treatment, chief among them the unity of consciousness.

Subsection 1.2: Objections to the Core Proposal

Explicit rejections of experiential combination are much more common than explicit endorsements, and implicit rejections are more common still. So we must consider reasons why someone might reject combinationism, to see whether these reasons apply to my core proposal.

Two particularly pressing objections are the incompatible characters argument and the boundary argument. The first claims that a composite subject could not have a unified consciousness, for each experience it got from one of its parts would have to display two different phenomenal characters, one reflecting its unification with the rest of the whole's experiences, the other reflecting its unification only with the other experiences of that one part; the second claims that a composite subject having unified consciousness would destroy the distinct consciousnesses of its parts, by depriving them of their essentially 'bounded' character.

I think that my core proposal resists both of these arguments, by analysing conscious unity as a relation by which one experience alters the phenomenal character of another, without thereby becoming shared with the other's subject. Thus the parts of a unified composite mind enjoy an experience that is

containment-bounded - there are things it does not include - but which also reflects the other experiences it is unified with, so as to have the same phenomenal character for both whole and part.

But there is also the opposite objection, that consciousness in something's parts is compatible with any sort of consciousness, or a lack thereof, in the whole. To address this 'explanatory gap', we must ask what what explains the principles of the core proposal. In particular, can we 'see why' TSE, CEI, and UA are true?

TSE is a claim of metaphysical possibility, and so can be supported just by showing flaws in arguments against it; UA is a claim of phenomenological equivalence, and so can be supported by phenomenological reflection. But CEI is harder, since it appears to assert a necessary connection between two distinct states of affairs. The core of the explanatory gap that critics allege faces combinationism lies here, in the transition from the consciousness of parts to the consciousness of wholes. The possibility of justifying it depends on what prior commitments a combinationist has: if they accept an *a posteriori* account of reductive explanation, on which facts about wholes can be explained by facts about their parts even without being deducible from them *a priori*, then CEI could be taken as a metaphysically contingent law of nature. Combinationism would still be distinct from Strong Emergentism because its composition principles were simple and general in application, or because they involved shared token properties, or because the properties they ascribed to whole and part were conceptually akin. Since CEI meets all these criteria, it has an excellent claim to being a non-emergent *a posteriori* principle of composition, if such things are possible.

If reductive explanation requires *a priori* deduction, however, then the status of CEI depends on whether 'experiential ownership' - the relation a subject bears to its experiences - is intelligibly reducible to other relations. Since a physicalist combinationist, or an experience-first primitivist combinationist, is likely to think it *is*, these combinationists can derive CEI, given TSE: for all that is to

be explained is that experiential ownership obtains between the composite and the experiences of its parts, and whatever underlying facts constitute ownership-facts can be built into the specification of CEI. But the subject-first primitivist *a priori* combinationist posits something irreducibly experiential in the ownership relation, and so to explain CEI they must appeal to another principle, BEI:

Basic-Experience Inheritance (BEI): A whole has a certain basic-experiential property whenever one of its parts does, simply in virtue of the part having that property.

BEI implies that a human being undergoes all the experiences of their parts, and that any whole that includes a human being undergoes that human being's experiences, albeit only in the 'basic' sense, stripped of any implication of unity and causal integration. Insofar as basic-experiential properties still involve something phenomenological, this commits the subject-first primitivist *a priori* combinationist to being an 'inclusionary' combinationist, accepting phenomenal overflow and as many mega-subjects as there are wholes containing conscious parts.

And why is BEI true? BEI follows from three premises: that basic-experiential properties are fundamental properties, that either wholes or their parts are entirely grounded in the other, and the principle HF:

Heritability of Fundamental Properties (HF): *If* a property is fundamental, and *if* one of the level-connecting views of composition is true, and *if* the heritability of that property would not yield demonstrable incoherence, *then* that property is both-ways inherited.

I argued in section 5 of chapter 3 that it is plausible that HF is an *a priori* truth, following necessarily from a proper understanding of 'fundamental' and 'composition'. Since combinationists will accept a level-connecting view of composition, and since it is only combinationists who regard basic-experiential properties as fundamental who need BEI to be *a priori*, a defence of HF suffices as a defence of BEI, and from there conceptual analysis of ownership yields CEI.

So the core proposal can close combinationism's explanatory gap in a few different ways, given the right premises. I have tried to present those necessary premises as convincingly as possible, but I do not expect everyone to accept them, and I am not absolutely confident of them myself. But I think they suffice to make combinationism a live option.

Despite my efforts to be ecumenical, I doubt that I have exhausted the logical space available for combinationism. A combinationist might reject all versions of my core proposal. For one thing, they might abandon any sort of an inheritance principle, instead explaining each mental property of a whole by a particular configuration of different mental properties instantiated by components. This allows for the retention of Strong Exclusivity and Strong Privacy, and may avoid some of the surprising implications of BEI. But to be explanatory, this type of combinationism would probably require specific and powerful analyses of mental states – i.e. a definite theory of what mental states we have, what their natures are, and how each can be constituted out of other mental states. This goes against the methodological neutrality enabled by my use of general principles like CEI.

Alternatively, one might look elsewhere for a bonding relation to generate conscious unity, rather than relying, as I have, on perceptual adumbration. I am not sure what other relation among experiences would constitute unity when seen from the whole's perspective, but there may be one.

If rival versions of combinationism are developed by others, along the above lines or along lines I have not anticipated, they may well be inspired less by the general idea of combinationism itself, and more by some specific idea about the mind's nature that has combinationism as an implication. These versions will still face the challenges that I have discussed and suggested responses to, and I hope that anyone developing such a view, even if they do not adopt any of the theses I have recommended, will find helpful my efforts to organise, clarify, and canvas solutions to the various conceptual difficulties involved.

Subsection 1.3: Extensions to the Core Proposal

The core proposal is meant to make sense of how some unspecified whole's consciousness might be explained by the consciousness of its parts, without reference to any particular sort of parts or wholes. But there are specific challenges that arise when we extend it to certain sorts of subjects, most notably to the microsubjects that might compose us, the megasubjects that we might compose, and any large parts of us that are sophisticated enough to entertain self-referential thoughts. In chapters 5, 6, and 7, I outlined some additional theses that combinationists might use to tame these problems.

Microsubjects raise the problems of blending and blurring - how a qualitatively homogeneous but structurally fine-grained mass of experiences can explain the qualitatively diverse but structurally coarse-grained consciousness we have. They also intensify the problem of structural mismatch through their role in the 'dancing qualia' problem, though I believe the refracted aspects proposal noted in the last section resolves that problem. To address the problems of blending and blurring, I offered the 'confused blends' proposal: that for straightforward physical reasons, microexperiences will be radically confused with one another, and under those circumstances they will compose an experience whose structure is not apparent to its subject but whose quality is a blend of its parts' qualities. The idea of phenomenal blending - based on familiar examples like the presence of redness in orange and pink - is meant to provide an intelligible way for the microexperiences to make a difference to the whole's experience, even to be present in it (on the inclusionary version of the proposal), without appearing as distinguishable elements.

The problem around megasubjects is simply that their existence seems implausible, yet some forms of combinationism seem to imply it. In particular, *a priori* combinationists who regard basic-experiential properties as fundamental must endorse BEI, which implies that any composite

containing a human being basically-has that human's experience. Moreover, combinationists who regard phenomenal unity as a fundamental relation, and who value simplicity and generality of laws, are under pressure from an anti-vagueness argument to accept the following claim:

Unrestricted Unity (UU): All experiences in the universe are phenomenally unified with each other, i.e. form a single phenomenal field.

However, I argued in section 6 that the forms in which combinationists may be committed to these views do not support the implications that make them seem implausible. Those who accept BEI should regard basic-experiential properties as dissociable from the forms of unity that would produce consistent, co-ordinated, rational overall behaviour, and those who treat phenomenal unity as fundamental should similarly regard it as distinct and dissociable from the forms of unity that would give each component a salient, informative, awareness of the experiences of the others.

Finally, component subjects capable of self-referential thought raise the problems of Oblivious Parts, Self-Reference, and Self-Identification. For each such subject we seem to face a conflict between their self-knowledge and their integration into a whole capable of self-knowledge. In chapter 7 I offered the 'anti-individualist' proposal, which combines the following claims:

- A subjects' impressions of whether an event occurs in their own minds, or by their own will, are not based on whether it really does, but on whether it harmonises with that subjects' psychology;
- First-person terms in their primary use have a dual reference, picking out both whichever subject we consider them relative to, and also whichever of the subjects using them is largest, most integrated, or otherwise privileged;
- We cannot self-identify relative to some of the large complex parts of our brain, and rationally should be agnostic about our own identity;
- There is no genuine knowledge that a subject agnostic about their own identity lacks, but only a certain sort of action-guidance, and this guidance is equally well provided by knowing which complex of harmoniously-interacting parts we belong to.

While I think the core proposal does a good job of addressing claims of explanatory gaps and latent incoherence, I am much less confident about the success of these extensions. One worry is that we may have to accept a range of very odd minds whose existence is unexpected and whose structure is unfamiliar. But we should have no problem accepting the existence of a wealth of unfamiliar minds, if it allows us to see our own specific type of mind as a natural product of a systematic and intelligible material universe. We could reasonably think that our experience acquaints us with both a fundamental intrinsic property – phenomenality – and a particularly human sort of structure, but that these two need not always go together.

What is most troubling about applying combinationism to our parts is that it forces us to lower our estimate of how much a mind must know about itself. According to inclusionary versions of the confused blends and refracted aspects proposals, there is much in our consciousness that we have not recognised and cannot report, both the wealth of distinct but radically confused phenomenal particulars that blend into each ‘pixel’ of our phenomenal field, and the reaches of causally isolated experiences that lie so far from the centre of this field that they are realistically unattendable. This will be a hard pill to swallow for anyone who thinks that ‘our consciousness’ just means ‘what we are conscious of’, and thus ‘what we have cognitive access to’: but then, that idea is the hallmark of exclusionary combinationism, which avoids these implications at the cost of losing the option of grounding CEI in BEI.

Perhaps more seriously, according to the anti-individualist proposal of chapter 7, the large parts of a human being are systematically blind to their status as parts, and as a result the human being is unable to ascertain whether they are the whole or just a large part. This may even involve a systematic sort of ‘introspective indiscernibility’ between having a certain experience and informatively adumbrating that experience when another part of the whole has it. This uncertainty is troubling, and

both inclusionary and exclusionary combinationists are pushed towards it if they allow for large sophisticated parts of a human being to be conscious in their own right.

Note, however, that this sort of substantial fallibilism is compatible with maintaining that consciousness is inherently self-knowing, that it essentially incorporates a deep sort of acquaintance with its own nature. It is just that this direct insight need not always enable attention, imagination, recognition, and conceptualisation, need not always be rightly interpreted, and need not always allow for self-identification. We already accept this, insofar as we accept the possibility of consciousness in creatures like worms and fish, who cannot think about their experience the way we do. Combinationism simply suggests that forms of this dissociation of insight from conceptualisation are present in human minds.

These implications will probably be enough to drive many people away from combinationism. And readers already firmly convinced of Anti-Combinationism may find many points in this work to hold up as *reductios* of the whole idea, or even as reflecting an unlivable and disturbing world in which individuals have vanished, being either ‘swallowed up’ by a greater whole or ‘torn asunder’ into microscopic parts. But I hope that at least some readers will find the costs of one or another form of combinationism bearable, and that those convinced of its falsity will find my attempt to defend a systematic version useful in articulating their disagreement.

Section 2: Introducing Mind-Fusion

The best way to get a sense of what it is like to be a component subject in a unified mind is to imagine becoming one, and so most of this chapter is devoted to a thought-experiment involving the fusion of two minds into one. My primary aim is to make vivid the abstract principles defended in previous

chapters; my secondary aim is to argue that combinationism expands our options for understanding survival and identity in cases of fission and fusion.

In subsection 2.1, I set up the thought-experiment: two human brains are implanted with communicating radio devices until they are as integrated as two parts of one brain. In subsection 2.2 I distinguish some of the observable outcomes which might result from this set-up, labelling them ‘merging’, ‘domination’, ‘dissociation’, and ‘dissolution’. In subsection 2.3 I argue that (given the ‘merging’ outcome) the thought-experiment qualifies as a ‘fusion case’ in the sense discussed in the literature on personal identity, relative to almost all theories of personal identity. In subsection 2.4 I review the rival analyses provided by that literature, focusing in subsection 2.5 on a largely neglected approach which I believe combinationism makes more attractive.

Subsection 2.1: Mars Needs Fused Humans

Suppose that some technologically-advanced Martians decide, for reasons best known to themselves, that they would like to be able to combine many humans into a sort of ‘hive mind’ creature, with a single consciousness controlling many bodies and drawing upon the cognitive resources of many brains. Suppose that a team of their best scientists gets funding for a series of experiments aimed at creating such a ‘multi-human’.

They do this by implanting specially designed pairs of electrodes into the brains of two human participants, such that when a certain circuit in one brain fires, it activates an electrode, which sends a radio signal to the other, which immediately generates a surge of current in the other brain. This allows activity in one brain to produce or influence activity in the other brain, and hence allows the thoughts and experiences of one subject to produce or influence thoughts and experiences in the other. If the

number and sophistication of these electrodes were increased enough, the causal connections between the two brains could eventually be as fast, as complex, and as reliable as the connections between the two hemispheres of either brain individually. At this point, if we did not distinguish a signal going down an axon from one sent *via* radio, we could meaningfully say there is a single, highly-integrated nervous organ, with parts in two different skulls.

Of course this description simplifies what will be an incredibly fraught and complicated process. The Martians are ahead of us technologically by two or three hundred years, but not by thousands, so while they can create and implant these electrodes, they have limited foresight about, and control of, what happens next. To prevent messy failures that only deplete their supply of humans, they run the procedure carefully and gradually, over a period of months or years, so that increases in the bandwidth of the implants are interspersed with periods of exploration, assimilation, and adjustment by the human participants, who can spend time working out what signals they can now send and receive and how to respond to these developments.

Moreover, let us suppose that the experimenters design the implants to mimic nerve cells as closely as possible. Obviously there are limits, since nerve cells do not emit or receive radio signals, but they may incorporate things like neurotransmitters, axons, ion channels, and so on. In particular, the manner in which the implants ‘multiply’, spreading to connect more and more circuits across the two brains, could be made responsive to the way in which the participants use it. Just as neuronal connections become stronger or weaker based on the history of their activity, so can the links between implants, and between each implant and its brain. This way the experimenters need not constantly be performing repeat surgeries: the first operation puts in a biomechanical device which thereafter grows and expands into the brain.

Subsection 2.2: Four Possible Observable Outcomes

So far I have described the various hi-tech interventions which our Martian scientists are making into the brains of their human subjects. The goal of these interventions is to create a single being, with a unified mind and coordinated, intelligent behaviour across both bodies, which combines the personality, memories, and values of both humans. But their success in this endeavour is not guaranteed. While I will focus on the ‘good cases’ in which they succeed, other results are possible, and the outcome is determined not simply by the set-up and techniques used, but also by the way the humans handle the process. They have to undergo a transformative, potentially traumatic experience, and their temperament, and attitudes to each other, will make the difference at each stage between experiencing it as communion or as invasion.

We may distinguish four ‘ideal types’ of outcome, allowing that actual outcomes may be intermediate between them, or entirely unexpected. The intended outcome can be called ‘merging’; this is when there is a unified mind that controls both bodies and is recognisably continuous with both original people. The combinationist can still say that this is actually one of *three* minds, since the originals can survive as parts of it. But they display no more independent thought or sense of individuality than the conscious parts of a human brain.

The second-best outcome for the Martians is to have a single mind controlling both bodies but displaying continuity with only one of the original two minds. In such a case, we must surmise that the other mind has been suppressed, assimilated, or somehow subsumed into the resultant being without enough of its personality being at all manifest in that being’s behaviour. Call this outcome ‘domination’. For instance, one of the two participants might be aggressive, defensive, and unwilling to allow another access to its thoughts, while the other is submissive, deferential, and values acceptance over autonomy. The development of the experiment might then involve the former constantly seeking

to interfere more and more with the latter's mental processes, while resisting any countervailing interference. By the end, one human has in effect 'colonised' and 'assimilated' the other's brain into itself, and thereby taken control of their body.

Third, we might end up with two recognisably separate minds, controlling different bodies or alternating in control of both bodies, despite the organic connectedness of their brains. This would be somewhat similar to a case of dissociative identity disorder: two psychologically distinct but internally integrated personalities control (simultaneously or sequentially) a single organic structure. These two would probably be recognisable as the original people, who had built up psychological barriers to replace the physiological ones that had previously separated them; a combinationist might still think there are, strictly, three minds here, but the composite mind they formed would be closer to familiar social mega-subjects like a club than to an ordinary human mind. Call this outcome 'dissociation'.

This might arise if both participants were very concerned to maintain the privacy of their own mental processes, but had little desire to explore or enter into those of the other. At each stage of the experiment, they might respond to the new way of influencing each other's minds by setting up, independently or cooperatively, policies and habits to minimise its effect.

Finally, the process might be too traumatic and too invasive for either subject to survive. They might both end up so deeply psychotic and fragmented that the resultant being is not recognisable as either, and perhaps not even as a single individual. This might mean that neither body's behaviour was coherently interpretable at all, or even that at a certain point both bodies collapse into catatonia or epilepsy, having somehow killed each other from within, and never wake up. One or both might even become traumatised to the point of violent paranoia, seeking out the other's physical body and stabbing or strangling it in order to silence the voices in their head (if there is a composite mind here, it is most similar to aggregative mega-subjects like 'all the snakes in Ontario'). Call this outcome 'dissolution'.

So to get the philosophically most interesting ‘merging’ result, we might need to run the experiment several times. But we may suppose that the Martians experimenters have as much concern for human life as human experimenters have for the life of rodents. My reason for sketching these four possible outcomes is that it will be illuminating to refer to them at various points, noting how the way the participants handle a particular aspect of the process might make one or the other outcome more likely. While I will focus on the responses which most conduce to the merging outcome, these will be best appreciated by contrasting them with those which conduce to domination, dissociation, or dissolution.

Subsection 2.3: Is it Fusion?

The next three subsections relate this thought experiment, and combinationism, to the literature on personal identity over time, a diachronic question which I have so far largely avoided. Readers interested only in the application of a strictly synchronic account of mental combination to the above thought experiment may skip them without much loss.

Is my imagined procedure enough to fuse two people into one? That is, does it meet the following three conditions:

- a) At the beginning, there are (at least) two people, and
- b) At the end, there is (at least) one person, such that
- c) The relations between each original person and the resultant person are sufficient, considered in themselves, to count each pair as identical, i.e. as the same person at two different times.

Assuming we grant a), whether b) and c) are satisfied depends on our theory of personal identity, and most popular theories of personal identity make my thought-experiment satisfy b) and c); the exception

is the theory that our identity depends on an immaterial substance, a view which conflicts with the weak metaphysical naturalism I have been assuming.

On one view, a person survives as long as their distinctive psychology, or something organically developed out of it, is realised (Shoemaker 1984, Noonan 2003). On another view, what matters is not psychological continuity but continuity of the *substrate* of psychology, whichever organ underlies the person's psychology (Unger 1990, Nagel 1986). This means a person survives as long as their brain survives and remains functional, even if it changes radically in what sort of psychology it supports. Another view focuses not on *psychology* but on *phenomenology*, defining the persistence of persons in terms of phenomenally-continuous streams of consciousness, or in terms of the substrate which enables them (Dainton 2008, Bayne 2010, Ch.12).

My thought-experiment would count as fusion according to any of these theories, if the 'merging' outcome is observed. The resultant being will be psychologically and phenomenally continuous with both original people, and the mechanisms that bring about this continuity have been designed to mimic those operative in a normal human brain. Merely being spatially scattered, or using electromagnetic waves rather than electrochemical depolarisations, should not preclude the formation of a psychologically and phenomenally unified mind.

A third view of personal identity identifies persons with animals, i.e. organisms with biological, not psychological or phenomenal, persistence conditions (Van Inwagen 1990, Olson 1997). Merely fusing the minds is then not enough to fuse the persons, if the organisms remain separate. However, most animalists say that the crucial thing for an animal's persistence is the 'control centre' of their self-regulation, which primarily means the brainstem, which coordinates breathing, heartbeat, digestion, sleep, and so on. And in my thought-experiment the connections can occur in both the cerebrum and the brainstem, so that the participants might end up with the sleeping and waking, eating

and breathing, hormones and heartbeat, of both bodies controlled by electrical patterns across both brainstems. They would then be a single organism by Olson's and Van Inwagen's standards.²

Subsection 2.4: Rival Descriptions of Fusion Cases

Suppose that we do have a fusion case. How should we understand it – do the participants survive, and if so, as who? An initial issue concerns how to interpret the plausible claims that there are at first two people, and by the end one person (and vice versa in fission cases). A combinationist will not say that there are *only* two subjects at the beginning, or *only* one at the end, because they allow that within each human subject there may be many other component subjects. They can however accommodate the intuitive counts, if they count by partial identity rather than by identity (cf. Appendix, subsection 6.2). The initial situation features two subjects with whom all the other subjects present overlap significantly, and the final situation features one subject with whom all the other subjects present overlap significantly.³

Assuming that there are, in some sense, two initial persons and one resultant person, how are these persons diachronically related? Here we are subject to three competing demands:

² Perhaps someone might maintain that a person must always be a single bodily structure, understood as requiring ligaments and membranes that literally tie all its parts together, or a single circulatory system that spreads nutrients to all its tissues, or something similar. Then my thought-experiment will not be a fusion case. But with the tools at our Martians' disposal, slowly and organically fusing two bodies can easily be done. Maybe the two bodies are attached to each other first, with mental fusion proceeding afterwards, so that they begin in the condition of conjoined twins. Or maybe both fusions run in parallel. Either way, no problem of principle arises.

³ The intuitive count of subjects might also be qualified by a 'four-dimensionalist', who distinguishes the person-stages which exist at a given moment from the temporally extended 'continuant persons' that they compose. On the classic four-dimensionalist approach to fission and fusion, we are right to say that there are at first two people, and then one, because we are counting not by identity but by coincidence-at-a-time. But if we count continuant persons by strict identity, then there are two persons all throughout the process – at the end they are coinciding, or 'sharing a person-stage'. Like combinationists, four-dimensionalists accept the intuitive count of persons as only true when employing a counting relation other than strict numerical identity.

1. Recognise continuity: the processes described are constructive, not destructive, and while the people involved change, this change is gradual, organic, and often beneficial in terms of overall capacities.
2. Recognise transformation: a major change has been effected, and the final outcome is importantly different from the initial state of affairs.
3. Coherence: any description must fit into an well-motivated logical framework, allowing us to recognise both the continuity and the transformation in a single consistent language.

There is extensive discussion of what description best meets these demands (see, e.g. Wiggins 1956, Parfit 1971, Lewis 1976, Nozick 1981, Perry 1971, Noonan 2001). Saying that one of the original two subjects survives as the resultant subject while the other ceases to exist can accommodate the continuity in one subject's development and the transformation in the other's, but must choose arbitrarily which to regard as the survivor, and thus fails to meet the third demand. And saying that both original subjects survive, each individually identical with the resultant subject though distinct from the other, recognises the continuity of the process while yielding logical incoherence by making identity non-transitive.⁴ The most popular option, saying that both original people cease to exist and that the resultant person is newly created with the memories of having lived for many years, can accommodate the sense of transformation, and can be made consistent by including a 'non-branching clause' in one's criterion of identity, but struggles to capture the sense of continuity in the procedure.

To my mind, the right thing to say is that both original people survive, and that they become the resultant person not individually but collectively, by composing it. They both live on in this new whole, despite neither being on their own identical to it. This option is not often explicitly defended (though see Moyer 2008 for a four-dimensionalist version), but I think it is the most pre-theoretically appealing option for inanimate things, i.e. for paradigmatically divisible things like rocks or statues (I defend this

⁴ This approach only really makes sense when given a four-dimensionalist reading, in which the non-transitivity of tensed-identity is made compatible with the transitivity of strict identity.

claim more fully in the Appendix, section 5). One merit of combinationism is that it lets us extend this attractive description of the fission and fusion of inanimate things to that of persons, an extension I spell out in this chapter.

I do not claim that this approach is preferable to all others; rather than comparing it to its rivals, I wish merely to defend it from objections. My claim is that if combinationism is true, this approach to fission and fusion is viable. It might still not be true, and combinationists might prefer some other approach. Combinationism does not directly imply *any* thesis about diachronic persistence or change; it is entirely synchronic in content. Consequently, combinationism does not in itself address any of the technical problems arising from the persistence of composite entities, e.g. whether a thing which loses a part becomes identical to the complementary part, which it had previously been distinct from. The aim of combinationism is merely to allow parallel treatment of such problems for mental and non-mental composites.

Subsection 2.5: The Mereological Approach to Fission and Fusion

In what follows I will exhibit and defend what I call the ‘mereological approach’ to fission and fusion.

We can more precisely characterise the mereological approach in terms of the following two theses:

Part-Persistence: In fusion cases, each original person becomes a part of the resultant person, and in fission cases, each resultant person was previously a part of the resultant person, and these facts, perhaps along with other facts, explain the sense in which these persons survive these processes.

Pair-Persistence: In fusion cases, the original persons composed a pair (i.e. a mereological fusion of two persons) which became a person, and similarly in fission cases, the original person becomes a pair composed of the two resultant persons, and these facts, perhaps along with other facts, explain the sense in which these persons survive these processes.

The mereological approach could be glossed as the disjunction of these theses, allowing for three versions, which affirm one, the other, or both. Note that neither thesis claims that the parts of persons, or pairs of persons, *are themselves persons* at the times when they are parts, or pairs, of persons. The mereological approach is consistent with saying this, but is also consistent with saying that ‘person’ is not an essential kind, so that a conscious subject may be a person at one time, but then become a non-person (e.g. a pair of persons) while remaining a conscious subject. Most definitions of ‘person’ require a degree of *over-all* intelligence and coherence, which things like pairs of persons generally lack.⁵

I find the mereological approach attractive, because it seems right to say that in fusion two people become one composite person, and that in fission one person becomes two people, and that this description captures the core of why these are not cases of people dying or being destroyed, i.e. not failures to persist. In particular, I am attracted to the idea that each participant in my thought-experiment survives by becoming one part of the resultant person, remaining a continuously conscious subject but ceasing to be a maximal, autonomous, person.

Why might we reject the mereological approach? One obvious worry is that surviving as a part of a person, or a pair of persons, is simply not comparable to surviving as a person, and gives us no interesting sense of continuity in fission or fusion cases. But both Part-Persistence and Pair-Persistence include the phrase ‘along with other facts’. The mereological approach does not claim that *any* survival of one’s parts, or survival as a part of something, is equally good. Clearly, someone who undergoes

⁵ Parfit rejects the mereological approach to fission (more precisely Pair-Persistence) on the basis of wrongly assuming that it requires the pair to be itself a person. He writes:

“Suppose the resulting people fight a duel. Are there three people fighting, one on each side, and one on both? And suppose one of the bullets kills. Are there two acts, one murder and one suicide? How many people are left alive? One? Two? (We could hardly say, ‘One and a half.’)”(1971, p.7)

Since the mereological approach need never have more than two persons on the scene, the duel is straightforwardly between two persons. We need only add that there is a pair of people fighting a duel, which is entirely natural. If one of the bullets kills, there is straightforwardly one murder. There then remains only one person left alive, and that person is “all that remains of” the pair, and thus also “all that remains of” the original pre-fission person.

fission survives in a stronger sense than someone who is eaten and digested into their component molecules, whether or not we choose to say that the latter person ‘has become a scattered mass of molecules’. The difference is that in one case, but not the other, the person’s parts continue to display much of the sophisticated psychological and phenomenal life of the original person. They keep doing the things that originally made them the parts of a person, even though they are now persons in their own right.

This leads into the real objection: that parts of persons and pairs of persons *lack psychological and phenomenal properties*. If this claim were true, then it would be hard to see how these entities could be identical with persons at other times on account of psychological or phenomenal continuity. This is where I believe combinationism is relevant: if Anti-Combination is true, then ascribing psychological or phenomenal properties to parts or pairs of persons will be superfluous and prodigal. If Anti-Nesting or Simplicity were true, there would be conclusive reason not to. I think this is the major objection to the mereological approach, so in the coming sections I will address this objection, drawing on the arguments of the previous chapters. I will defend the ascription, to parts and pairs of persons, of psychological and phenomenal properties of the sort that could allow them to be continuous with persons. In doing so, I take myself to be defending the viability of the mereological approach to fission and fusion.

Section 3: Fusion from the Perspective of the Parts

In this section I discuss my thought-experiment of fusion with a focus on the two human individuals involved. My aim is to show that what happens to them can be understood as an extreme form of various familiar relational phenomena: they are connected with another person, who is at each stage a

strictly discrete entity and whom they at first perceive as ‘other’ but come eventually to perceive as an extension of themselves.

Subsection 3.1: The Early Stages

At first, the two participants will be related somewhat like two conversational partners, or two people with pagers. Each can, by thinking a certain way, produce a certain kind of experience in the other. Depending on where the implants are first put, this might be one seeing lights when the other thinks hard about math, or one hearing words whenever the other feels sad, or something else. Each participant may struggle, at first, to distinguish ‘normally occurring’ experiences from those produced by the other. But suppose we give them a supportive environment, where they can talk normally with each other (so as to ask “do you feel anything when I do *this*?”), and have time and inclination to practice controlling and interpreting the implants. This will probably let them devise a mutually-understood ‘language’, and come to perceive implant-generated experiences as signals betokening another mind, just as we perceive words, hand gestures, or facial expressions.

In learning this language, the participants would be employing the correlation or lack thereof between their own volitions, the reported volitions of the other, the experiences they undergo, and the reported experiences of the other (thereby conforming to the patterning principle of chapter 7, section 1). Assuming that each was honest and open, they could distinguish experiences arising spontaneously, experiences voluntarily produced by them, and experiences produced voluntarily or unintentionally by the other. But they could do this only because these various experiences were generated by ‘unsynchronised’ systems, which varied independently.

With time, they will also be able to reliably discern, from incoming signals, what effects they are producing in the other – paralleling the ability to, for instance, read in someone’s face how they feel about one’s utterances. Even before this they can guess, infer, or wonder about their effects on the other – including how the other is judging them based on ‘hearing their thoughts’. Each may thus become ‘self-conscious’ about whether the electronic relaying of their thoughts is impressing, amusing, disgusting, etc. the other. They may then make efforts to reduce or control the amount of information they send out, either by avoiding the thoughts which send out signals, by using feedback information to find ways of thinking those thoughts without being detected, or by learning which thoughts the other finds hardest to identify. Each might, that is, try to develop a ‘poker brain’ in the same way that we can develop a ‘poker face’.

Participants could also try to control the other’s knowledge of their mind by asking the other to deliberately ignore the signals they receive, to direct their attention away from them. If they trust each other, they may simply request this, much as we might request that someone look away while we change clothes or type a password. If these ways of politely ignoring the other, and the ways of establishing a ‘poker brain’, became habitual, they might eventually lead to the ‘dissociation’ outcome, with the individuals surviving as two separate minds in effectively one brain. But if they do not trust each other, then in order to ensure their privacy, each participant must ascertain whether the other is attending to the signals they are receiving *via* the implant. But then each must, to protect their own privacy, invade the other’s privacy. This lays the ground for a conflict which might end in ‘domination’, if one wins conclusively, in ‘dissociation’, if both manage to repel the other, or in ‘dissolution’, if each psychologically cripples the other.

Fortunately, the participants may display emotional responses besides defensiveness. In the right circumstances, humans strongly desire both to know others and to be known by them. So if the

right circumstances can be contrived, the participants may relish the connection which their implants give them, and spend much of their time engaged in silent but energetic conversation. No doubt they will also want a degree of privacy, so expedients like the ‘poker brain’ will be employed to some extent, but not so as to become automatic and inflexible. What will secure this happy result? Primarily it will be the temperaments of and relationships between the participants selected – perhaps for best results they should be a pair who have already established a strong and stable friendship or romantic partnership, who feel comfortable exposing their own mental lives and are enthusiastic about getting to see the other’s. Psychological health and stability will also be important, to handle productively the tensions and arguments which will inevitably arise when two people, even people who love each other, are installed permanently inside each other’s skulls.

Subsection 3.2: Is this Electrical Telepathy?

Suppose that the experimenters have, by luck, wise choice, or trial-and-error, selected two stable and mature human beings who are willing to be merged with each other but refuse to either conquer or be conquered. Is the communication they become capable of ‘telepathy’? Do they, at any point, have telepathic access to each other’s minds – and do the parts of our brains have telepathic access to each other’s minds?

If telepathy means direct, *unmediated* awareness of another’s experiences, then no: the access is mediated by the electrical connections (radio waves or nerve signals) which link the two subjects. This means that the access is fallible, for something may interfere with the signals travelling along those connections. In this sense, what the two participants have is no more telepathic than is an everyday spoken conversation – one subject’s thoughts are encoded in some form of energy signal that is picked up and decoded by the other.

Is the electrical communication perhaps somehow ‘less filtered’? It might be, especially at first. When we talk, we can exert a lot of control over the signal that gets to the other person, crafting it to be misleading or uninformative. But when the implants are first put in, there is none of that: each receives a signal that varies based simply on what the other is *actually* thinking or feeling, not on what they wish to claim they are thinking and feeling. But deliberate control, and with it insincerity and deception, can arise simply from each coming to understand how the connection works. So in a sense we might say that the participants have telepathic access at first (when they are probably unable to make sense of it) but then lose it as they become more aware of the implants’ workings.

In a third sense, telepathy would mean literally sharing experiences. Are there single token mental events which both subjects ‘have’? To some extent this will depend on what it is to ‘have’ an experience, the topic of chapter 3, section 4. In the sense of ‘control’, the two participants may share an experience, if the behaviour of both is guided by it. But this differs only in degree from the way that my decision might guide the behaviour of many other people, via my verbal instructions to them. The more important point is that two discrete subjects cannot both metaphysically underlie a single experience, since a single event cannot be occurring simultaneously in two substrates. Whatever is going on in one brain will be an experience of that subject, and not the other, and if an experience is realised in a dispersed manner, spread across both brains, then it belongs to neither individually, for neither is sufficient to underlie it: rather, that experience belongs to the composite of both, and each has some part of it, some component experience.⁶

So I conclude that in any strong sense, telepathic communication does not (and probably cannot) occur between discrete subjects. They can communicate in a fast, reliable, and informative

⁶ If an experience begins in one brain and then propagates into the other, we might say both subjects had it, but not simultaneously (cf. Unger 1990, pp.177-184). But insofar as we can divide events into temporal parts, we might prefer to say that the experience is composed of a sequence of experiences, each of which belongs to either the first brain, or the second, or to neither.

way, and they may in some cases do so with little room for insincerity or pretence. Yet they do not know each other's experiences *directly*, the way they know their own experiences, and they do not literally share experiences. This is a consequence of accepting (albeit in weak forms) the privacy and exclusivity of experience. If the two participants are discrete, Weak Privacy precludes them knowing the same experiences directly, and Weak Exclusivity precludes them sharing it.

Because the procedure involves no telepathy, we could not use it for direct inter-subject quality comparisons. Suppose we connect one normally sighted participant and one blind participant. Or suppose we connected two participants who were, unbeknownst to us, spectrum-inverted relative to each other, one seeing green whenever the other would see red, and so on.⁷ What would happen?

Maybe one participant would report, with surprise, having a kind of experience whose quality they could not previously have imagined. But this might also happen when two experientially-normal subjects are connected, if the signals each received stimulated their brain in previously unknown ways. Or maybe they would simply report experiencing familiar qualities under new circumstances or in new arrangements – because the incoming signal caused the same kind of brain activity as normal sensory stimuli. In either case, the type of quality which the receiving subject experiences supervenes on events going on in their brain, and thus depends on the experiential capacities of that subject and their brain. They may find previously-unknown capacities unlocked by the procedure, but this is still not the same as them directly sharing the experiences of the other participant.

Subsection 3.3: Achieving Conscious Unity

⁷ Spectrum inversion is often presented as two subjects being physically identical but phenomenally different, but even if this is conceptually and metaphysically possible I doubt that it is nomologically possible. I am not thinking of this kind of spectrum inversion, but rather of two people who differ physically in some subtle way which changes their phenomenology while giving little outward indication of doing so.

Our two participants will come to enjoy unified consciousnesses when they stand in the relation which constitutes unity. Here I assume the analysis of chapter 4, according to which this relation is the A-relation, defined in terms of adumbration, the phenomenological structure in which a subject is aware of something given as intimately connected with something not given.

How we think about the *progression* of A-relatedness in this procedure will depend on what we think is true at the start, when our two subjects are simply two normal humans, perhaps isolated and perhaps communicating normally. As noted in chapter 6, section 2, anyone who regards phenomenal unity as a fundamental relation will be under pressure from an ‘anti-vagueness’ argument to posit it running universally among all experiences. If we accepted this sort of universalism then we should say that right from the beginning, each participant’s consciousness involves a phenomenology of pure openness, and thereby referred to the experiences of the other. If we rejected Universalism, we would instead say that phenomenal unity appeared at some moment when the connections among their experiences were sufficiently informative, though it may be indeterminate at which precise moment they first qualified.

What about richer forms of unity? In chapter 4 I proposed that richer forms of unity involved mutual, informative, veridical, adumbration of each other by the parts’ experiences. Crucially, to qualify as conscious unity each part must be adumbratively aware, not of the other’s experiences as external events, but of how things seem to the other. That is, each participant must see the world through the other’s eyes, and have an awareness of this other perspective constantly in the background of their own perspective.

In our thought-experiment, the increasing richness of, and sensitivity to, the signals travelling between the two brains could easily generate increasingly informative adumbration of this sort. As the participants learn to tell what kind of experiences in the other mind produce what experiences in their

own, they will come to perceive the latter experiences as the revealed aspect of - the expression of - a mental process that is not fully given but whose content they increasingly can discern. By extension, the absence of certain experiences in their own minds, and the particular combinations of experiences they do have, will also come to seem meaningful, adumbrating a whole other stream of consciousness. Eventually, when each experience of either participant conveys something not just of *their* other experiences, but of the experiences of the other participant, we will be able to call the whole set unified because each experience will be felt as merely one fragment of a conscious whole that includes both minds.

Through this process, we will also find pairs of experiences in the two participants interacting to yield more complex experiences. For instance, when one of the participants perceives something, the signals received by the other may activate a memory of a similar thing perceived in the past, and the signals of this memory received by the first may then contextualise and colour their perception of this new thing just as their own memories would. This growing disposition to think together will likely be accompanied by a growing difficulty in thinking separately: when relevant thoughts from the other subject spring to both minds so quickly and readily, it will be hard for either not to be influenced by the other's on-going thoughts.

In chapter 6, subsection 2.4, I offered the analogy of 'internalisation': when someone has exercised a formative influence on us, we can come to involuntarily and unreflectively see the world as they would see it, with this seeing becoming a constant background to our own seeing. What stops that from counting as conscious unity is that it is in virtue of a past interaction, not a present one, which is precisely what the procedure we are imagining changes. Each participant's constant background awareness of the other's perspective is a product of an ongoing interaction - though this need not mean that it snaps into existence simply because the implants are turned on. If the flood of new stimuli is

unwelcome, it will prompt defensive measures, or cause a traumatic breakdown in one or both psyches. But even if it is welcomed, there will still have to be a process of attunement and mutual learning, to help the two minds communicate more effectively rather than to help one recreate the other in their absence.

Subsection 3.4: Responsibility and Soul-Searching

As the procedure progresses, the two participants are likely to spend less time attending specifically to one another, and more time attending jointly to external things. In the early ‘conversations’, each focuses on how to convey things to the other and on how to interpret the other’s own expressions. Their goals are to learn about the other, or to cause the other to believe certain things (true or false) about them. This contrasts with shared attention, where they are not attending to each other, but are rather aware of each other in the background of something they are both attending to, and which they are attending to partly because they know the other is attending to it.

To illuminate this shift, we might compare two people first meeting, engaged in ‘getting to know’ one another, with two old friends jointly considering a shared problem. In the latter case, each may ask the other for their view on a particular part of the question, and will maintain a constant awareness of what the other knows, may not know, can do, refuses to do, and so on, but only in the same way that they maintain a constant awareness of their own capabilities and knowledge, without having to focus attention on them (Cf. mindset pragmatics, Stalnaker 1970, Hellie forthcoming-a).

The growth of shared attention corresponds to one aspect of the growing dispositional unity between their consciousnesses: their experiences tend more and more to transfer attention, so that if one focuses on something, the other’s attention will also be drawn to that thing (even if that thing is known

to them only *via* signals from the other). This will contribute to a growing difficulty in assigning responsibility to one participant or the other. Each time one begins to focus on an action, the other becomes aware of this, and has a few related ideas, which the first immediately becomes aware of, and so on. Eventually every action performed by either body will be the product of multiple rounds of feedback between both minds, making it all but impossible to isolate the exact contribution of each.

Similar things will occur with rapid and habitual actions, because at some point it will be possible for one participant to initiate actions in the other's body, perhaps acting via a momentary urge it can prompt in the other's consciousness or perhaps bypassing the latter. Even if the second participant is able to inhibit the action just as they can inhibit their own, at this stage they will likely have become fairly comfortable with each other and not automatically prone to inhibiting every impulse they receive from the other. For instance, if one participant is graceful and perceptive and the other clumsy and oblivious, the first may start adjusting the second's posture to avoid the trips and spills they were previously prone to. The second, seeing no reason to block out this helpful intervention, allows it to become habitual, so that every action performed with the second body is a blend of contributions from both minds.

But even while it becomes harder to assign responsibility to just one participant, it also becomes less practically important to do so. Normally we care which of two people did something, because we need to know what to expect from those people in the future. But if neither participant will act separately in the future, this prospective need becomes less urgent. It is also normally important to assign responsibility so as to allocate rewards and punishments, but with the two participants so closely linked in body and mind, and likely strongly caring about and empathising with each other, any pain, joy, or inconvenience caused to one will affect both.

There will also be long-term effects on each participant's personality. Just as people often change to reflect and accommodate their families, the participants will be prone to absorb values, beliefs, and habits from each other – while any sharply opposing or incompatible traits are likely to be removed, either violently (in a case tending toward domination) or by persuasion and negotiation (in a case tending toward merging). Any persistence of sharp conflict increases the odds of dissociation or dissolution.

Drawing on my defence of the patterning principle in chapter 7, section 1, we might say that their background psychologies and background wills are more attuned to each other. If the participants perceive events as their own voluntary actions to the extent that those events harmonise with their psychology, then they will start experiencing each other's actions as their own. For instance, when I decide to picture a yellow flamingo, and the image of a yellow flamingo arises, the close correspondence between what I intended and what happens makes me feel that I imagined the flamingo voluntarily. Now if one participant decides to picture a yellow flamingo, which is noticed by the other, who is generally better at visual imagination, their habitual response might be form an image of a yellow flamingo and transmit it to the first participant. The first participant, experiencing an intention followed by a matching image, will likely feel as though they have voluntarily imagined that flamingo (cf. Perky 1910). The distinction between doing something oneself, and doing it with the other's help, becomes increasingly irrelevant as these options become indistinguishable in speed and reliability.

Arguments over responsibility may still arise, when something goes wrong and each tries to blame the other. But these are unlikely to end with any clear answer: rather, the participants need to be willing to let go of the question. Only if they can think of their actions as 'ours', rather than trying to carve out 'mine' and 'yours' in each case, will the procedure lead to successful merging.⁸ In the terms

⁸ Something similar to this is already found in the practice of social institutions which codify policies of collective responsibility, e.g. the British cabinet.

of chapter 4, the salience of concealment must be low, and each participant's adumbration of the other must be experienced as adequately informative, if they are not to perceive each other as threats. This is partly about *getting* more information through the implants, but also partly about *needing* less information, and being willing to trust the other even without exhaustively knowing every detail of their processing.

Can each participant not still consciously decide to 'take charge' of their own mind and make a decision that is their own, not shared? Supposing the other is supportive and does not deliberately interfere, isn't individual responsibility still available, by carefully 'screening out' incoming signals?

There are three difficulties facing such efforts at individual responsibility. First, as noted above, it will grow harder and harder to discern which thoughts come from outside and which from inside. Second, many internally-generated thoughts will now reflect the past influence of the other, and may even require consultation with the other in order to properly understand them (for instance one participant may have been persuaded of a certain belief by the other, but remember the conclusion better than they remember the arguments).

The third and most interesting difficulty is that autonomous decision-making requires distinguishing, not only between what is one's own and what is another's, but also between what is one's own and what is a random passing whim or chance thought. Occasionally I might get the urge to slap an annoying person, but it does not follow that if I gave into that urge every time it arose I would be acting more autonomously, for it may be that a concern for civility and mutual respect is a far more important part of 'who I really am' than this occasional urge.

Because acting autonomously means finding our real will, amongst a certain amount of random statistical 'noise', we need to engage in 'soul-searching'. By this I mean a kind of calm taking stock of one's thoughts and feelings, a conscious effort to be open to all our desires and to discern which are

stable and which are fleeting. This effort at ‘self-consultation’ contrasts with the sort of blinkered focusing where we just pursue to completion what we have embarked upon, pushing aside all new thoughts and feelings that emerge, attending only to the implications of a particular line of thought or the means to a particular end. While this makes us more likely to successfully complete our task, it impedes efforts to establish our true wishes, and increases the risk of devoting ourselves to something we do not really want.

So acting autonomously requires ‘soul-searching’, which requires openness to one’s ‘whole mind’, a lowering of thresholds for admission to attention. But in our thought-experiment, this kind of openness will actually invite thoughts stimulated (even if unintentionally) by the other. The attempt to block out or ignore thoughts coming from the other requires the opposite of soul-searching: vigilantly keeping watch over one’s thoughts and driving away any that do not fit certain criteria. Thus the distinction ‘talking to the other person vs. doing it myself’ comes increasingly to line up with the distinction ‘soul-searching vs. blinkered focusing’. We can find parallels to this in real life: talking with someone else is often the best way for us to work out what we really want, as long as they are supportive and open-minded. The participants in the thought-experiment just take this to an extreme.

Subsection 3.5: Do the Participants Survive?

I will end this section by defending one of the presuppositions of the mereological approach to fusion: that the two original participants survive as parts of a person, retaining many psychological and phenomenal characteristics. Many people would describe the merging outcome by saying that the two participants are no more; they no longer exist, having been ‘absorbed’ or ‘dissolved’ into the whole.

The mereological approach can accommodate this claim of the participants' non-existence in at least two superficial senses, but implies they are strictly false.

The first superficial sense is that the two participants have ceased to be salient things, or things which it is useful to think in terms of. It is no longer sensible for someone trying to understand the situation to organise their thoughts around 'participant 1' and 'participant 2'. The second superficial sense explains this fact; each participant has lost a significant degree of 'independence', in that their body's actions are no longer primarily controlled by their own mental processes, and their mental processes are no longer primarily controlled by their own mental processes a moment before. But despite these facts about the two participants, a defender of the mereological approach will maintain that they still, strictly speaking, exist. The success of fusion is not marked by a change in which things exist, but by a shift in the salient divisions, whereby the distinction between the two individuals becomes increasingly irrelevant. Each one's story is not a story of something being destroyed, but of something growing and forming new connections to an external thing – an external thing which, if the experiment succeeded, probably managed to elicit intense feelings of friendship, love, and acceptance.

Here are two ways to support this analysis. The first is to simply point out that in all fundamental intrinsic respects, the type of thing happening in each brain at the end is the same as what was happening at the beginning. Neurons fire, stimuli produce responses, information is processed and filtered and integrated to enable a coherent and meaningful worldview. Why think that these neural events no longer support a conscious state of their own?

One might say that they no longer support a conscious state of their own because they contribute to supporting a larger conscious state. But this follows only given the rejection of combinationism, which claims precisely that supporting a conscious state of their own can be a way of contributing to a larger conscious state. Or one might say that their psychological integrity has been

destroyed, and they no longer display the same psychology as before. But any theory of persistence must account for the way that our psychologies change over time, and one of my major aims in this section has been to show that what happens to these participants is just an extreme version of things that happen to all of us when we become intensely related to someone – shared attention, long-term shifts in personality, or finding that we best understand our own desires by discussing them with another.

Second, in addition to the above reasons for accepting psychological and phenomenal continuity between the original people and the eventual parts, we can support the mereological approach by an argument analogous to that discussed in subsection 3.3 for phenomenal unity. Either the participants survive, or there was a precise moment when they disappeared, or it is vague matter exactly they vanished. Since it seems impossible to find a non-arbitrary millisecond for each participant when their consciousness ‘went dark’, the mereological approach has the advantage of an appropriately gradual analysis, which need not impose a sharp boundary onto a smooth continuum.⁹

Section 4: Fusion from the Perspective of the Whole

In this section I shift to consider the whole composed of the two human participants. Its perspective is, at first, much less familiar to us, and harder to make sense of. Nevertheless, I argue, we have conceptual tools available that can give us some idea of what it is like to be a pair of people mid-way through this process, and perhaps even what it is like, if anything, to be a pair of people prior to this process – i.e. what it is like to be any of the many pairs of people which actually exist.

⁹ What about saying that the participants cease to exist, but it is vague exactly when? If this vagueness is metaphysical, then at certain points it will be metaphysically indeterminate whether a subject exists or not. But if this vagueness results from there being different admissible ranges of application for the participant’s names, then we must treat the existence of a conscious subject as reducible to some underlying dimension, and the defender of the mereological approach can claim that they have identified the relevant underlying dimension, namely the relative independence of the many overlapping subjects involved.

Subsection 4.1: The Initial Pair

Having defended the presuppositions of Part-Persistence at the end of the last section, I will begin this section with a defence of the presuppositions of Pair-Persistence, namely that pairs of persons exist and possess psychological and phenomenal properties.

First note that subject-first and experience-first combinationists will conceive of ‘pairs of persons’ differently, because they conceive of persons differently. For the former, a person is a physical organism (or organ), and so a pair of persons is just a pair of organisms (or organs) whose members happen to be intelligent and rational. For the latter, a person is constituted by a set of suitably related experiences, and so a pair of persons is an aggregate of experiences forming two suitably related clusters. On either conception, do such entities exist? Well, do pairs of things in general exist? We cannot simply appeal to ordinary language here, because the way that we talk about pairs can seem sometimes plural (‘that pair *are* really getting on my nerves’) and sometimes singular (‘there *is* a pair of lawyers outside’), and moreover because grammatical number is an unreliable guide to real singularity, at least in English (consider ‘a pair of pants’, or ‘they’ as neuter singular).

What we need to consider is whether pairs have a place in our best overall ontology, and here there is much dispute. One particular argument, analogous to the anti-vagueness arguments of subsections 3.3 and 3.5, presents unrestricted composition (which would imply that pairs exist) as the only way to avoid vagueness about existence. The thought-experiment of personal fusion makes this argument especially vivid, by exhibiting a gradual progression from the participants being merely a normal pair (and hence, we suppose for *reductio*, not composing anything) to them being highly integrated (and hence, we suppose, composing a whole). At what moment does there begin to be

something – how much integration does there need to be, precisely? Any precise answer will seem arbitrary, and so it seems there must have been something from the beginning.

Even if we do suppose that pairs exist, the question is whether they have any psychological or phenomenal properties, since the best way to make sense of a pair becoming the eventual person (or of a person becoming a pair, in the case of fission) is to say that there is continuous instantiation of such properties. Now, many mental properties seem bizarre to attribute to a pair of persons, once we are clear that this ascription is not to be merely distributive: obviously we can say ‘there’s a happy pair’, meaning that each member individually is happy, but can the pair be happy *itself*?

But recall chapter 3’s distinction between ‘full-ownership’ and ‘basic-ownership’ of experiences, the former defined as the relationship subjects have to ‘their’ experiences, and the latter defined as the additive elements of the former. In chapter 6, section 1, I argued that if we suppose basic-ownership is dissociable from full-ownership, we can endorse BEI while explaining away the appearance of non-mentality in large disunified composites. This same analysis can be used to explain how a pair of persons can have phenomenal properties: not all the same ones as its component persons, but the basic-experiential ones. The pair contains and underlies the same experiences as its parts – but they, and not it, also have their overall behaviour intelligently guided by those experiences, and (depending on the truth of UU) they and not it may have their experiences in a unified way.

Psychological properties are trickier, because their causal role in guiding intelligent behaviour seems so much more central to them. Here I think the defender of the mereological approach ought to adopt a slightly different strategy. In chapter 3 I observed that the causal element of ownership was a matter of degree: a given set of beliefs and desires, together with all associated phenomenology, may explain more or less of the events occurring in a certain entity. Even in the best cases (e.g. the mind controlling a human brain) there will be a residue of unexplained events, like the random circulation of

molecules through different tubes, or the object's slight deformation by the earth's gravity and rotation. Usually we only ascribe a mental state to an entity if that mental state, and others it operates together with, controls 'a great deal' of the events occurring in that entity; hence we can say that a human being believes a certain film is good, but not that the earth's biosphere believes that film is good, because that belief explains much of what occurs in the human but a miniscule fraction of what occurs in the biosphere.

But suppose we instead relativise ascriptions to a degree of control. That is, we might say that the above belief is instantiated to a high degree by the human being, and to a very low degree by the biosphere. This would let us say that the pair of people instantiates the psychology of each person about half as much as the person themselves. As the experiment progresses, the mental states of each participant acquire a greater and greater control over events happening in the other brain and body, and thereby come to be instantiated by the whole pair to a greater and greater degree.¹⁰

Thus we can say coherently that if the pair of persons exists at the beginning of the procedure, it has phenomenal and psychological properties – though it has the former in the basic rather than full sense, and has both only to a fairly low degree. Similarly, we can say that in fission, a person has continuity of psychology, though they end up instantiating psychological properties about half as strongly, and having experiences only in the basic sense. Whether this is enough to say that one entity has phenomenal and psychological continuity from the beginning to the end of the procedure will depend on exactly what sort of properties need to be retained, and to what degree. But there is at least

¹⁰ It is very plausible that the threshold for when psychological states control the whole pair's behaviour sufficiently to be ascribed to it by normal semantic standards coincide with the thresholds we normally employ for ascribing collective mental states to groups. These thresholds seem to involve, at a minimum, common knowledge among the group members, and mutual conditionality - each intends to do their part only if the other does their part, and knows that the other knows this. Cf. Searle 1990, Bratman 1999, Pettit & List 2011.

the possibility of finding some psychological and phenomenal criteria of persistence according to which a pair can persist as a person, or a person as a pair.

Subsection 4.2: Unity and the Coalescence of the Phenomenal Field

Even supposing that the pair exists, and has psychological and phenomenal properties, does it have unified consciousness – do all the experiences it basically-has form one phenomenal field, one ‘thing it is like’ to be that entity? On the account of unity given in chapter 4, this turns on whether its two sets of experiences are A-related.

In subsection 3.3 I argued that combinationists face a vagueness argument pressing them to be either reductivist or universalist about phenomenal unity, either analysing phenomenal unity as ‘a sufficient degree of dispositional unity’, where ‘sufficient’ is semantically or epistemically vague, or else positing it from the very beginning. The alternative is for phenomenal unity to be metaphysically vague, or for the transition to occur at some precise but seemingly arbitrary point.

Whenever the composite begins to enjoy phenomenal unity, it has a single phenomenal field, and we can describe the process from then on in terms of the changing configuration of items in this field. In chapters 4 and 5 I envisaged experiential elements arrayed in the phenomenal field according to their tendencies to interact, with greater proximity indicating greater sensitivity. On this picture, the composite’s field will have two large clusters of closely-unified experiences at some significant distance from each other, with a few elements in each close to elements in the other. These few

elements are the experiences which activate, or result from, the implants in their brains, and as the implants develop these points of proximity will extend, pulling the two clusters closer together.¹¹

Subsection 4.3: Attention, Introspection and Self-Consciousness

Talking about what it is like to be the composite will naturally make us wonder whether and how the composite can introspect. Similarly we might wonder when it can attend to things as a single whole, and when it can conceptualise its own experiences as its own.

In the early stages of the procedure, each person has good access to its own experiences, but quite limited access to those of the other brain. Each will thus form introspective impressions of only part of the composite's phenomenal field: or at least, introspective impressions in which the rest of the field features hazily and peripherally. Similarly, if attending to something is a matter of placing it at the centre of one's phenomenal field, then at first the composite cannot do this, for each person can only place an experience at the centre of one cluster or the other within the field. This is still a sort of 'introspection by the composite', or 'attention by the composite': the composite is introspecting on, or placing something at the centre of, half its experiences at a time, using a different brain each time. But this falls short of what we might think of as 'proper' introspection and attention, which involve the whole phenomenal field. These, I will argue, become possible for the composite only when it can act simultaneously with both brains, and connect its two acts properly.

Let us start with attention. If one participant's attending to something serves to draw the other's attention to it, and vice versa, then the central elements of each cluster have great causal proximity with

¹¹ It should be recalled that there are as many dimensions to this field as are needed to consistently represent the different distances among the elements, and it might be non-Euclidean in some respects. Consequently we should not try to too hard to visualise this image of two clusters growing closer and closer together.

each other, represented geometrically by their closeness to the other. By the two centres of the clusters moving together, they come to form an element which will be made central to the overall field. Thus the composite has, using both of its brains, placed something at the centre of its phenomenal field, i.e. attended to it.

For proper introspection, the composite needs to not only have each brain introspect but have each enquire with the other about the other's part of the field, and incorporate what they are told into their own view. It may never get a complete survey of the phenomenal field in a single brain, but this does not mean it never gets a complete survey: the complete survey is distributed across the two brains. This allows the composite to think about its own thoughts, but it does so by means of each participant separately identifying and then putting together 'what *I* am experiencing' and 'what *they* are experiencing'. These two impressions will be distinguished as introspective and testimonial, respectively, and so the composite can think of itself as such only on the basis of first thinking of each part as a distinct part.

But as the procedure progresses, the consequent 'we-thoughts' will follow the antecedent 'I-thoughts' automatically, as any attempt at introspective stock-taking by one participant automatically prompts the other not only to also take stock, but to listen to its partner and share its own results. With increasing automaticity, it will become less salient to each that it is hearing things from someone else. By contrast, the impressions in each brain of what 'we' are experiencing will grow more salient, not to mention easier to focus on, for segregating out only the experiences from one's own brain will come to take more and more effort. Eventually, for a participant to introspectively review all and only what they individually are experiencing, without attending also to what the other is experiencing, will come to seem both pointless and nearly impossible, compared to reviewing everything that in both minds. The

participants come to be conscious of themselves only as a whole, not as parts. And thereby the whole becomes conscious of itself as such.

Subsection 4.4: Blending and Blurring, Redux

While in one sense the composite has gained new capacities throughout the procedure, there is also a respect in which it has steadily lost capacities. In particular, it has lost various capacities to do one thing *without* a certain other thing happening. The implants ensure that certain mental or bodily actions taken with one body will have automatic consequences in the other body – such as the other participant knowing what was done, or feeling some emotion, or supplying some useful or disruptive feedback. It is often worthwhile and overwhelmingly tempting to prioritise speed and efficiency over carefully scrutinising every step in a mental process, and so when there is no strong reason to keep things separate, a useful pattern of action may become so habitual that its components cannot occur separately.

This loss of capacities is a form of *confusion*, as defined in chapter 5. The composite mind becomes unable to perform one mental act without simultaneously performing another, just as each participant's mind may be unable to attend to the microexperiences of its microscopic parts without simultaneously attending to many others. At first this confusion will be weak and shallow: shallow because the effect in the second mind might still potentially be inhibited or avoided, and weak because even once both mental actions have occurred, the separateness of the two brains' attentional systems will allow for distinct attention. But over time, as the links become stronger and faster, and as the two brains' capacities for attention become more and more coupled together, there may be instances where it becomes in practice impossible even to attend separately. Obviously this will not happen for all mental events – there are benefits to preserving some separation of function, if only so as to know at

each moment which body's eyes are generating which visual experiences. But pairs of events which there is no reason to keep distinguishing may become strongly, robustly, and symmetrically – i.e. radically – confused.

Since confusion is subject-relative, this does not guarantee that the parts suffer the same confusion. They may suffer confusion between a given thought or experience of theirs and the *feedback* that it automatically prompts from the other brain, but this is still confusion among events in their own brain, contrasting with confusion among events in the two brains.

This raises the interesting possibility of the whole experiencing phenomenal qualities which its parts cannot, blended out of, and thus nothing over and above, the qualities experienced by its parts. To see this, imagine that one subject is rather like a spectrum-invert relative to the other, outwardly the same but seeing different qualities - seeing gred, grue, and grurple where the other sees red, blue, and purple. For these to blend they must be both phenomenally unified and radically confused; this is unlikely to happen with perception (due to the separateness of eyes) so we should focus on imagination. Consider the whole visually imagining the colour blurple, a blend of gred and blue, by imagining gred in one brain, and blue in the other, without separating the two. What is this like for the parts, e.g. the subject who can see and imagine blue but not gred? Since their blue-experience is part of a complex content with the other's gred-experience, it adumbrates that other experience in some fairly informative way. Yet since the link is not telepathic, they do not experience and cannot even imagine gred, and by extension do not experience and cannot imagine blurple. Yet at the same time, this joint imagination is entirely satisfactory within the psychological economy of the merged minds, generating no sense of dissatisfaction or frustration in either part.

Thus the subject imagining blue has an experience which adumbrates gred informatively, without actually instantiating the character of gred, and yet generates no further curiosity or sense of

lacking access to the quality gred. This is a hard description to make sense of, but not impossible. I think the best way to make sense of how the adumbration can be highly informative without actually conveying the character of gred is to think of it as a sort of acquaintance that enables recognition but not recall; the subject will recognise next time the other subject is imagining gred, but cannot themselves visualise or otherwise capture the quality – as when we say “I don’t know how to describe it, but I would recognise it if I saw it again”. And the best way to make sense of its producing no sense of frustration or curiosity is to appeal to the Patterning Principle: the subject imagining blue thinks they can imagine gred, and indeed blurple, because whenever they try to (by resolving to imagine ‘that quality’) this prompts the other to imagine gred, thereby composing blurple, and all the feedback that any subject (composite or component) gets is ‘success: quality visualised’. In fact much of that feedback is coming to each component subject from the other, but the degree of volitional harmony they have built up makes this fact impossible to detect from the inside.

Concluding Remarks:

When I think about the procedure our two imaginary humans have undergone, I find combinationism more attractive. If minds cannot be parts of minds, then there must be some moment when two become one, and I find it hard to accept any such abrupt transition in a gradual process. Even if minds can be parts of minds, but component minds do not explain the mentality of their composite, then when we have explained the behaviour of the composite in terms of the two original minds closely and automatically co-operating (as I think we can), it would be superfluous to posit a further distinct mind belonging to the whole. Yet if the whole gives as much indication of being an intelligent conscious subject as any of us do, then it seems arbitrary to deny it that status. I find myself forced to look for some account on which the genuine consciousness of the parts and the genuine consciousness of the

whole are not only compatible but are two sides of the same coin. Over the course of this work I have attempted to find such an account, and to articulate both why that account is attractive and why it might be rejected.

Glossary of Abbreviations:

Absolute Phenomenal Context: The phenomenal context of an experience e = the set of experiences with which e is co-conscious.

Agential Integration from Salient Concealed and Non-Salient Concealment (SCNC): A subject's experiences are disposed to guide its behaviour in an integrated way whenever, and in proportion as, they adumbrate each other's contents as highly salient, while finding no salience in their concealment.

Anti-Combination: The experiential properties of a conscious subject cannot be explained by the experiential (and topic-neutral) properties of, and relations among, its parts.

Anti-Nesting: A conscious subject cannot have parts which are themselves conscious subjects.

Basic-Experience Inheritance (BEI): A whole has a certain basic-experiential property whenever one of its parts does, simply in virtue of the part having that property.

Complex Contents from Informative Adumbration (CCIA): A subject's experiences jointly represent complex contents whenever, and in proportion as, they informatively adumbrate each other's contents.

Composition Principle (schematic): If a set of parts p_1, p_2, p_3, \dots possess properties F_1, F_2, F_3, \dots and stand in relations R_1, R_2, R_3, \dots then they compose a composite c which possesses property G .

Conditional Experience Inheritance (CEI): A whole has an experiential property whenever one of its parts does and that part is appropriately related to its other parts, simply in virtue of the part having that experiential property and being appropriately related to the other parts.

(Conditional) Inheritance of x : A whole has property x whenever one of its parts does (and when, moreover, that part is appropriately related to its other parts), simply in virtue of the part having x (and being appropriately related to the other parts).

Containment-Boundedness: A set of experiences is containment-bounded if and only if there are some experiences it does not contain.

Different Contexts (DC): An experience of a whole is unified with different experiences than any experience of its experientially-different parts.

Dual Reference: " $\varphi(I_B)$ " means " $\varphi(I_p)$ and $((I_R = I_p)$ or $(I_R$ have a certain relation to $I_p)$)"

Epistemic Optimism: A subject who commits no 'epistemic wrongdoing' (e.g. never believes something for which they have insufficient evidence, etc.) will not inevitably and systematically believe falsehood.

Essential Boundedness (EB): The set of experiences belonging to any subject is bounded.

Experiential Combinationism: The experiential properties of a conscious subject may be fully explained by the experiential (and topic-neutral) properties of, and relations among, its parts.

Experiential Combinationism_N: The experiential properties of a jointly-conscious collection of things may be explained by the experiential properties of, and relations among, those things.

Experiential Compositeness: A conscious subject may have parts.

Experiential Compression (EC): For two experiences to belong to a subject, they must be distinguishable (i.e. not confused) with respect to the 'bringing into consciousness' operation, relative to that subject.

Experiential Nesting: A conscious subject may have parts which are themselves conscious subjects.

Experiential Simplicity: A conscious subject cannot have parts.

Explanatory Gap between Subjects (EGS): It is never the case that the existence of a number (one or more) of subjects of experience with certain phenomenal characters, standing in independently intelligible relations, renders intelligible the existence of some other subject of experience.

Heritability of Fundamental Properties (HF): *If* a property is fundamental, and *if* one of the level-connecting views of composition is true, and *if* the heritability of that property would not yield demonstrable incoherence, *then* that property is both-ways inherited.

Human Boundedness (HB): Each normal human has a bounded set of experiences.

Knowledge by Ownership (KO): Having an experience is necessary and sufficient for being able to know it directly.

Location Inheritance: A whole 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.

Mass Additivity: The mass of a composite entity is equal to the sum of the masses of its parts.

Monistic Conditional Experience Inheritance (CEI_M): A whole has an experiential property whenever one of its parts does and that part is appropriately related to its other parts, and that part has that experiential property simply in virtue of the whole having that experiential property, and the part being appropriately related to the other parts.

Negative Epistemic Boundedness: A set of experiences is negatively epistemically bounded just if its members confer on their subject no knowledge of other experiences outside that set which is not dependent on knowledge of something non-experiential.

Nihilistic Conditional Experience Inheritance (CEI_N): Some things have an experiential property collectively whenever one of them does so individually and that thing is appropriately related to the other things, and the things have that experiential property collectively simply in virtue of the particular thing having that experiential property individually, and being appropriately related to the other parts.

No Summing of Subjects (NSS): It is never the case that the existence of a number (one or more) of subjects of experience with certain phenomenal characters *a priori* entails the existence of some other subject of experience. (2009a, p.302; a slightly different formulation appears at 2009b, p.130)

Organisational Invariance (OI): “experience is invariant across systems with the same fine-grained functional organization” (Chalmers 1995b, p.310)

Organisational Invariance of the Attended (OIA): attended experience is invariant across system with the same fine-grained functional organization.

Pair-Persistence: In fusion cases, the original persons composed a pair (i.e. a mereological fusion of two persons) which became a person, and similarly in fission cases, the original person becomes a pair composed of the two resultant persons, and these facts, perhaps along with other facts, explain the sense in which these persons survive these processes.

Parity Claim (PC): If we were connected to another subject in the way that our parts are connected to each other, we would not regard that subject as a distinct entity, but instead would ascribe their decisions, thoughts, and experiences to ourselves.

Part-Persistence: In fusion cases, each original person becomes a part of the resultant person, and in fission cases, each resultant person was previously a part of the resultant person, and these facts, perhaps along with other facts, explain the sense in which these persons survive these processes.

Patterning Principle (PP): Our spontaneous impressions of whether an event is external or internal, and of whether it is our voluntary action or not, are determined by the patterns of correspondence and divergence we detect between it and other things.

Phenomenal Essentialism (PE): The phenomenal character of an experience is essential to it.

Phenomenal Holism (PH): The phenomenal character of an experience depends partly on its phenomenal context, i.e. by which experiences it is unified with.

Positive Epistemic Boundedness: A set of experiences is positively epistemically bounded just if any subject who experiences all and only those experiences is thereby enabled to know that they experience all and only those experiences.

Power Inheritance: A whole has a given causal power whenever any of its parts do, simply in virtue of the part having that power.

Power Sharing: Particular token causal powers can belong simultaneously to two different entities.

Security of Self-Reference: A subject consciously thinking a first-person thought always refers to themselves by the first-person term in that thought – there is no risk of it referring to something else, or failing to refer.

Strong Exclusivity (SE): A single experience cannot belong to multiple distinct subjects.

Strong Privacy: A single experience cannot be directly known by multiple distinct subjects.

Strong Publicity: A single experience may be directly known by multiple discrete subjects.

Strong Sharing (SS): A single experience may belong to multiple discrete subjects.

Subject-Relative Phenomenal Context: The phenomenal context of an experience e for a subject s = the set of experiences of s with which e is co-conscious for s .

Token-Sharing of Experiences (TSE): Particular experiences can belong simultaneously to two different entities.

Token-Sharing of x : Particular tokens of property x can belong simultaneously to two different entities.

Unification-by-Adumbration (UA): A subject has a unified experience of X-and-Y-together whenever they experience X, and experience Y, and these experiences are A-related.

Unification-by-Experiential-Adumbration (UEA): A subject has a unified experience of X-and-Y-together whenever they experience X, and experience Y, and these experiences are EA-related.

Unification-by-Objectual-Adumbration (UOA): A subject has a unified experience of X-and-Y-together whenever they experience X, and experience Y, and these experiences are OA-related.

Unified Composites from A-Related Parts (UCAP): A composite subject enjoys conscious unity whenever its component subjects are A-Related.

Unified Composites from EA-Related Parts (UCEP): A composite subject enjoys conscious unity whenever its component subjects are EA-Related.

Unified Composites from OA-Related Parts (UCOP): A composite subject enjoys conscious unity whenever its component subjects are OA-Related.

Unity Incompatible with Boundedness (UIB): A subject’s set of experiences is bounded only if its members are not unified with experiences outside that set.

Unity in Part and Whole (UPW): For the experiences of a composite subject to be unified, the experiences of its component subjects must be unified with each other,

Unrestricted Unity (UU): All experiences in the universe are phenomenally unified with each other, i.e. form a single phenomenal field.

Weak Exclusivity (WE): A single experience cannot belong to multiple discrete subjects.

Weak Privacy: A single experience cannot be directly known by multiple discrete subjects.

Weak Publicity: A single experience may be directly known by multiple distinct subjects.

Weak Sharing (WS): A single experience may belong to multiple distinct subjects.

x - y Connections: Something can have property y simply by having properties x_1, x_2, \dots

Appendix: Why Experiential Combination Matters, and to Whom

<u>Section 1: From Mental Unity to Metaphysical Simplicity</u>	p.350
<i>Subsection 1.1: The Achilles Argument, its History, and its General Structure</i>	p.351
<i>Subsection 1.2: Rejecting the ‘Emergence’ Option</i>	p.352
<i>Subsection 1.3: Rejecting the ‘Combination’ Option</i>	p.354
<u>Section 2: Homunculi and the Simplicity Diagnosis</u>	p.356
<i>Subsection 2.1: Two Homuncular Systems</i>	p.357
<i>Subsection 2.2: The Blockhead and the Chinese Room</i>	p.358
<i>Subsection 2.3: The Internalised Chinese Room and the Psychofunctional Homuncular Systems</i>	p.360
<u>Section 3: Panpsychism and the Combination Problem</u>	p.362
<i>Subsection 3.1: Physicalism, Psychophysical Laws, and the Combination Problem</i>	p.363
<i>Subsection 3.2: Different Varieties of Panpsychism</i>	p.365
<u>Section 4: Split Brains and Split Persons</u>	p.366
<i>Subsection 4.1: Pathological Splitting</i>	p.367
<i>Subsection 4.2: Everyday Fragmentation</i>	p.368
<i>Subsection 4.3: How Experiential combination Helps</i>	p.370
<u>Section 5: Fission and Fusion</u>	p.372
<i>Subsection 5.1: Fission, Fusion, and the Simple Soul</i>	p.372
<i>Subsection 5.2: What makes for Unproblematic Fission?</i>	p.375
<i>Subsection 5.3: Consequences of Rejecting Anti-Combination</i>	p.377
<u>Section 6: Problems of Multiplication</u>	p.378
<i>Subsection 6.1: How the Problems Arise</i>	p.379
<i>Subsection 6.2: The Tolerability of Object-Multiplication</i>	p.381
<i>Subsection 6.3: The Difference between Persons and Objects</i>	p.383
<i>Subsection 6.4: The Role of Anti-Combination - ‘Additionalism’ about Experience</i>	p.384
<u>Section 7: Anti-Nesting Among Materialists</u>	p.386
<i>Subsection 7.1: Anti-Nesting, Anti-Combination, and Parsimony</i>	p.387
<i>Subsection 7.2: Arguments Against Anti-Nesting</i>	p.389
<u>Section 8: Collective Mentality</u>	p.390
<i>Subsection 8.1: Collective Intentionality and Collective Consciousness</i>	p.391
<i>Subsection 8.2: Why Deny Collective Consciousness?</i>	p.392
<i>Subsection 8.3: The Role of Anti-Combination - ‘Additionalism’ about Experience</i>	p.394
<u>Section 9: Nihilism and Non-Singular Mentality</u>	p.395
<i>Subsection 9.1: An Argument Against Nihilism</i>	p.395
<i>Subsection 9.2: Are Nihilism and Combinationism Compatible?</i>	p.397
<i>Summary</i>	p.399

In this appendix, I will explain how several existing debates in philosophy of mind are substantially impacted, sometimes explicitly and sometimes implicitly, by the truth or falsity of the three mereological denials:

Experiential Simplicity: A conscious subject cannot have parts.

Anti-Nesting: A conscious subject cannot have parts which are themselves conscious subjects.

Anti-Combination: The experiential properties of a conscious subject cannot be explained by the experiential properties of, and relations among, its parts.

In section 1, I review the roles of Experiential Simplicity and Anti-Combination in a historically popular argument against materialism; in section 2, I review the role of Experiential Simplicity in driving a range of influential intuitions; and in section 3, I review Anti-Combination's role in a major contemporary argument against panpsychism. In section 4 I discuss various cases of synchronically divided people, moving in section 5 to diachronically divided people, i.e. people who split in two, or conversely who fuse together. In section 6 I consider what I call 'problems of multiplication', a family of puzzles whose most notable member is Peter Unger's 'Problem of the Many'. In section 7, I review recent defences of Anti-Nesting by materialists, and in section 8 I briefly consider some of the issues raised by collective mentality. Finally, section 9 discusses the relevance of combinationism to mereological nihilism.

One theme that comes out repeatedly in this discussions is the links among the mereological denials: accepting the weaker ones, such as Anti-Combination, pushes us towards accepting the stronger ones, such as Experiential Simplicity. This underlines the importance of my investigation of Anti-Combination, even when these like Simplicity and Anti-Nesting receive more widespread discussion.

Since the influence of the mereological denials is sometimes implicit and even unacknowledged, in some sections I adopt a somewhat hermeneutic stance, not reporting explicit endorsements of the mereological denials but rather elucidating how they play an tacit role in the framing of a debate or in the evaluation of options. Obviously this sort of claim is always open to objection, but my confidence is increased by the fact that some authors assert the explanatory

indivisibility of the mental in passing, as when Lowe simply declares that “the self patently does not consist of a plurality of lesser ‘selves’ acting cooperatively, despite the picturesque ‘homuncular’ descriptions of mental functioning advanced by some philosophers. Such descriptions are not intelligible if taken literally.”(1996, p.39) If it seems to some authors obvious enough to assert without argument, it may well seem to other authors obvious enough to leave implicit and unarticulated.

Section 1: From Mental Unity to Metaphysical Simplicity

There is a long tradition of arguing for Experiential Simplicity as a step towards the rejection of materialism. Since it seems central to almost all material things that they be divisible into material parts, Experiential Simplicity would preclude subjects of experience being material things.¹

Not all arguments for simplicity are relevant to combinationism. In particular, some arguments seek to show that all real substances must be simple, not just minds (see, e.g. Leibniz 1967, pp.88-89). What is relevant is the claim that something distinctive about our mental life would be impossible, were we composites. An argument of this sort appears in the *Critique of Pure Reason*, where Kant calls it ‘the Achilles of all dialectical inferences of the pure doctrine of the soul’ (A351): apparently invincible but in fact fatally flawed. This label has been employed by recent interpreters, in particular in Lennon and Stainton’s anthology ‘The Achilles of Rationalist Psychology’ (2008, throughout), and

¹ Strictly speaking, one might reconcile Experiential Simplicity and materialism by means of atomism, holding that each mind is a single indivisible material atom. While this view may have been held by the ancient atomists (cf. Skrbina 2005 pp.33-34), it has become less and less plausible as our understanding of matter, and of the brain, has advanced. As far as physical science can tell, there is no material particle which remains in the brain through a person’s life, nor does any atom seem to be remotely privileged as the site where ‘it all comes together’. If we are individual particles, then we are outwardly indistinguishable from trillions of other particles performing similar roles in the brain – and must expect to lose our sophisticated human-like consciousness unexpectedly at some point, when we are excreted from the brain and then from the body. Insofar as this view is extremely bizarre, and is the only way to reconcile Experiential Simplicity with materialism, we may say that Experiential Simplicity still poses a major threat to materialism.

Mijuskovic's 1984 book 'The Achilles of Rationalist Arguments.' The following discussion draws heavily on this previous scholarship.

Subsection 1.1: The Achilles Argument, its History, and its General Structure

The first clear presentation of 'the Achilles' comes in Plotinus (1956, pp.255-258, 342-356) with subtly different versions then appearing in Proclus (1963, p.163) and Avicenna (1952, pp.47ff). It becomes extremely popular during the early modern period, with a highly abbreviated version appearing in Descartes' Meditations (1985, Volume 2, p.59), and other versions being given by Butler (1860), Mendelssohn (2002), Clarke (1978, Volume 3, p.759), Bayle (1991, pp.128-134), and other less well-known figures like Francois Lamy (1701, pp.118-119) and Ralph Cudworth (1837). Kant judges the argument a failure, but only because its conclusion (Experiential Simplicity) and the negation thereof are equally beyond the possibility of human knowledge. The argument survives into the 19th and 20th centuries in figures like Lotze (1894, p.158) and Brentano (1987, pp. 290-301), before receding somewhat from view before the advance of both anti-metaphysical attitudes and reductive versions of materialism.

Many of these authors go on to argue from the mind's simplicity to its immateriality, immortality, or both, but I will not directly examine these further steps.² The exact form that the argument takes, and the forms of mental unity appealed to, differ among different authors, but a recognisable structure underlies many disparate versions, in which we suppose for *reductio* that mental life inheres in a composite, distinguish an exhaustive array of possibilities, and show that each must fail.³ Not all versions review all the options⁴, but it is charitable to suppose that this stems not from

² In the terms employed by Lennon and Stainton, my focus is on the 'narrow Achilles', not the 'broad Achilles'.

³ This is close to what Lennon and Stainton call the 'narrowest Achilles', though I distinguish the options in a slightly different way than they do.

failing to notice them, but from assuming that their problems are sufficiently obvious to go without saying.

The argument proceeds thus: if a composite had mentality, we might explain this by reference to mentality in only some of its parts, or by reference to mentality in all of its different parts, or independently of the mentality of its parts. Let us provisionally call the first option ‘partition’, the second ‘combination’, and the third ‘emergence’ (without implying any definite affinity with any particular account of ‘emergentism’, strong or weak).

Partition is the most obviously unsatisfactory, for it merely pushes the question back one step – does the part responsible for the whole’s mental life itself have parts? If so, we may repeat the question, and if not, then mental life has been traced to a simple entity after all. What is more significant is why both ‘emergence’ and ‘combination’ are also rejected

Subsection 1.2: Rejecting the ‘Emergence’ Option

Several authors disregard the ‘emergence’ option entirely, and take experiential combination to be simply ‘what follows’ from materialism. Here are three examples:

[S]uppose that the composite were thinking; then every part of it would be a part of the thought, but the parts would first contain the whole thought only when taken together. (Kant 1781, A352)

If we could be divided, each separate part of us would have the power of consciousness in itself; and there would follow separate consciousnesses, which is contrary to our hypothesis.

⁴ For instance, Avicenna never mentions what I will call the ‘emergence’ option, and Plotinus never mentions what I will call the ‘partition’ option, though both discuss the others.

We, therefore, the living beings that we call ourselves, are indivisible. (Hinxman Duke 1847, summarising an Achilles argument in Butler 1897, quoted p.85)

If [the soul] had the nature of body... there would be a particular soul... answering as a distinct and independent entity to every local experience. (Plotinus 1956, p.257)⁵

But why is the alternative (in which the composite has mentality independently of whether the parts do) rejected? When reasons are given, they usually involve claims like the following, which Rozemond (2008) calls ‘homogeneity principles’:

...every Power or Quality that is or can be inherent in any System of Matter is nothing else than the Sum or Aggregate of so many powers or qualities of the same kind, inherent in all its Parts. (Clarke 1978, VIII p.759)

...the action of a composite... is an aggregate of many actions or accidents, which is distributed among the multitude of substances... (Kant 1781, A352)

...no acting power can be produced in the whole, the source of which is not to be found in the constituent parts... (Mendelssohn 2002, s.II)⁶

These claims might be read as requiring that every property of a whole be found in the parts, but that would be an uncharitably extreme reading susceptible to many easy counterexamples.⁷ Instead, they seem to be expressing something like ‘general combinationism’, the idea that the part-whole relation is explanatory, however we understand the relevant sort of explanation (cf. Chapter 2, especially subsection 2.1 and section 3). Even if this reading does not capture everything intended by all the authors, it is plausibly a minimal requirement shared among them.

⁵ These passages do not explicitly say that the many consciousnesses of the parts explain that of the whole. But we may infer that intention, from the fact that postulating the many is taken to be justified by recognising the whole’s consciousness, together with the assumption that the whole’s consciousness is in need of explanation.

⁶ See also Bayle 1991, p.129: “From the supposition that each atom is not animated, it follows that a collection of atoms feels nothing...”, and Plotinus, *Enneads* p.343, “...it [is] impossible, that life and soul should result from a congeries of lifeless and soulless things; or that mindless things put together, should beget mind...”, here translated and endorsed by Cudworth 1837.

⁷ Powell 1990, pp.94-95, offers precisely this reading, and provides ‘circularity’ as just such an easy counterexample.

On this reading, the proponents of the Achilles argument need a further premise to entirely rule out the ‘emergence’ option. An opponent might accept the homogeneity principles, but claim that the non-mental properties of the parts may transparently explain the mental properties of the whole. To rule this out requires a premise denying that the mental can be transparently explained through the non-mental, i.e. an assertion of the irreducibility of the mind, which I believe the authors in question plausibly would accept. Given these two premises (‘irreducibility’ and ‘homogeneity’, or in my terms anti-physicalism and generalism combinationism), and given the inadequacy of ‘partition’, we can validly move from Experiential Compositeness to Experiential Combination – or, as, the Achilles argument does, from Anti-Combination to Experiential Simplicity.

Subsection 1.3: Rejecting the ‘Combination’ Option

The combination option is often divided into two versions: the ‘whole-in-part’ view and the ‘part-in-part’ view. On the former, when a composite object has some mental state, all of its parts have that mental state, so that each part has the whole mental life of the composite. On the latter view, each part has only a partial form of the composite’s mental life, which ‘add up’ when taken together. There are specific objections given to each version, but the most fundamental objection to both is that “representations that are divided among different beings... never constitute a whole thought” (Kant 1781, A353). This claim is repeatedly brought out by means of an analogy between the parts of a composite subject and a plurality of separate people, examples of which have already appeared in chapter 1, subsection 2.1, and chapter 3, subsection 1.1. As I argued in chapter 3, the most charitable interpretation of this move is that it functions to make vivid the appearance of an explanatory gap between component subjects and composite subjects (thereby supporting the principle I there called

EGS). Insofar as it is the claim of an explanatory gap that lies behind this rejection, the success of the proposals I have made means the failure of the Achilles argument.

This reading of the argument would fit neatly with my earlier suggestions about the homogeneity principles used to rule out composites having mental life even though their parts do not: in both cases the argument relies on the idea that properties of a whole should be transparently explainable through its parts. The analogy to groups of humans serves to show that attributing mental life to the parts will do nothing to entail or explain the mental life of the whole.

Several more specific reasons are given for rejecting particular forms of experiential combination. The part-in-part view is sometimes criticised on the grounds that some particular sort of mental state could not arise from other mental states. So it may be claimed that multiple simple ideas could not add up to any more complex idea, so that the whole's thoughts would be just as crude and basic as those of its smallest parts (e.g. Bayle 1991, p.130). Alternatively, it may be claimed that it is simple ideas in the whole that are most difficult to explain, since they cannot be divided up into parts to attribute to parts of the thinker (e.g. Avicenna, 1952, p.48). Or it might be claimed that many unclear or confused ideas could not add up to any clearer or more distinct ones (e.g. Mendelssohn 2002, s.II).

Another argument turns on the intensity of feelings: if a given feeling is 'spread out' among many parts, and each receives only a fraction of its intensity, then the whole will be unable to add them together to form an intense feeling (e.g. Bayle 1991, p.131). A final argument focuses on how inter-modal contrasts and comparisons could be made if each part of the composite were only aware of one or other sensory modality (e.g. Plotinus 1956, pp.346-347).

Most of these specific objections have been addressed implicitly by some part of the present work - for instance, the concerns over inter-modal integration and complex thoughts touch on the representational unity of consciousness, discussed in chapter 4, while the concerns over simple ideas

touch on the possibility of phenomenal blending, discussed in chapter 5. The objection to the whole-in-part view is the opposite: it avoids all the above specific difficulties at the cost of being unsatisfying as an explanation. If it has to ‘build in’ the whole’s entire mental life at the level of the parts, little is gained in our understanding of the basis of the mind, since the mental life of the parts will appear just as in need of explanation as that of the whole.

While these arguments often appeal to relatively diverse and theory-specific doctrines about mental structure, I have tried to suggest that there is a common thread to many versions: the feeling that properties of composites must be explained through their parts, and that consciousness cannot be thus explained. That is, the common-thread argument has as its premises general combinationism, anti-physicalism, and Anti-Combination. The defence of combinationism is thus a defence of the coherence of substance-monistic primitivism about consciousness, i.e. of views like property dualism and neutral monism that reject both physicalism and substance dualism.

Section 2: Homunculi and the Simplicity Diagnosis

While few contemporary philosophers of mind explicitly endorse Simplicity, they often appeal to intuitions about thought experiments in which the composite nature of some entity is salient. David Barnett has argued that the best explanation for some of these intuitions is that “our naïve conception of a conscious being demands that conscious beings be simple” (2008, p.309). This section discusses some of these thought-experiments, evaluating what role mereological considerations may be playing, with Barnett’s claim as a convenient point of reference.

Note one initial problem for Barnett’s diagnosis: in everyday life people seem quite happy to attribute consciousness to animals which are, to all appearances and by many accounts, composite. Barnett gives two responses: first, “when we say that human bodies are conscious, we simply mean that

human bodies are occupied by conscious beings”(p.313); second, “as our bodies are ordinarily presented to us, it is easy for us to ignore their composite aspect”, for when the parts “appear to be spatially continuous with one another, the whole body presents itself to our minds, not as a system of independently existing parts, but rather as something like an ‘extended simple’.”(p.314)

In an attempt to make these claims more compelling, Barnett imagines seeing a human body through a series of ‘magical goggles’ which allow us to see its fine structure down to the level of subatomic particles. Looked at in this way, Barnett claims, our willingness to ascribe consciousness to this widely scattered system evaporates, and it seems no better a candidate for consciousness than the bizarre creatures philosophers have imagined. It is to these imaginations that I now turn.⁸

Subsection 2.1: Two Homuncular Systems

First consider two famous thought-experiments from Block 1992: the Homunculus-Head and the Nation of China. In both scenarios, the outward behaviour of an intelligent human person is functionally simulated by a vast collection of individual people, all able to see an array of inputs analogous to the person’s sensory inputs and a bulletin board that records the analogue of the person’s present internal state. In the ‘Homunculus Head’ case, these many people are miniscule in size, and locked in the hollowed-out cranium of an outwardly-normal human body. In the ‘Nation of China’ case, these many people are the actual citizens of some sufficiently large human society, the ‘bulletin board’ is a series of satellites, and their ‘body’ is normal-sized and controlled remotely.

⁸ I am focusing only on cases which directly and vividly involve mereological considerations. Barnett tries to extend the Simplicity diagnosis to account even for intuitions such as the explanatory gap between matter and consciousness, and the possibility of zombies. To the extent that such a diagnosis is plausible, I think discussing it would merely repeat the material discussed in section 1.

Each ‘homunculus’ has a very simple, inflexible, job, defined by a single input, a single output, a single internal state, and a single change of internal state. If the bulletin board is displaying the right state, and the right input comes in, that homunculus must implement their allotted output and change the internal state. By allocating a homunculus to each square of the hypothetical machine-table which entirely captures the input-output function of a human being, we could specify a system functionally identical to that human being, and thus according to functionalism – the doctrine that all mental state types are identical with some functional type – the stipulated system must be conscious.

Block hopes that these cases undermine functionalism, because “there is *prima facie* doubt whether [such a system] has any mental states at all – especially... ‘qualitative states’”(p.76), a doubt that then transfers to functionalism itself. More forcefully, he later refers to the attribution of consciousness to such systems as “an absurdity” (p.79). Barnett argues that our responses to these cases really reflect the Simplicity intuition. I think there is something to Barnett’s diagnosis, though I suspect it is over-stated. One reason for caution is that Block gives something close to a direct response to a mereological construal of his thought-experiments, namely the ‘Tiny Alien Spaceships’ thought-experiment described in chapter 5, subsection 3. Barnett responds that we are only inclined to attribute consciousness to the system of tiny spaceships because we attend to its strong isomorphism with the system of elementary particles composing a human being, and attribute consciousness to the latter due to our ordinary failure to attend to its composite aspect.

Subsection 2.2: The Blockhead and the Chinese Room

Here is a second reason for caution about reading Block’s examples as related to the Mereological Denials. Block 1981 offers a different thought-experiment, which serves a similar argumentative function, but does not involve a subject composed of subjects. This is the ‘Blockhead’, a humanoid

device controlled by a rudimentary mechanism searching a vast look-up table which contains every possible one-hour-long intelligent-seeming English conversation. If allowed to search this galaxy-spanning list at superluminal speeds, the device could easily simulate intelligent conversation (for an hour) merely by finding on its list a ‘canned’ conversation which matches its present one up to the last-received utterance. This device, Block thinks, is obviously not intelligent – its appearance of intelligence derives from the intelligence exercised by its designers in exhaustively distinguishing the cogent from the non-cogent strings. The point is that intelligent behaviour does not entail intelligence, nor consciousness (insofar as intelligence is a sign thereof); the structure of the internal mechanisms involved may be such as to ‘defeat’ any initial assumption of intelligent causes for intelligent behaviour.

The intuition that the Homuncular systems are not conscious might derive from similar considerations – not the component subjects, but the structure of information-processing. Just as the real intelligence in the Blockhead lies in the design of the huge look-up table, so we might think the real intelligence in the Homuncular systems lies in the design and allocation of the fantastically precise individual roles – the running of the system gives an illusion of intelligent behaviour because of this ‘stored’ intelligence.⁹ So any negative intuition here admits of two alternative explanations: Barnett’s ‘simplicity’ reading and Block’s intended ‘anti-functionalist/anti-behaviourist’ reading.¹⁰

⁹ Jackson 1993 offers an alternative view of what is wrong with Blockhead – that its intelligent-seeming behaviours at different moments are entirely causally independent of each other, because the content of one line of the table is independent of the others. This is equally germane to reading our reaction to the Homunculus-Heads as reflecting something other than the Simplicity intuition: what each homunculus does is entirely independent of what the others do, except insofar as they change the bulletin board. The instructions assigned to one homunculus do not constrain the instructions given to another.

¹⁰ Potentially there might be many other candidate explanations to consider: for instance, Rory Madden has argued that the best explanation is that “our naïve conception of a conscious subject demands that conscious beings be topologically integrated” (2012, p.2), where integrity is treated as a new primitive notion, stronger than mere spatial contiguity (so that two people shaking hands do not form an integrated whole; see Schwitzgebel 2014 for arguments against this principle, there called ‘contiguism’). Here I focus on the most influential rival explanation – that which emerges from Block (and Searle’s) own work.

Something similar goes for another famous thought-experiment, John Searle's 'Chinese Room' (1980/2003), which Barnett also diagnoses as an instance of the Simplicity intuition. Here we imagine a monolingual English-speaker imprisoned in a room with a huge look-up table correlating strings of Chinese with each other. The content of this table is such that when one string is sent into the room as 'input', the string which the table instructs the human to send out as 'output' is an appropriate, intelligent, response. The system can pass the Turing test in Chinese, but the human operator *ex hypothesi* understands no Chinese at all. Does the system itself (human+room+table+...) understand Chinese? Searle thinks it intuitive that it does not, and Barnett suggests that this is because it is so obviously a composite. Searle's own diagnosis, however, focuses not on the mereology of the system but the way it processes information: it manipulates symbols merely as syntactic units, not as meaningful representations.

So with both Block's and Searle's thought-experiments, there are rival explanations available for the intuitions they elicit. However, both authors go on to offer adjusted versions of their thought-experiments which alter the balance of plausibility between Barnett's diagnosis and their own – though in different directions.

Subsection 2.3: The Internalised Chinese Room and the Psychofunctional Homuncular Systems

In response to opponents who attribute understanding to the Chinese-Room system as a whole, Searle imagines the human operator "memoriz[ing] the rules in the ledger and the data banks of Chinese symbols, and [doing] all the calculations in his head" (1980/2003, p.337). Searle thinks it remains just as intuitive that this person does not understand Chinese; but now the composite nature of the 'system' is no more obvious than that of a normal Chinese speaker, so Barnett's Simplicity diagnosis does not apply. Thus, insofar as the no-understanding intuition about the original Chinese-Room system is

preserved in this adjusted version, Barnett cannot use that intuition in his case for the Simplicity intuition.

Conversely, Block's adjustment strengthens the case for the Simplicity diagnosis. He imagines a Homuncular system whose homunculi are related in such a way as to match the parts of the actual human brain in all respects which empirical psychology can discern. This is a significant change of internal structure, and Block thinks that it does make a real difference to the plausibility of attributing many mental states to the system – in particular, he thinks it is plausible to attribute propositional attitudes, memories, beliefs, and so on. But “there is perhaps as much doubt about the qualia of this homunculi-headed system as there was about the qualia of the homunculi-headed Functional simulation discussed earlier”(1992, p.306).

If the original intuition (that such an attribution is not only ‘doubtful’ but ‘an absurdity’) persists into this case, then the most obvious explanation seems to be mereological – the only notable difference between the ‘psychofunctional Homunculus-Head’ and a normal human is that the former contains component subjects. So this intuition is germane to Barnett's Simplicity diagnosis. That is, we might plausibly think that the negative intuition about the initial Homuncular systems' consciousness was over-determined: it reflected both the Simplicity intuition and also a judgement that the structure of information processing involved was inappropriate. Having removed the latter factor, any residual negative intuition may be attributed to the Simplicity intuition.

In fact, Barnett himself offers a similar thought-experiment. He first observes that while we can “consider what it might have been like to be Descartes as he wrote the Meditations, or to be Hobbes as he fled the English Civil War”, it seems absurd to “consider what it might have been like to be this pair of philosophers during these events”(2008 p.312). To remove the confounding factor of degree and form of interaction, Barnett imagines taking Descartes and Hobbes, shrinking them to the size of

someone's cerebral hemispheres, training them to exactly imitate the role played by those hemispheres, and then putting them into someone's head as a replacement for the hemispheres. *Ex hypothesi*, the person continues to behave exactly as before, and all psychology above the level of the hemisphere remains true of them. Yet they are now the limiting case of a psychofunctional Homunculus-Head. Barnett thinks it is still absurd to think that the pair themselves might be conscious.

In conclusion, it seems to me that some but not all of the intuitions Barnett discusses admit of non-mereological explanations. If we refuse to attribute consciousness to Searle's Chinese Room and Block's original Homunculus-Heads, this may well be due to their manner of information-processing, not their mereological structure. However, negative intuitions about the psychofunctional Homunculus-Heads, and Barnett's hemisphere-imitating-Early-Moderns, probably reflect the Simplicity intuition. This supports a limited endorsement of Barnett's suggestion that the idea of the simple soul still exercises an influence in contemporary philosophy of mind, even though most philosophers reject it. But Barnett proceeds no further: he takes intuitions as evidence of naïvely-accepted principles, but does not seek for or provide arguments that might explain the attractiveness of those principles. Given the connections among the mereological denials, noted throughout this appendix, I think it is a live possibility that the intuitive attractiveness of Experiential Simplicity depends on a prior, implicit, acceptance of Anti-Combination.

Section 3: Panpsychism and the Combination Problem

Much existing discussion of experiential combination stems from the recent revival of interest in panpsychism¹¹; indeed, I have taken the term 'combination', as meaning explanatory, non-emergent,

¹¹ I will understand '(naturalistic) panpsychism' as the doctrine that all which fundamentally exists has both physical (i.e. accurately described at least in part by physics) and conscious properties. This is a stronger claim than both 'micropsychism', the doctrine that this holds of at least some fundamental physical entities (cf. Strawson 2006), and

composition, from the ‘combination problem’ supposed to face panpsychism (Seager 1995, p.280). To a significant degree, the viability of panpsychism turns on its explanatory potential, and to a significant degree its explanatory potential relies on experiential combination. Consequently many critiques of panpsychism involve arguing for Anti-Combination, and the arguments that arise in this process have been a major source of material for this dissertation. In this section I will try to briefly explain the dialectical position and background of panpsychism, and how it relates to combinationism.

Subsection 3.1: Physicalism, Psychophysical Laws, and the Combination Problem

A certain widespread narrative about reasons for considering panpsychism starts with dissatisfaction about physicalism as an explanation of consciousness (Nagel 1986, Seager 1995, Chalmers 1995, Strawson 2006). If a non-mental understanding of physical matter leaves an ‘explanatory gap’ between physics and consciousness, and all physicalistic attempts to ‘close the gap’ are abandoned, one natural alternative is to accept the fundamentality of consciousness. Yet consciousness does not seem to actually vary independently of the physical, suggesting that they are linked by *a posteriori* nomological necessities, even if not by *a priori* conceptual necessities. The resultant picture is one on which fundamental psychophysical laws ‘bridge the gap’, connecting physical and mental properties in an epistemically opaque way, just as different physical properties are related by fundamental physical laws.

So far this is not necessarily panpsychist; strong emergentists might also posit fundamental psychophysical laws. The next move is to point out that fundamental laws tend to be simple and general, allowing for a great variety of forms to be built up gradually from a small set of widespread

‘panprotopsychoism’(cf. Chalmers 2013b), the doctrine that all fundamental entities have features somehow ‘akin to’ consciousness.

basic elements – not attaching a basic element to a precisely-specified sort of rare and complex structure. Moreover, it is arguably this simplicity and generality that makes physics an explanatorily satisfying framework – by contrast, the more narrowly applicable ‘emergence laws’ posited by emergentists seem unsatisfyingly *ad hoc*. Consequently, we should expect psychophysical laws to put consciousness more or less everywhere: if the mental is fundamental (Cf. Blamauer 2011), then panpsychism is the natural conclusion.

This is not the only argument for panpsychism; there are other considerations about intrinsic vs. relational natures, and about reconciling physical with mental causation, which have been appealed to (Strawson 2006, Coleman 2009, Rosenberg 2004). But this two-stage argument (labelled by Seager 1995 the ‘genetic argument’ for panpsychism) is my interest here, and this argument saddles panpsychism with a certain explanatory burden. Because it seeks to add new fundamental laws only at the basic level, it still needs to explain how complex things like human minds are ‘built up’ from the basic experientiality of matter. In short, it must be able to provide the sort of transparent, intelligible explanation of human-level consciousness that physicalism could not.

Unfortunately, there is a widely-voiced fear, coming from both opponents of and sympathisers of panpsychism, that this explanation is impossible: that experiences simply do not combine in the necessary way. Sometimes this appears simply as an admission of bafflement at the idea (e.g. Nagel 1986, p.50), but just as often it is claimed as conceptually evident that no set of facts about a multitude of related minds can explain a composite mind: as Van Cleve puts it, “the prospect of this seems no brighter than of the mental following from the purely physical.”(1990, p.219, cf. chapter 3, section 1) Others have argued that even with experiential combination, panpsychism would necessarily yield a different sort of experience than we have, one in which the underlying multitude would be vividly apprehended:

If there are billions of phenomenally-qualified ultimates bustling away in the composite structure of my mind... the different phenomenalities of each one ought to be accessible to introspection. Yet this is precisely what we do not find. (Coleman 2012, p.143, cf. chapter 5, section 1)

Subsection 3.2: Constitutive and Emergentist Panpsychism

Not all panpsychists are combinationists. Indeed, while most panpsychists deny Simplicity (because they attribute consciousness to familiar composite objects like brains) and Anti-Nesting (because they attribute consciousness to the simpler parts of such objects), there are still exceptions. Leibnizian ‘monadists’ could affirm Experiential Simplicity by maintaining that all and only fundamental physical simples are conscious, and each of us is one of them, while Tononi (2009, 2012) provides an example of a panpsychist (or something extremely close to it) affirming Anti-Nesting.

But even among panpsychists who accept Experiential Compositeness and Experiential Nesting, we can distinguish those who do and those who do not endorse Experiential Combinationism. There are a number of subtle boundaries to draw here, which have been marked with a number of different labels, such as ‘constitutive’ and ‘emergent’ (Chalmers forthcoming-b), ‘reductive’ and ‘emergentist’ (Goff 2010), ‘constitutive’ and ‘non-constitutive’ (Mørch 2013), or ‘constitutive’ and ‘causal’ (Mørch ms). Sometimes, by contrast, the distinction drawn is between different sorts of emergence, such as ‘weak’ and ‘strong’ (Chalmers 2006), ‘conservative’ and ‘radical’ (Seager forthcoming), or ‘weak’, ‘strong’, and ‘brute’ (Mørch ms).

Some of these distinctions concern the panpsychist’s metaphysical claims - does microexperience *ground* macroexperience? Others concern the panpsychist’s explanatory claims - does microexperience *explain* macroexperience, and if so by what standard of intelligibility? Different panpsychists disagree over the relationship between these two questions (e.g. does a claim of grounding

imply *a priori* deducibility). But overall a rough division can be marked out between a camp of panpsychists who seek a tighter relationship (explanatory and metaphysical) between conscious wholes and parts, and those who prefer to loosen this relationship as much as is compatible with retaining some theoretical advantage for panpsychism over non-panpsychist forms of physicalism and radical emergentism. In this work I have used the labels ‘constitutive’ and ‘emergentist’ for these two broad camps, taking the former to be committed to combinationism.¹²

This dependence on experiential combination is sometimes used as an argument within the panpsychist camp, against constitutive versions and in support of emergentist versions. It is also often used as in attacks on panpsychism as a whole, by critics who assume that the only or best forms of panpsychism must be constitutive. One common reason for this assumption is the perception that non-constitutive panpsychism offers no explanatory advantage over non-panpsychist emergentism. Whatever the merits of this assumption, the important point is that ‘the combination problem’ has been used as a weapon against constitutive panpsychism, both by other panpsychists and by external critics. Insofar as this work has shown combinationism to be defensible, it has thereby defended constitutive panpsychism.

Section 4: Split Brains and Split Persons

In this section I consider cases intermediate between what we would normally count as one person and what we would normally count as two persons, which challenge the synchronic individuation of

¹² Note that in these terms, it is entirely possible for a thing-monist variety of panpsychism (as found in e.g. Jaskolla & Buck 2012, and in such historical figures as Spinoza and Schopenhauer) to count as ‘constitutive’, if the consciousness of the subordinate parts can be derived from that of the whole by epistemically transparent principles. As I argued in subsection 3.2 of chapter 1, the important thing is not whether wholes or parts are ‘prior’, but whether a total description given at one mereological level is sufficient to yield a description at other levels without any additional laws or facts being added. This is reflected in the widespread recognition that monism may reformulate the combination problem, but still faces a version of it: explaining how the one cosmic mind generates many finite minds.

persons. I will argue that this kind of intermediate case is much easier to accommodate and make sense of within a framework which admits experiential combination, and conversely is problematic within a framework committed to Anti-Combination.

Subsection 4.1: Pathological Splitting

Difficulties of individuation arise particularly in two complementary pathological cases: the split-brain syndrome and dissociative identity disorder.

The details of the split-brain case are complex, and have been ably explained by others (Gazzaniga et al. 1962, Sperry 1964, Nagel 1971), so I will merely note the essentials: when the major direct connection between the two cerebral hemispheres is severed, the patient behaves normally in everyday life but shows dissociated responses when given dissociated stimuli. That is, if some stimulus is presented only to the sense organs connected to one hemisphere, the other remains unaware of it, and the patient responds (intelligently, flexibly, and to all appearances consciously) only with the means available to that hemisphere, such as one hand, and speech in the case of the dominant hemisphere, while simultaneously ignoring it (intelligently, flexibly, and to all appearances consciously) with the means available to the other hemisphere.

There is thus a lack of mutual availability between what seem to be two streams of processing, each associated with different identifiable physical structures, but with little difference in personality. Dissociative Identity Disorder (previously called Multiple Personality Disorder) presents a complementary profile: multiple 'alters' which may differ radically in personality, including memories, but which typically alternate in control of the whole body and all its organs. (More complex cases

where different alters claim to be simultaneously present, or to have ‘telepathic’ access to each other’s thoughts, have also been reported).

There are many views on how to describe these cases. They might involve two people sharing a head, or merely a single person whose thoughts and experiences are dissociated from each other in interesting ways. We might take the same line on both cases or we might, impressed with the differences between them, describe them in different ways – e.g. Tye 2003 rejects a ‘two person’ account of the split-brain case in part because he thinks that Dissociative Identity Disorder offers a more compelling picture of what ‘two people in one body’ looks like. A particularly interesting approach to the split-brain case is suggested by Bayne & Chalmers 2003: a single subject has a single phenomenal field within which are two functionally divided clusters of functionally unified experiences. In a slogan, there is phenomenal unity without access-unity. Other approaches involve the right hemisphere being a highly sophisticated but wholly unconscious automaton, or the patient’s mind being unified most of the time and then split only during experiments, or conscious unity being nontransitive, so that some experiences are disunified despite being unified with the same other experiences. So far no proposal has achieved consensus.

Subsection 4.2: Everyday Fragmentation

A lot of non-pathological cases suggest the beginnings of ‘internal division’, when mental conflicts become so intense as to prompt descriptions like ‘I am at war with myself’, or ‘I am enslaved by my passions’. This phenomenon is summarised by Isaiah Berlin thus:

‘I am slave to no man’; but may I not be a slave to nature? Or to my own ‘unbridled’ passions? Are these not so many species of the identical genus ‘slave’ -- some political or legal, others moral or spiritual? Have not men had the experience of liberating themselves from spiritual slavery, or slavery to nature, and do they not in the course of it become aware, on the one hand,

of a self which dominates, and, on the other, of something in them which is brought to heel?
(1957, p.17)

Berlin is critical of this line of thought, but documents its extensive history, in Kant, Rousseau, Plato, the Stoics, and others. In some writers the inner division does not seem to be strictly a mereological one: when Kant distinguishes the transcendental and the empirical self, the relation between them is clearly something more abstruse than that of two parts of one whole. But in other writers, of whom Plato and Freud are the most notable examples, it does seem to lead to an account of the mind as really composite, containing reason, spirit, and appetite (2000, p.111 ff) or ego, superego, and id (1923). And even writers who do not speak of particular parts do sometimes discuss of the importance of establishing and maintaining a singleness of psyche, i.e. of avoiding inner division (Korsgaard 2009, Frankfurt 1987). Others analyse even everyday decision-making as involving the collaboration of ‘homunculi’, quasi-agential systems in the brain (Selfridge 1959, Dennett 1991, cf. De Sousa 1976).

It is not clear how much of this talk is best taken as asserting a literal compositeness in human minds, and even if a given statement does assert compositeness, it is not clear what type of entity the parts are meant to be – Platonic ‘appetite’, for instance, seems to be a mental being of some sort, with goals that guide its actions, but should we think of it as a conscious subject with its own phenomenology? And if I fail to achieve genuine unity of self, what or who are the things which exist in place of the single self which does not exist? I certainly do not mean to insist on a literal, metaphysical, reading of all such talk, but I think it indicates a realm of interesting questions about the individuation of persons, just as the pathological cases in their own way raised such questions.

Finally, alongside these real cases we can place a variety of interesting imagined cases employed in philosophical thought experiments. Parfit 1984 slightly adjusts the split-brain case by giving each hemisphere autonomous control over half the body, allowing ‘the patient’ to simultaneously pursue two mentally intensive tasks at once with different hands (pp.246-249).

Moreover, he imagines being able to switch quickly and safely between a divided and an undivided mode, putting additional pressure on the idea that in the divided mode there are two distinct people (and connecting this example with the fission and fusion cases discussed in the next section). Van Inwagen 1990 describes ‘Cerberus’ and ‘Supercerberus’, single organisms with multiple cerebra or multiple whole brains, which may be either tightly coupled or operating independently (pp.191-194).

Subsection 4.3: How Experiential combination Helps

The questions raised by these cases are complex and multifarious, and go well beyond the mere truth of combinationism. I do not wish to take a stance on the proper description of each of the normal, pathological, and hypothetical beings described. Rather, I think that how we approach questions about them is systematically shifted by the truth or falsity of combinationism. This is best seen by considering an entity for which we are not intuitively inclined to support an analogue to Anti-Combination, i.e. something we naturally regard as explanatorily divisible.

For instance, consider the question: how many *brains* does the split-brain patient have?¹³ We might say ‘two’, or we might (more likely) say ‘one’: the plausibility of the two answers could be adjusted by imaging progressively greater surgical disconnection. But even when we are not sure which answer to give, this uncertainty seems merely semantic: rather than being baffled by the situation, we are merely unsure how to describe a physical situation which we intuitively understand fairly well. This is because we can always shift away from a language of discrete countable brains, and explain what is going on in terms of the neural parts and their relations to one another. We could say, for instance, that all the normal parts of a brain are present, but they are no longer interacting electrochemically in the

¹³ Readers who do not regard brains as explanatorily divisible should run the example using some inanimate object partly bisected, like ‘two’ pieces of paper connected only by a small bit the scissors missed.

way they were before. And insofar as brains are explanatorily divisible – insofar as knowing that brain parts exist and knowing how they are related suffices to explain everything about whole brains – this tells us everything we need to know.

If conscious minds too were explanatorily divisible, then we might similarly duck the question of number and give an adequate explanation of the mental situation in terms of a change in the relations among two mental parts which were already there in the normal case. But if we accept Anti-Combination, a description in terms of parts will simply not address our questions about the presence or absence of a single whole mind, and conversely a description in terms of a single mind will leave open whether or not there are a plurality of component minds. The two modes of description stand independent, and hence we face a stark choice between them: as Nagel puts it, “Something... in the ordinary conception of experience, leads to the demand for an account of these cases which the same conception makes it impossible to provide.” (1971, p.409)

As well as removing this sense of deep difficulty, a combinationist framework would often be more flexible, because relations among parts can be matters of degree, whereas being one mind or two minds cannot (it is little help to say that the split-brain patient has one-and-a-half minds). Moreover, this lets us finesse the question of whether, and at what point, experiences of internal division literally threaten the metaphysical singleness of the person. If all persons are, even in the best cases, structures of interacting sub-persons, then a loss or gain of agreement and coordination among these components can involve both a literal division in the mind, and also a mere change of degree in the character of a single composite mind. There need be no sudden shift from one mind having difficulties to two minds in uneasy co-operation, but only a gradual degeneration in the quality of interactions.

So an account of experiential combination would provide greater conceptual flexibility in treating these cases. Questions which are hard for explanatorily indivisible minds but intuitively

unproblematic for explanatorily divisible bodies become equally easy for minds if we make them explanatorily divisible.

Section 5: Fission and Fusion

The last section reviewed difficulties in synchronic individuation; this section shifts to difficulties in diachronic individuation in cases of ‘fission’ and ‘fusion’. In these cases, though we can count persons easily enough at the beginning and end of a certain process, we struggle to connect those counted before with those counted after. As before, experiential combination does not directly determine what to say about such cases, but it significantly impacts the costs and benefits of different options.

Subsection 5.1: Fission, Fusion, and the Simple Soul

Fission here means that relations supposedly sufficient for numerical identity hold between a single person at one time and two or more persons at later times; fusion means that such relations hold between a single person at one time and two or more persons at earlier times. I will sometimes just talk about fission for convenience, but everything said will apply also to fusion.

What counts as fission or fusion depends on the criteria for personal identity. For example, if having appropriately-caused apparent memories (what Shoemaker 1970 calls ‘quasi-memories’) of a past person’s experience is sufficient for being that person, a fission case need only involve two people who simultaneously quasi-remember the same past person’s experiences (see Williams 1956, and, according to Noonan 2003, pp.51-52, Leibniz 1981, p.245). But if identity requires sameness of *brain*, then producing apparent memories in some other brain is irrelevant, and a fission case must involve splitting a brain, so that two different people end up with enough neural tissue to qualify as having ‘the

same brain' as the same past person. A case like this, inspired by the split-brain syndrome, has become the standard example, appearing prominently in the work of Wiggins (1967), Shoemaker (1984), and Parfit (1971, 1984).¹⁴

A variety of accounts can be offered of how to analyse such cases, holding variously that the original person does not survive at all and thus should fear fission, that they survive as one or other resultant person, that they survive as both somehow, that they survive as both yet are identical to neither individually, or that they were never a single person to begin with. But all options come with some significant drawback. Thus it has been seen as an attraction of some theories, in particular those on which the person is a simple immaterial substance of some kind, that they need not allow fission cases to arise at all (cf. Noonan 2003, pp.15-17). On such views, defended by, e.g. Swinburne, 1997 and Chisholm 1976, persons are metaphysically indivisible, and such things as continuity of brain and memories are merely usual indicators of, but not constitutive of, the persistence of such persons. Thus while the relations that provide evidence of personal identity can be split, those constitutive of it are not, and there are no genuine fission or fusion cases.

But why is this a theoretical advantage? For even if human persons are metaphysically indivisible, and hence immune to fission and fusion, the same surely cannot be said for all things. Ships, rocks, axes, planets, droplets, waves, amoebae, bushes, and all the rest of the world's furniture are never claimed to share the indivisibility of persons, or to possess any deeply unified soul. For these objects, therefore, fission cases remain possible – indeed, in many cases actual. So what theoretical gain is there in precluding fission for persons? Why would this be a theoretically worthwhile move,

¹⁴ Some theories refuse to prioritise the cerebrum (which can easily be bisected), instead taking only continuity of brainstem, or of whole body, to be sufficient for personal identity (as on the 'animalist' views of Olson 1997 or Van Inwagen 1990). Such views still admit fission cases: the cerebrum is bisectable because there it has significant functional redundancy, and we might in principle artificially create such redundancy in the brainstem or whole body. Once the subject had been thus 'fattened' they could be divided; cases along these lines are described by Unger (1990, pp.287-290), drawing on Zuboff (1982).

unlike such risible expedients as claiming that only *Canadian* persons are metaphysically indivisible? Precluding fission for one particular subset of persons does nothing to *remove* the problem of fission, but merely restricts its scope slightly. By contrast, precluding fission for persons is often regarded as an effective (albeit costly) way of removing a problem.

The natural explanation is that the problem of fission for persons is not merely one instance of a general problem of fission for material things, but a distinctive problem which does not arise for material things. That is, persons (and by extension conscious subjects in general) differ essentially from other objects in a way that would make their fission, were it possible, distinctively problematic. The identity conditions of persons hold them to a higher standard than those for material objects.¹⁵ I think a major reason why fission and fusion of persons seems so problematic is Anti-Combination. When Anti-Combination is held together with Simplicity, so that persons are both explanatorily and metaphysically indivisible, no problem arises. But if one abandons Simplicity while retaining a tacit commitment to Anti-Combination, fission cases are possible yet incomprehensible.

Some authors say this explicitly. For instance, Zuboff 1990 says of a brain-chopping sort of case:

[W]e may easily think of the brains themselves in terms of fractions. Thus, though we might be a bit puzzled about [the identity of whole brains], if we like we can just fall back on talking about there being half of the original brain with you and half now over there with the other. But one could never talk about the subject or his experience like that. (p.41)

Similarly, Unger discusses the doctrine that “A subject is absolutely indivisible. A subject's body and brain certainly may be divided... But the subject himself cannot possibly be divided.”(1990, pp.41-42) Though he argues against this doctrine, he repeatedly attests to its intuitive pull on him.

¹⁵ To use a phrase introduced by Butler (1740/1860, p.325) and employed by Chisholm (1976, p.76 and elsewhere), persons persist in a ‘strict and philosophical sense’, while other objects persist in only a ‘loose and popular sense’.

Subsection 5.2: What makes for Unproblematic Fission?

The role of Anti-Combination in making person-fission seem less tractable than fission of other things is best seen by contrasting the significance of *extrinsic determination of identity* in each case. This is an issue that arises especially on the ‘no-branching’ approach to fission, according to which the original survives as neither of the resulting persons, and hence ceases to exist. Shoemaker, Parfit, Nozick, Van Inwagen, and others have all advanced something like this view. The problem is that this move allows for identity to be determined by ‘extrinsic’ factors, in a way which is intuitively troubling. That is (to adapt the formulation defended by Noonan, 2003, p.137), it allows for a single series of events to constitute either the continuous survival of a single person, or else the death of one person and the immediate production of another, depending on what happens somewhere else, to someone else who is not involved in those events.

For instance, in the brain-bisection case, if the left half-brain were destroyed and the right transplanted into a new body, the original person would survive in that body. But if the right half-brain were also transplanted into a new body, the original person would not survive, and someone else would wake up in the new body. The events involving the left half-brain have every intrinsic factor relevant to personal survival, but may not count if something else happens independently in another room. To many this has seemed inappropriate (see e.g. Williams 1956, 1970, Noonan 2003, Ch.7, 12).

Now contrast this with the case of something explanatorily divisible, like a rock (I take the following examples from Unger 1990, pp.164-166). Suppose we split it in half. Now we have two rocks, and intuitively neither is the original. Suppose we had taken the same rock, and sanded away precisely one half of it; intuitively the remaining portion is the original rock, now smaller, despite the fact that it is not intrinsically any different from one of the two fission-products in the first case, which

was not identical with the original. Paradox? Extrinsic determination of identity? While we may be uncertain what exactly to say of the original rock's fate, there is no deep intuitive difficulty. For we have a language available that can describe everything that happened in a non-paradoxical manner: we can say that the original rock was composed of two rocky halves.

When the rock is split, these two physical objects simply stop being physically connected as they had previously been; each persists through the process, and neither's identity depends on extrinsic factors. When we sand away half of it, one of these two physical objects is destroyed, while the other remains intact – again, the persistence of each depends entirely on changes in its intrinsic properties. It seems to me that these descriptions give rise to no profound metaphysical puzzlement: they are straightforward and perspicuous. They do leave some uncertainty about what to say has happened to original rock, but this is merely a question of how to describe events which we perfectly well understand. But the same does not hold true of person-fission, because knowing what happens to a person's parts seems to leave open the fate of the person themselves.

Moreover, there is a way to describe the whole rock's fate which does not involve any objectionable extrinsic determination of identity. We do this by saying that what has happened in the splitting case is that the rock has become a scattered object; it persists, but no longer counts as a single rock (since those must be all in one piece). It is now a pair (i.e. a mereological fusion) of rocks, a change in status concomitant to each half's change from the category of 'integrated portion of a rock' to that of 'rock'. Now, if we allow that the rock in the second case persists as its one remaining half, it remains true that the extrinsic fact of the other half's existence determines whether the remaining half is or is not identical to the original rock. But look at what is being determined extrinsically here: not whether that half-sized bit of stone is the original rock or is an entirely discrete object, but whether it is *all of* the original rock, or *merely one part of* it. But this is equivalent to whether or not there is *more of*

the original rock, a question we could never cogently have expected to be determined intrinsically. It obviously cannot be an intrinsic fact about one part of me whether there are any other discrete parts; that logically must depend on the existence of *other* things.

By saying that the rock persists not as a rock but as a pair of rocks, we have ‘domesticated’ the troubling kind of extrinsicness found in no-branching accounts of person-fission, by turning it into the natural and expected extrinsicness of whether there is more of something. We can do this because for the half-sized rocky object, the alternative to being identical with the original rock is not being an entirely different thing, but being part of it, and to that extent ‘retaining’ the intimate connection which, in the absence of other parts, would have sufficed for a judgement of identity. If combinationism is false, however, the alternative to ‘being the original person’ will be ‘being an entirely new person’, severing that intimate connection and making the extrinsic determination seem much more problematic.

Subsection 5.3: Consequences of Rejecting Anti-Combination

Rejecting anti-combination would not in itself resolve the issues around personal fission and fusion. But it would allow us to extend to persons whatever resolution we might think we have in the case of rocks and other divisible physical things, by adopting the ‘mereological approach’ to fission and fusion described in chapter 8, subsection 2.5. We might describe the two fission-products or fusion-ingredients as parts of the one person they came from or became, and thereby finesse the difficulty over extrinsic determination. We might say that a person survives fission as a pair of persons (and that fusion turns a pair into a person), or we might not say that but maintain that ‘what matters’ is preserved because their parts survive (and that fusion preserves ‘what matters’ because each person survives as part of a person). Or, of course, we might say that it matters little exactly which description

we employ, for the substantial truth is given by an account of what has befallen the parts of the persons involved, leaving indeterminate which person-level description best fits this.¹⁶

Obviously there are still significant questions and puzzles. For one thing, different accounts of what a person is will yield different accounts of what a part of a person is: if persons are conscious bodies, the relevant parts of them may be conscious organs, or conscious half-organs or sub-organs; if persons are essentially psychological beings defined by their memories, commitments, and personality traits, then the relevant parts of them may instead be sub-personas, clusters of drives and values which need not be realised in any specific organ. Another issue concerns the qualitative similarity of parts. If someone splits into two indiscernible copies, what parts of them survive as each copy? Conversely, if two people merge, the resultant person might display recognisable personality traits from each, allowing for at least some cogency in observationally identifying the ‘parts’ each has become. But it might not: they might be seamlessly blended, or the two original people might have been so intensely similar that whatever seems like an echo of one is equally an echo of the other. While I cannot here try to answer these questions, I think they should be taken as indicating fertile conceptual space for investigation.

Section 6: Problems of Multiplication

The previous sections considered abnormal cases in which it seems intuitively that something odd is happening to the subjects involved. By contrast, the arguments considered in this section involve perfectly normal and everyday situations. These arguments, drawn from the work of Trenton Merricks

¹⁶ Combinationism allows a sort of reductionism about persons which is similar to, but subtly different from, the reductionism defended by Parfit, who argues that all description in terms of subjects is a superficial gloss on a fundamental description in terms of experiences and bodies. Combinationism does not directly imply or preclude this, but does naturally suggest a position where the fundamental description is in terms of subjects of experience, just smaller simpler ones.

and Peter Unger, suggest that under standard materialist assumptions, any situation we would count as containing “one person” actually contains multiple overlapping persons. This is held to be sufficiently absurd to motivate revising or abandoning those standard materialist assumptions so as to avoid such multiplication. Significantly, both authors claim that multiplication of overlapping persons is absurd and objectionable, while multiplication of inanimate objects is not.

Subsection 6.1: How the Problems Arise

I distinguish three arguments of increasing complexity, each turning on divisibility: the first, divisibility into salient, functional parts like heads and brains; the second, divisibility into arbitrary parts like top halves; the third, divisibility into overlapping parts with equal claim to be ‘the whole’. In each case, three plausible premises seem to imply that these parts are conscious subjects in their own right, thinking and feeling simultaneously with the whole.

Those premises are: that the parts in question are intrinsically suitable to be conscious; that consciousness is an intrinsic property; that the effects of the other parts on the parts in question are not such as to harm, suppress, or disrupt its consciousness-generating activities. The first premise is supported by the observation that the parts in question would plausibly remain conscious if the other parts were removed, as long the right external support (supplies of nutrients, air, and water, etc.) were provided. The second and third premises are intuitive; it seems that I could be conscious even if everything I think about was in fact illusory, and it seems that what the neck, say, provides to the head is entirely beneficial, not disruptive. But if all three premises are true, it is hard to resist the conclusion that the parts in question are presently conscious.

The first sort of multiplication problem involves parts which ordinary language already recognises. For instance, Merricks claims that admitting the existence of brains leads to an “unacceptable multiplication of thinkers”:

[W]ithin the region occupied by a human organism there is a conscious human organism and a conscious brain. The brain is not identical with the organism; they differ in properties and parts. So... there are at least two conscious entities within that region... (2001, p.49)

The second sort of multiplication involves not just familiar, well-organised parts of things, but all the other parts they have, such as small parts (e.g. ‘my cells’), arbitrary sections (e.g. ‘my left half’), complements of more intuitive parts (e.g. ‘all of me except my left foot’), and maybe even scattered parts (e.g. ‘my left foot plus three cells from my right ear’). Plausibly, for many such parts, I could continue to exist, and to think, if I came to consist of nothing more than them – for instance, if I became ‘all of me except my left foot’, I would still be a conscious thinker.

The third form of multiplication, laid out canonically in Unger 1980 and sometimes called ‘the Problem of the Many’, involves all-but-indistinguishable entities which differ only in their ‘questionable’ microscopic parts. For instance, consider an atom which is only indeterminately part of me, perhaps just barely adhering to a skin cell. Such peripheral atoms seem to be neither determinately part of me, nor determinately outside me. Now consider the entity which contains all of me except this atom, and the entity which contains this atom as well. Not only do both seem to be conscious thinkers (by the argument above), they also both seem to be ‘equally strong candidates’ for being me – there is no non-arbitrary reason to privilege one as a person and relegate the other to being a mere person-like entity. Any plausible criterion will either admit both or exclude both, so how can we decide which of the two thinkers is ‘me’ and which is ‘most of me’?

From the above reflections Unger concludes that if persons are divisible material things, then either there are trillions of people where we thought there was only one, or there are none at all. Judging the first option absurd, he has at times either denied the existence of people, including himself (1979), or endorsed substance dualism (2004).

Merricks' response is more subtle, since he remains technically a substance monist. But in order to avoid admitting that "my thoughts are not mine alone, but shared" (p.49), Merricks denies the existence of brains, other internal organs, arbitrary sections, and all-but-one of the all-but-indiscernible entities involved in the Problem of the Many. My atomic parts exist, but they only compose one thing, which has special emergent conscious properties. Thus consciousness is intrinsic, but my foot-complement, were it to exist, would differ from me intrinsically, not just in lacking a foot but in lacking certain irreducible mental powers not determined by the microphysical facts. Merricks admits that we could never identify, from the outside, which precisely-specified set of atoms is the one which composes the person, and which sets compose nothing: thus although he is not a substance dualist, his solution, like Unger's, posits non-physical facts which uniquely attach to a particular being but cannot be inferred from even the fullest physical knowledge.

Subsection 6.2: The Tolerability of Object-Multiplication

Merricks and Unger both adopt views which avert the threatened multiplication of minds, but since neither extends these views to other things, they must still accept the multiplication of most material objects. This sort of multiplication is considered acceptable: but why?

Here is one answer: our practice of counting objects is fundamentally a tool for organising, guiding, and facilitating our perception and understanding of our surroundings. When I say "look at

that one table” or “there are three tables in this room”, what I should intend, and perhaps all I do intend, is to guide people’s attention to, and inform people about, the nearby matter and its suitability for resting things on. The difference between ‘two tables’ and ‘one table’ is not really a *numerical* difference, but a difference in the things that can be done with the matter in question – e.g. whether it can be placed in two separate rooms and remain useful. Since we are not really concerned with the numbers of things, we should not be worried by the multiplication of things. Resources for the development of this approach can be found in Lewis 1993. There he reviews several proposed responses to a more-or-less Ungerian¹⁷ problem, rejecting some and endorsing others, in particular endorsing the appeal to ‘partial identity’.

What is partial identity? Lewis, following Armstrong (1978) suggests that “the real opposite of identity is... not distinctness in the sense of non-identity, but rather distinctness in the sense of non-overlap”. Where two things overlap, they “are not entirely identical, not entirely distinct, but some of each.”(p.33) That is, the sharing of parts is treated as a partial form of identity. With a larger or smaller overlap, then, things can be said to be more or less identical.¹⁸ What is useful about this way of speaking is that it lets us interpret everyday talk of “two things” as meaning not “two *distinct* things” but “two wholly or largely *discrete* [i.e. non-overlapping] things”.

This offers a way of interpreting numerical statements about multipliable entities: rather than interpreting “there is one chair here” as “there is a chair here such that all chairs here are identical to that chair”, we should interpret it as “there is a chair there such that all chairs here significantly overlap with that chair.” Similar treatments will be possible for other numbers. Of course, how much overlap is ‘significant’ will be vague and context-sensitive. But that is to be expected: most of our descriptions of

¹⁷ Lewis ties together Unger’s discussion with a similar discussion by Peter Geach, 1967, which employs a similar problem to motivate a different view, namely revising the identity relation to make it sortal-relative.

¹⁸ It is not entirely clear what we should say about a thing and one of its proper parts (am I more identical with a given 10% of myself than are the bodies of two conjoined twins who overlap 50%?) but we may set that aside.

things are similarly vague and context-sensitive. For crowds, a bit of overlap is par for the course; for chairs, absolute discreteness is the expected standard. The important thing is that the apparent conflict between our everyday numerical judgements and the arguments for multiplication is resolved by replacing judgements of *distinctness* with judgements of *discreteness*.¹⁹

Subsection 6.3: The Difference between Persons and Objects

If we can learn to live with widespread multiplication of objects, why not with multiplication of persons too? Merricks and Unger both deny that persons are on a par with objects in this regard: it is this difference between multiplication of inanimate objects (tolerable) and multiplication of conscious subjects (intolerable) that drives them to posit a profound ontological divide between the two. But why is multiplication of conscious subjects specifically so absurd?

One argument is that if subjects were ‘multipliable’, our self-knowledge would be undermined, since we could not know *which* of the many subjects we were. I do not know, for instance, if people really refer to *me* with their words and thoughts about ‘Luke’, or to some overlapping person. Merricks claims that “there seems to be no way to tell whether one is the ‘maximal’ object... or just one of the countless wannabes.”(2001, p.103; similar points are made by Unger, 1980, p.461-462). In chapter 7 I argued that while there is something to this concern about self-identification, but that given certain theses about the metaphysics and semantics of indexical reference, and the ethics of self-interest, a limited inability to self-identify is unproblematic.

But even beyond these epistemological concerns, both authors display sheer intuitive revulsion at multiplication of persons. The repugnance of this conclusion is arguably at the heart of Merricks

¹⁹ I am using partial identity and vagueness in a slightly different way from Lewis, who suggests we “count by almost-identity”, supplemented by vague reference, rather than counting by “insignificant discreteness” with ‘significant’ treated as a vague predicate. But the spirit of the proposals is, I think, very similar.

2001, where he describes it as “simply incredible”(p.95). Indeed, the need to avoid positing “a mighty host of conscious, reflective, pain- and pleasure-feeling objects now sitting in my chair”(2001, p.95) is his primary argument for consciousness being something over and above the microphysical. But he rejects analogous arguments about inanimate physical objects, because “unlike the claim that there are two conscious beings wearing my shirt, the claim that there are two objects upon the pedestal [where we would normally say ‘a statue stands’]... is not absurd.”(p.106)

Similar judgements of relative absurdity appear in Unger’s later work²⁰. Discussing a situation that we would normally describe as ‘one person eating chocolate’, he says:

The thought that there are, in my situation, many individuals all experiencing is far more disturbing, to my mind, than the thought that, in this situation, there are many entities each chewing, or many each digesting. With the digesting of some chocolate candy, say, there seems little more at stake than just a matter of our choosing what forms of words to use... With an experiencing being, by contrast, matters seem very different. (2006, pp.378-379)

Subsection 6.4: The Role of Anti-Combination - ‘Additionalism’ about Experience

I believe that a large part of this differential absurdity stems from Anti-Combination. Because the doctrines that can make multiplication tolerable turn on partial identity, they cannot be extended straightforwardly to explanatorily indivisible persons. The proposal that we count things by discreteness, not distinctness, made sense because two things which share 99.9% of their parts are, to all intents and purposes, the same thing – the shared parts, and their powers, are what interest us, not what they are grouped with. To put it another way, the second thing is not really anything ‘*in addition to*’ the first – or rather, all that is additional is the 0.1%, which is negligible. But if conscious minds are

²⁰ Unger’s views have changed substantially over his career – in his early work (e.g. 1979a, 1979b, 1980) he is a nihilist, in his middle-period work (e.g. 1990) a common-sense-physicalist, and in his later work (e.g. 2004) a substance dualist.

explanatorily indivisible, so that the consciousness of a whole cannot be reductively analysed into the consciousness of its parts, then even when two subjects share 99.9% of their physical parts, all the consciousness of the second is something entirely new, not to be explained in terms of a negligible addition to the already-recognised consciousness of the shared portion. Its experiential features – its perspective, its phenomenal field, its stream of consciousness – are entirely ‘in addition to’ those of the first subject. Call this ‘Additionalism’ about experience.

Given Additionalism, we cannot allow that I and a trillion other subjects can be ‘counted as one’ because our streams of consciousness ‘overlap almost entirely’, or that we should treat our talk of ‘persons’ as just a device for talking about the mental parts, or mental ‘stuff’, that distinct persons comprise, and may almost entirely share. This is what makes multiplication of thinkers seem intolerable: each additional thinker must be regarded as a whole new person, implausibly squeezed into the same head, rather than as just a new and slightly different way of drawing a boundary around the same mental reality.

Of course, if Experiential Simplicity is true, then the problems do not arise in the first place, just as if Experiential Simplicity is true, then fission and fusion cases for persons cannot arise. If we hold fast to there being an important difference between conscious subjects and inanimate objects in all these cases, perhaps what this should teach us is that conscious subjects are not divisible things. What is difficult, it appears, is to accept Anti-Combination while denying Experiential Simplicity (a lesson suggested also by the discussions in sections 1 and 7).

Additionalism flows from Anti-Combination. Undermining Anti-Combination will therefore undermine Additionalism, and with it much of the intuitive support for Unger’s and Merricks’ arguments. The problem of self-identification is not directly removed by denying Additionalism, but it is removed by the anti-individualist proposal of chapter 7. Thus a successful defence of

combinationism removes the absurdity of multiplication of subjects. It does not necessarily make such multiplication wholly unproblematic, but it can show it to be no *more* absurd than multiplication of physical things in general.

Section 7: Anti-Nesting Among Materialists

In section 1 I discussed a historically popular argument deriving Experiential Simplicity from Anti-Combination. In contemporary philosophy of mind, explicit defences of Simplicity are rare²¹, but there is some debate over the weaker principle I have called ‘Anti-Nesting’: that while a subject of experience may be composite, its parts cannot themselves be subjects. Anti-Nesting is defended by Putnam (1965) and Tononi (2012), and criticised by Block (1978) and Schwitzgebel (2014).

Putnam’s influential formulation of functionalism stipulates, in his suggested schematic definition of ‘pain’, that “No organism capable of feeling pain possesses a decomposition into parts which separately [satisfy this definition]” (1965/2003, p.215), this stipulation being intended “to rule out such ‘organisms’... as swarms of bees as single pain-feelers.” (p.216) But, as Schwitzgebel notes, Putnam “doesn’t explain why this possibility is absurd for actual swarms of bees, much less [for any] possible future evolutionary development of a swarm of conscious bees...” (2014, p.11)

A more developed defence of Anti-Nesting comes in Tononi’s explanation of what he calls ‘the exclusion principle’, part of his ‘Information Integration Theory of Consciousness’. Consciousness is integrated information, but not all integrated information is conscious; if a system is contained within a

²¹ For exceptions, see Chisholm 1991, Swinburne 1984, and Lowe 1996. In each case, the defence is equivocal, cautious, or heavily qualified, and the reasons often draw heavily on either the historical arguments covered in the previous section, or on considerations about personal identity that were covered in section 6.

supersystem with equal or greater informational integration, or contains a subsystem with greater, it cannot be conscious.

Subsection 7.1: Anti-Nesting, Anti-Combination, and Parsimony

Tononi's defence of this principle involves two major themes: phenomenology and parsimony. First, he appeals to the phenomenological claim that "experience is exclusive" (p.59), already discussed in chapter 4 section 1. I argued there that any plausible sense of the 'boundedness' of experience fails to provide an argument for Anti-Combination or Anti-Nesting. Physical objects can overlap and contain each other, but also have definite borders, and can be said to 'encompass what they do, and nothing more.' And the fact that we lack conscious access to a given process hardly entails that the process is not itself conscious – clearly it could be conscious in another mind, and even for our own mind there might be processes that are phenomenally conscious but not access-conscious. The argument makes sense only with the unmotivated premise that were two consciousness things to compose a third, each would be conscious of everything the other was conscious of.

Tononi's other justification for the Exclusion Principle is that to allow nesting or overlapping consciousnesses would be unparsimonious, since given the 'most conscious' complex (the one with the most integrated information), we gain nothing explanatorily by positing additional, 'less conscious', complexes.²² But (as Schwitzgebel points out) it is not obvious how parsimony considerations apply to part-whole relations. Most wholes are supposed to have more-or-less the same causal powers as their

²² Tononi claims that "We recognize this principle [of parsimony] intuitively when we talk to each other: most people would assume that there are just two consciousness (complexes of [high informational integration]) that interact a little, and not also a third consciousness (complex of [lower informational integration]) that includes both speakers."(pp.67-68) But this is unconvincing, since Exclusion is only needed to rule out such a third consciousness if we impose no lower limit on how much informational integration is needed for consciousness. Tononi's theory imposes no such lower limit, departing starkly from common sense. So Tononi is claiming that common sense recognises Exclusion, when Exclusion is only needed to avoid a consequence which common sense would never have predicted.

parts, but taking this to be a case of causal competition or ‘overdetermination’ (enough to motivating denying the existence of one or the other) is a significant departure from common sense.

In fact, I believe that parsimony-based arguments for Anti-Nesting rely on Anti-Combination: it is Anti-Combination which makes experiential nesting seem unparsimonious. If positing experiential parts does not in itself explain the experiential whole, then the latter, being unexplained, must be a further posit. That is, it counts against parsimony just as it would had we not posited the parts. But then if (as is likely) the explanatory power of the parts largely pre-empts that of the whole (they cause the same effects, rationalise the same actions, etc.), having both will seem a needless multiplication of entities. This is just another manifestation of ‘Additionalism’ about experience, as described in the previous section: given Anti-Combination, each mind is an additional posit.

Conversely, if experiential combination is possible, then the whole is no additional posit at all: ‘positing’ the whole is merely drawing out the consequences of already-positing parts and their relations. There is then no offence against parsimony, any more than in accepting the existence of macroscopic physical objects, having already accepted the existence of their component atoms.²³ Indeed, it may well be that ruling out Nesting requires offending against ‘theoretical simplicity’, i.e. introducing additional independent stipulations (such as Tononi’s Exclusion Principle) into one’s basic theory. It may be ‘cheaper’ to allow for a profusion of individuals, than to complicate the theory to exclude them (Cf. chapter 2 section 2).

²³ Merricks argues that even this is unparsimonious: macroscopic physical objects are superfluous given their parts, and we should deny their existence. This reflects the fact that Merricks holds what chapter 1 called an ‘autonomous levels’ view of composition: the ‘wholes’ he speaks of are something over and above their parts, and their parts are something under and below them. I am assuming that parts and wholes do not generally compete (causally, explanatorily, or in terms of parsimony), and enquiring whether *experiential* parts and wholes can meet this standard: Merricks thinks parts and wholes always compete, and is to that extent outside my intended audience.

Subsection 7.2: Arguments Against Anti-Nesting

Block and Schwitzgebel both offer thought-experiments meant to undermine whatever intuitive appeal Anti-Nesting may have²⁴. Block's thought-experiment has already been noted in relation to the problem of 'dancing qualia' in chapter 5 subsection 3.3: suppose that a normal human being, just by eating and breathing, absorbed microscopic conscious beings into their body and brain: surely they would remain a conscious being, in defiance of Anti-Nesting.

Schwitzgebel describes beings he calls "Antarean Antheads", outwardly elephantine aliens from a planet near Antares whose 'brains' are in fact a swarm of minute insects, individually conscious but of rudimentary intelligence, living inside a 'mobile hive' which their aggregate wriggings and scent signals control intelligently just as our bodies are intelligently controlled by the aggregate synaptic firings of our trillion neurones. He suggests that for humans to re-interpret the Antheads' actions and utterances (in which they come across as "sanitary, friendly, and excellent conversationalists", 2014, p.9) as non-conscious and non-intelligent when the mechanisms responsible were discovered, would be grossly implausible, not to mention an offensive sort of human chauvinism.

In both thought-experiments, a certain sort of leverage is exerted: granting that we are materialists, and reject Simplicity, we will attribute consciousness to humans and aliens made out of ordinary matter. Given that, would it not be absurd to completely reverse that attribution because of the additional presence of consciousness in some parts of that matter? This strategy will fail against an opponent who accepts Experiential Simplicity, since they will not grant the comparison. But it seems to show that accepting Anti-Nesting while denying Experiential Simplicity is an uncomfortable middle

²⁴ As Block points out, crude versions of Anti-Nesting might admit of even more direct refutation, if *any* conscious part is enough to deprive the whole of consciousness: for then the consciousness of anyone with a sentient internal parasite, or carrying a sentient foetus, will be in question. The elaborate thought-experiments target some more sophisticated version, which specifies that the conscious parts must not "play a crucial role" in the consciousness of the whole (Block 1992 p.77).

position. Moreover, since I argued in subsection 7.1 that the only compelling argument for Anti-Nesting relies on Anti-Combination, this any difficulty in maintaining Anti-Nesting without Simplicity will mean a difficulty in maintaining Anti-Combination without Simplicity.

Section 8: Collective Mentality

In this section I discuss *collective mentality*, arguing that Anti-Combination contributes significantly to a general aversion to collective phenomenal consciousness even among those who accept other forms of collective mentality.

Philosophical discussion of collective mentality takes off from the widespread tendency in natural language for the behaviour of social collectives to be described and explained in mentalistic terms. Both populations and institutions can be spoken of as believing things, wanting things, or doing things in order to attain given ends (as can small groups like ‘Tim and Bob’). The initial question for philosophers is how to interpret this language – whether it expresses commitment to a distinctive phenomenon, or is merely a figure of speech. The deflationary attitude is expressed by Quinton:

Groups are said to have beliefs, emotions, and attitudes and to take decisions and make promises. But these ways of speaking are plainly metaphorical. To ascribe mental predicates to a group is always an indirect way of ascribing such predicates to its members... To say that the industrial working class is determined to resist anti-trade union laws is to say that all or most industrial workers are so minded. (1975, p.17)

Certainly it seems that much mentalising talk of collectives fits this ‘distributive’ model, where the group is said to be P simply because all or most members are P. And even some cases where the very same predicates do not apply to both group and members may admit of an equally deflationary construal if they apply merely ‘as if’ predicates. To say that an institution behaves *as if* it wanted a

certain outcome need carry no more commitment to mentality than to say that evolution operates as if it wants organisms to be better adapted to their environments.

However, it is far from clear why all mentalising talk about groups should be taken in a deflationary way. Consider cases where a stable, highly-organised institution extracts information about some matter, processes that information through a series of internal events, and uses it to guide institutional policy. There is certainly no necessity that any single member of the group should be apprised of all the information, and privy to all the deliberations, and also involved in all of the executions, and yet the overall process seems like it might merit literal mentalistic description. Consider, moreover, hypothetical advanced societies, which might enhance the speed, reliability, and density of communication channels greatly, so as to remedy any deficits in that regard with present day social groups.

When we focus on these kind of highly-organised groups, their mere being-composed-of-mental-parts becomes more and more salient. Insofar as this status makes us feel that their mentality *must* be merely figurative, it is reasonable to suspect the role of one of the Mereological Denials.

Subsection 8.1: Collective Intentionality and Collective Consciousness

While some philosophers, like Quinton, have been dismissive of collective mentality, others have devoted much theoretical effort to analysing it. These analyses sometimes involve commitment to genuine, though reducible, collective intentionality, though different accounts locate the distinctiveness of collective intentionality in different aspects of intentionality: some in the *content* of what each individual member intends (Bratman 1999), some in the *mode* of intentionality directed on that content

(Searle 1990), and some in the possession of the intentional state by a collective *subject* (Velleman 1997). And different accounts disagree on whether collective intentionality, however construed, can be reduced to individual intentionality – i.e. intentionality which is individual in content, mode, and subject. Depending on analysis, collective intentional states may be a special type of intentional state possessed by an individual, intentional states each token of which is shared, i.e. had collectively, or an intentional state belonging to a collective.

Collective intentions have received the most attention, but people also ascribe collective beliefs, desires, and so on – i.e. all the cognitive and conative apparatus of agency. Pettit & List 2011, for instance, argue at length for recognising group agents who hold beliefs and desires which rationalise their actions, doing so in a way that is in practice, but not in principle, irreducible to the beliefs, desires, and actions of individual members. They claim that this locates them between Hegel (their emblem of wholly irreducible, strongly emergent, collective agents) and Hobbes (for whom collective agents reduce straightforwardly to individual agents authorised by others).

However, genuine conscious experience in social groups is widely rejected by contemporary philosophers. Even those authors who go furthest in defending a range of genuine collective mental states stop short of collective consciousness. For example, Gilbert (2002) and Huebner (2011) both argue in support of collective emotions, but do so in large part by trying to break the link between emotion and consciousness, arguing that genuine emotions may be devoid of phenomenology. Thus they prevent “the implausibility of collective consciousness... impugn[ing] the possibility of collective emotions.”(Huebner 2011, p.102) Similarly, Pettit and List, while defending the genuine existence of collective agents, attribute no phenomenology to these agents.

Subsection 8.2: Why Deny Collective Consciousness?

Why is it so obvious that social groups are not conscious? Usually this is taken as unargued common-sense, or a basic intuition. But it is actually hard to see what could justify such an intuitive belief. When we believe that something lacks consciousness, this is usually because its behaviour does not display the kind of flexible, intelligent, goal-seeking that characterises the behaviour of higher animals. But in favourable cases, social groups can display just such a pattern of behaviour, so something must be defeating this *prima facie* evidence of consciousness. Sometimes we discount apparent evidence of consciousness when we discover that the mechanisms that produced it are actually structured in an insensitive or inflexible way, aiming merely to simulate intelligence, as in computer programs like ‘cleverbot.com’, which are designed to give a superficial impression of being a conversational partner by deploying a series of stock phrases (cf. subsection 2.3 of this appendix). But this defeater is not present in the social case, for social groups often seem to gather, integrate, and deploy information about their surroundings in a flexible, responsive, open-ended way.

Alternatively, the defeater might be some difference of degree between the processing going on in a social group and that going on in a human or animal brain: for instance, the signals among its components travel slower, or need to go further, or carry less information. Schwitzgebel argues against this, suggesting that large, institutionally dense nation-states like the USA are at least on a par with ordinary mammals in informational complexity – in the relevant respect, “a rabbit brain is not clearly more sophisticated”(2014, pp.20-21). Moreover, even if a difference in degree of sophistication justified denying that collective consciousness is *actual* – because all presently-existing social groups are less sophisticated in their structure than a rabbit’s brain – it would not justify denying that collective consciousness is *possible*, since technological or social transformation might make up this deficit. And it is also not clear that it justifies the claim that collective consciousness is *obviously* and *definitely* not actual, i.e. that supposing it is absurd.

Subsection 8.3: The Role of Experiential Combination - 'Additionalism' about Experience

I believe that a major contributor to the widespread denial of collective consciousness is Additionalism about conscious experience, which is a consequence of Anti-Combination. In the words of Searle, "Most empirically minded philosophers think that [collective mentality] must reduce to individual [mentality]." (1990, p.404) There is a wish to avoid irreducible, strongly-emergent group minds which stand above individual minds; such beings are regarded as mysterious, unparsimonious, and even ethically threatening. This may allow for groups to have genuine intentionality, if that intentionality can be reductively explained through their parts. But it does not allow for collective consciousness, because consciousness in a whole is always something over and above consciousness in its parts. For instance, Petit & List 2011 devotes much space to distancing themselves from the 'animation theory', on which "group agency [is] the product of an... organicist force" which is "equally [as] mysterious" as the vitalist's explanation of life (p.9); they take a sort of 'weak emergentist' position, on which group agency is a distinctively useful way of regarding social groups, which enables explanations which would otherwise be in practice impossible, or at least very hard, but involves no new supra-personal forces.

Yet if Anti-Combination is false, then this conflict between reducibility and genuine collective consciousness disappears: social groups can be nothing over and above their members, while still being conscious in just the same way that individual members are. Of course, the detailed analysis of how collective consciousness arises from individual consciousness will be complex, and probably admit of as disagreement as there has been over the right analysis of collective intentionality. And combinationism would not necessarily entail that collective consciousness is ubiquitous or even actual; it merely shows that the possibility is not absurd. In doing so it serves also to underwrite the application

of social analogies to the structure of the human mind, allowing us not only to illuminate the social world by reference to our inner experience, but also to illuminate the world of neural signals by reference to our experience of communication, coordination, and cooperation.

Section 9: Nihilism and Non-Singular Mentality

Since experiential combination involves parts and wholes, its scope of application depends on which things compose wholes. Those who hold ‘universalism’ about composition will have many more wholes composed of, and composing, conscious subjects, and so may approach the question of experiential combination with different theoretical needs from those who hold that composition is restricted in some manner, so that some sets of things do, while others do not, compose wholes.

But one view of composition has an especially pressing need for experiential combination. In the view of both its proponents and its detractors, mereological nihilism, the view that no sets of things compose wholes, demands either physicalism or something like combinationism in order to provide a plausible account of consciousness.

Subsection 9.1: An Argument Against Nihilism

Nihilist theories generally seek to reconstrue all our talk about composite objects so as to quantify only over simple entities arranged in and behaving in various ways. But even philosophers very close to being nihilists, who develop and defend such reconstructions for our talk of non-conscious entities, doubt that any such ‘paraphrase’ or ‘translation’ is possible for talk of conscious properties.

For instance, Peter Van Inwagen (1990) and Trenton Merricks (2001) both assert that mental activities cannot be performed ‘cooperatively’ by multiple agents, while physical activities can. When a baseball shatters a window, Merricks says, “every atom arranged baseballwise causes something, and when what one of them causes is added to what each of the others causes, the ‘sum’ is the shattering of the window”(p.111) Physical effects can be seen as the aggregate effect of many microscopic activities of physical parts. Yet “it does not seem that... when what one of my atoms does is added to what each of the others does, the ‘sum’ is my consciously deciding.”(p.111)

Similarly, Van Inwagen says that where we thought there was a shelf supporting books, or the sun shining, there are in fact “simples... arranged shelfwise [that] cooperate to support weight” and “simples... arranged siderially [that] cooperate to produce light”(p.118). Yet “I do not see how we can regard thinking as a mere cooperative activity... surely, planning for tomorrow or feeling pain cannot be activities that a lot of simples can perform collectively” (p.118) Mental activities demand a singular subject, which must be either composite or simple, and if simple must plausibly be immaterial (cf. the first footnote to section 1). Since both philosophers reject substance dualism for independent reasons, they conclude that there must be at least some composites, namely conscious composite organisms.

Sider, defends nihilism against this allegation, but does so essentially by registering being unpersuaded Van Inwagen’s somewhat dogmatic presentation of the disanalogy between mental and physical activities:

Perhaps van Inwagen’s belief in metaphysical singularity has something to do with the character of conscious experience? A subject’s simultaneous experiences are experienced by that subject as being in some sense part of *one* conscious episode, and as experienced by a single subject. But it is unclear why these aspects of phenomenology could not be due, metaphysically, to states of particles. (2012, p.27)

I believe that there are two issues driving this dispute. One is whether consciousness and other mental phenomena can be reduced to a non-mental bases; the other is Anti-Combination. Merricks and Van Inwagen accept both, and so must reject any nihilistic account of human consciousness. A non-combinationist who was also a physicalist could consistently be a nihilist, analysing facts about consciousness into facts about the physical properties of and relations among fundamental particles. But a non-physicalist nihilist must, I think, reject Anti-Combination, and so my inquiry into grounds for Anti-Combination will bear on whether there is a sound argument against mereological nihilism here. If combinationism is true, one major reason for rejecting nihilism disappears.

Subsection 9.2: Are Nihilism and Combinationism Compatible?

It may at first seem strange that combinationism, a doctrine making reference explicitly to parts and wholes, should serve to defend nihilism, the explicit denial of parthood relations. But the type of relations which combinationism claims between parts and wholes are precisely those relations which the nihilist is able to give their eliminative account of: if all the facts about mental wholes are fully explicable through facts about their parts, then nothing is lost if the former facts are eliminated in favour of sole reliance on the latter. If the parts, so to speak, do all the work, then cutting out the whole becomes defensible if we have independent motivations (e.g. parsimony) for doing so.

To clarify, distinguish two (exclusive but not exhaustive) theses:

Non-Singular Mentality: all mental properties are such that they may be possessed collectively by many things, without being possessed by any single thing.

This contrasts with:

Singular Mentality: all mental properties are such that they require a genuinely singular thing to bear them (regardless of what else bears them).

I take it that all nihilists who are also naturalists are committed to Non-Singular Mentality, since they attribute experiential properties to the collections of particles which we would normally describe as human beings. If experiential properties are instantiated at all, then naturalistic nihilists must attribute them to groups.

Now consider ‘combinationism_N’, which I define as follows:

Experiential Combinationism_N: The experiential properties of a jointly-conscious collection of things may be explained by the experiential properties of, and relations among, those things.

This is just normal experiential combinationism with the mereological terms replaced with nihilist-friendly stand-ins (indeed, if composition is in some sense identity, then combinationism_N is in some sense equivalent to combinationism). My claim is that nihilists who do not wish to identify each of us with a simple substance should be committed not just to Collective Experiences, but also to combinationism_N. This is because their general strategy of reconstruing apparent talk of composites in terms of simples will translate combinationism into combinationism_N, while being unable to thus translate other, non-combinationalist, views of consciousness like Van Inwagen’s and Merrick’s.

For instance, if we fleshed out combinationism with some claim like:

A group of subjects compose a composite subject if and only if their experiences all stand in the A-relation to the experiences of the others.

Then the nihilist could flesh out combinationism_N with the following claim:

There are no composite subjects - there are just micro-subjects arranged subject-wise.

Whenever we think that an composite subject (like a human being) exists, we are misapprehending what is really true, namely that there is a set of micro-subjects which have experiences and all of whose experiences stand in the A-relation to those of the others.

If they then stipulated the a definition in terms of A-relatedness for the phrase ‘are arranged subject-wise’, they might express this by saying that there are no composite subjects, there are just simples arranged subject-wise. They could then maintain that *if* the original combinationist analysis is

correct, then their revisionary-seeming claims like the one above are actually innocuous, since the things they eliminate were nothing over and above the things they keep.

Summary:

If I am correct in the interpretive and dialectical claims made in this appendix, then Anti-Combination, the denial of experiential combination, has a huge impact on the way we think about minds. Anti-Combination is, for a start, crucially involved in two different arguments against materialism, the historically popular ‘Achilles’ argument, and the more recent argument from the need to avoid multiplication of subjects. Moreover, Anti-Combination determines whether we can reasonably accept collective consciousness, or consciousness in psychofunctional homuncular systems, or consciousness in organisms like a swarm of bees. It makes a number of problem cases profoundly less tractable for persons than for inanimate objects, and it dooms panpsychism as an explanatory project.

Of course I am probably not entirely correct in all these claims, and at each point there are potentially other ways to defuse or circumvent any consequence of Anti-Combination. But hopefully I have said enough to persuade the reader that the defence of combinationism that I have offered is important to our understanding of many disparate issues in the metaphysics of mind.

Bibliography

- Alter, T., and Nagasawa, Y., (2012). *What is Russellian Monism?* *Journal of Consciousness Studies* 19, pp. 67-95.
- Anzai, A., Peng, X., Van Essen, D. (2007). *Neurons in monkey visual area V2 encode combinations of orientations*. *Nature Neuroscience* V10:10, pp.1313-1321
- Aristotle (2007). *Nicomachean Ethics*. Original publication ~350 BCE, 1911 translation by D. P. Chase. Cosimo Inc.
- Armstrong, D. M. (1963). *Is Introspective Knowledge Incorrigible?* *The Philosophical Review*, Vol. 72, No. 4, pp. 417-432
- Armstrong, D. M. (1978). *Universals and Scientific Realism*. 2 vols. Cambridge University Press.
- Avicenna (1952). *The Book of Healing*. Originally published 1027, in Rahman, F. (trans.) *Avicenna's Psychology*. Oxford, Oxford University Press.
- Ayer, A. J. (1940). *The Foundations of Empirical Knowledge*. New York: Macmillan.
- Baars, B. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press
- Baillargeon, R., & De Vos, J. (1991). *Object permanence in young infants: further evidence*. *Child Development* 62:6, pp.1227-1246
- Barnes, E., and Williams, J. R. G., (2011). *A Theory of Metaphysical Indeterminacy*. In Bennett, K., and Zimmerman, D. (eds.), *Oxford Studies in Metaphysics Volume 6*. Oxford University Press, pp.103-148
- Barnett, D. (2008). *The Simplicity Intuition and Its Hidden Influence on Philosophy of Mind*. *Nous* 42:2, pp.308-335
- Basile, P. (2010). *It Must be True – But How Can it Be? Some Remarks on Panpsychism and Mental Composition*. *Royal Institute of Philosophy Supplement* 67, pp.93-112
- Baxter, D. (1988). *Many-One Identity*. *Philosophical Papers* V17:3, pp.193-216
- Baxter, D. and Cotnoir, A. (eds.) (forthcoming). *Composition as Identity*. Oxford University Press
- Bayle, P. (1991). *The Historical and Critical Dictionary*. Original Publication 1697, Selections trans. and introduced by Richard Popkin. Indianapolis: Hackett
- Bayne, T. (2010). *The Unity of Consciousness*. Oxford: Oxford University Press
- Bayne, T. & Chalmers, D. (2003). *What is the Unity of Consciousness?* in A. Cleeremans (ed.), *The Unity of Consciousness: Binding, Integration, Dissociation*, Oxford: Oxford University Press, pp.23-58, accessed online 2014 at <http://consc.net/papers/unity.pdf>
- Bennett, K. Material coincidence and the grounding problem. *Philosophical Studies*, 118(3):339-371, 2004.
- Bennett, K. (2011). *Having a Part Twice Over*. *Australasian Journal of Philosophy*, pp.1-21
- Bennett, K. (forthcoming). *'Perfectly Understood, Unproblematic, and Certain': Lewis on Mereology*. In *Blackwell Companion to David Lewis*, edited by Barry Loewer and Jonathan Schaffer
- Berkeley, G. (2008). *A Treatise Concerning the Principles of Human Knowledge*. (Originally published 1710), Arc Manor LLC
- Bhaskar, R. (1978) *A Realist Theory of Science*, 2nd ed. Brighton: The Harvester Press
- Bhattacharya, J., and Petsche, H. (2005). *Phase synchrony analysis of EEG during music perception reveals changes in functional connectivity due to musical expertise*. *Signal Processing*, Volume 85:11, pp. 2161-2177
- Blakemore, S., Wolpert M., and Frith, D. (2002) Abnormalities in the awareness of action. *Trends in Cognitive Sciences* Vol.6 No.6 237-242
- Blamauer, M. (2011). *Taking the Hard Problem of Consciousness Seriously: Dualism, Panpsychism and the Origin of the Combination Problem*. in M. Blamauer (ed.) *The Mental as Fundamental. New Perspectives on Panpsychism*. Ontos

- Block, N. (1992). *Troubles with Functionalism*. Originally Savage, C. W. (ed.), *Perception and Cognition: Issues in the Foundations of Psychology*, University of Minneapolis Press, 1978. Reprinted with revisions in Beakley, B., and Ludlow, P. (eds.), *The Philosophy of Mind: Classical Problems/Contemporary Issues*. MIT Press.
- Block, N. (1981). *Psychologism and Behaviourism*. *The Philosophical Review* V90:1, pp.5-43
- Block, N. (1995). *On a Confusion about a Function of Consciousness*. *Behavioral and Brain Sciences* V18:2, pp.227-287
- Block, N. & Stalnaker, R. (1999). *Conceptual Analysis, Dualism and the Explanatory Gap*. *The Philosophical Review* V108, pp.1-46
- Block, N. (2003). *Mental Paint*. In *Reflections and Replies: Essays on the Philosophy of Tyler Burge* edited by Martin Hahn and Bjorn Ramberg, MIT Press
- Block, N. (2005). *Two Neural Correlates of Consciousness*. *Trends in Cognitive Sciences* Vol.9 No.2, pp.46-52
- Block, N. (2007). *Overflow, Access, and Attention*. *Behavioral and Brain Sciences*, V30:5-6, pp.530-548
- Block, N. (2011). *Perceptual consciousness overflows cognitive access*. *Trends in Cognitive Sciences* xx, pp.1-9
- Bower, T.G.R. (1974). *Development in infancy*. San Francisco: Freeman
- Brentano, F. (1987). *The Existence of God: Lectures given at the Universities of Wozzburg and Vienna, 1868-1891*. ed. and trans. Krantz, S., Nijhoff International Philosophy Series
- Briscoe, R. (2011). *Mental Imagery and the Varieties of Amodal Perception*. *Pacific Philosophical Quarterly* V92:2, pp.153-173
- Broad, C. D. (1925). *The Mind and its Place in Nature*. London: Kegan Paul
- Broad, C. D. (1953). *Self and Others*, Herbert Spencer Lecture, Oxford University. Reprinted in *Broad's Critical Essays in Moral Philosophy*, ed. David R. Cheney, 1971.
- Burge, T. (2000). *Reason and the First-Person*. In *Knowing Our Own Minds*, Crispin Wright, Barry C. Smith, and Cynthia Macdonald, Oxford University Press, pp.243-271
- Burke, M. (1994). *Dion and Theon: An Essentialist Solution to an Ancient Puzzle*. *The Journal of Philosophy* V91:3, pp.129-139
- Butler, J. (1860). *The Analogy of Religion, Natural and Revealed, to the Constitution and Course of Nature*. Originally published 1736, edited, with an analysis, by J.T. Champlin. Boston: J. P. Jowett and Co.
- Cameron, R. (2011). *Composition as Identity Doesn't Settle the Special Composition Question*. *Philosophy and Phenomenological Research*, V84:3, pp.531-554
- Campbell, J. (1999). *Schizophrenia, the Space of Reasons, and Thinking as a Motor Process*. *The Monist* V82:4, pp.609-625
- Cartwright, N. (1980). *Do the Laws of Physics State the Facts?* *Pacific Philosophical Quarterly* 61, pp.75-84.
- Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Oxford University Press
- Carruthers, G. (2012). *The case for the comparator model as an explanation of the sense of agency and its breakdowns*. *Consciousness and Cognition*, 21, 32-47.
- Carruthers, P. (2004). *Reductive explanation and the 'explanatory gap'*. *Canadian Journal of Philosophy* V34:2, pp.153-174
- Carruthers, P., & Schechter, E. (2006). *Can panpsychism bridge the explanatory gap?* *Journal of Consciousness Studies*, 13(10-11), 32-39
- Chalmers, A. (1987). *Bhaskar, Cartwright and Realism in Physics*. *Methodology and Science* 20, pp.77-96
- Chalmers, D. (1995a). *Facing Up to the Problem of Consciousness*. *Journal of Consciousness Studies*, V2:3, pp.200-219
- Chalmers, D. (1995b). *Absent Qualia, Fading Qualia, Dancing Qualia*. In T. Metzinger (ed.), *Conscious Experience*, Imprint Academic, pp.309-328

- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- Chalmers, D. (2003). *Consciousness and its Place in Nature*. In S. Stich and F. Warfield (eds) *Blackwell Guide to the Philosophy of Mind*, Blackwell
- Chalmers, D. (2006). *Strong and weak emergence*. In P. Clayton and P. Davies (eds.) *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion*. Oxford University Press
- Chalmers, D. (2009) *The Two-Dimensional Argument Against Materialism*. In B. McLaughlin (ed.) *The Oxford Handbook of the Philosophy of Mind*. Oxford University Press
- Chalmers, D. (2012a). *Modality and the Mind-Body Problem: Reply to Goff and Papineau, Lee, and Levine*. Accessed September 2013: <http://consc.net/papers/modality.pdf>
- Chalmers, D. (2012b). *Constructing the World*. Oxford University Press
- Chalmers, D. (forthcoming-a). *The Combination Problem for Panpsychism*. In *Panpsychism*, (eds) Bruntrup, G. and Jaskolla, L., Oxford University Press, draft accessed January 2014 <http://consc.net/papers/combination.pdf>
- Chalmers, D. (forthcoming-b). *Panpsychism and Panprotopsychism*. In *Consciousness in the Physical World: Essays on Russellian Monism*, (eds) Nagasawa, Y. and Alter, T., Oxford University Press, draft accessed January 2014 <http://consc.net/papers/panpsychism.pdf>
- Chalmers, D. & Jackson, F. (2001). *Conceptual analysis and reductive explanation*. *The Philosophical Review* V110:13, pp.315-361
- Cheng, P. (1997). *From Covariation to Causation: A Causal Power Theory*. *Psychological Review* V104:2, pp.367-405
- Chisholm, R. (1976). *Person and Object: A Metaphysical Study*. London: G. Allen & Unwin
- Chisholm, R. (1991). *On the Simplicity of the Soul*. *Philosophical Perspectives* V5, pp.157-181
- Chuard, P. (2007). *Indiscriminable Shades and Demonstrative Concepts*. *Australasian Journal of Philosophy*, 85 (2): 277–306.
- Chudnoff, E. (2013). *Gurwitsch's Phenomenal Holism*. *Phenomenology and the Cognitive Sciences* 12:3, pp.559-578.
- Clark, A. (1989). *The Particulate Instantiation of Homogeneous Pink*. *Synthese* 80, pp.277–304
- Clarke, S. (1978) *The Works*, originally published 1738, reprinted New York: Garland Publishing
- Clarke, T. (1965). *Seeing Surfaces and Physical Objects*. In Max Black, ed., *Philosophy in America*, Cornell University Press
- Coleman, S. (2009). *Mind Under Matter*. in David Skrbina (ed.), *Mind that Abides*. Benjamins
- Coleman, S. (2012). *Mental Chemistry: Combination for Panpsychists*. *Dialectica* V66:1, 137–166.
- Coleman, S. (2013). *The Real Combination Problem: Panpsychism, Micro-Subjects, and Emergence*. *Erkenntnis* January 2013 <http://link.springer.com/article/10.1007%2Fs10670-013-9431-x>
- Coleman, S. (forthcoming). *Panpsychism and Neutral Monism: How to Make Up One's Mind*. In Bruntrup and Jasolla (eds), *Panpsychism*, Oxford: Oxford University Press.
- Copeland, B. (1993). *The Curious Case of the Chinese Gym*. *Synthese*, V95:2 pp. 173-186
- Corbetta, M, Miezin, F., Dobmeyer, S., Shulman, G., and Petersen, S. (1990) *Attentional modulation of neural processing of shape, color, and velocity in humans*. *Science* V248, pp.1556–1559
- Cotnoir, A. (2013). *Composition as General Identity*. In K. Bennett and D. Zimmerman (eds.) *Oxford Studies in Metaphysics*, volume 8. Oxford: OUP, pp.295–322
- Creary, L. (1981). *Causal Explanation and the Reality of Natural Component Forces*. *Pacific Philosophical Quarterly* 62, pp.148–157.
- Crick, F., and Koch, C. (1990). *Towards a Neurobiological Theory of Consciousness*. *Seminars in Neuroscience*, V012: pp 263-275

- Cudworth, R. (1837). *The True Intellectual System of the Universe: Wherein All the Reason and Philosophy of Atheism is Confuted, and Its Impossibility Demonstrated*. Originally published 1678, reprinted in Works, New York: Gould & Newman
- Dainton, B. (2000). *Stream of Consciousness: Unity and Continuity in Conscious Experience*. London and New York: Routledge
- Dainton, B., *Précis of 'Stream of Consciousness'*, Psyche V10:1, 2004
- Dainton, B. (2008). *The Phenomenal Self*. Oxford: Oxford University Press
- Dainton, B. (2010). *Phenomenal Holism*. Royal Institute of Philosophy Supplement 67, pp.113-139
- Dainton, B. (2011). *Review of Consciousness and its Place in Nature*. *Philosophy and Phenomenological Research* V83:1, 238-261.
- David, N., et al. (2008) The “sense of agency” and its underlying cognitive and neural mechanisms. *Consciousness and Cognition* 17 (2008) 523–534
- De Courcillon, Louis, *Sur la Religion* (1684) by Louis de Courcillon, Abbe de Dangeau (1643-1723), and his close friend Francois Timoleon, Abbe de Choisy
- Delaney, C. (1972). *Sellars's Grain Argument*. *Australasian Journal of Philosophy* 50:1, pp.14-16
- Della Rocca, M. (2008). *Spinoza*. Routledge
- Dennett, D. (1978). *Where am I?* In *Brainstorms*. MIT Press
- Dennett, D. (1991). *Consciousness Explained*. Penguin Press.
- Dennett, D. (1992). *The Self as Centre of Narrative Gravity*. in F. Kessel, P. Cole and D. Johnson (eds.) *Self and Consciousness: Multiple Perspectives*. Hillsdale, NJ: Erlbaum
- Descartes, R. (1985) *Meditations on First Philosophy*. Originally published 1641, in *The Philosophical Writings of Descartes*, 2 vols., Cottingham, J., Stoothoff, R., and Murdoch, D. (trans and eds.), Cambridge: Cambridge University Press
- Desmurget, M., Reilly, K.T., Richard, N., Szathmari, A., Mottolese, C., & Sirigu, A. (2009) Movement intention after parietal cortex stimulation in humans. *Science*, 324, 811-813
- De Sousa, R. (1976). *Rational Homunculi*. In Rorty, A. (ed.) *The Identities of Persons*, University of California Press, pp.217-238
- Diaz-Leon, E. (2011). *Reductive Explanation, Concepts, and a priori Entailment*. *Philosophical Studies* V155, pp.99–116
- Doré F. (1986). *Object permanence in adult cats (Felis catus)*. *Journal of Comparative Psychology* V100:4, pp.340–347.
- Dretske, F. (1994). *Introspection*. *Proceedings of the Aristotelian Society* 94, pp.263-278
- Dubner, R., and Zeki, S. (1971). *Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey*. *Brain Research* 35:2, pp.528–32
- Duke, Henry H. (1847) *A Systematic Analysis of Bishop Butler's Treatise on the Analogy of Religion to the Constitution of Nature, so far as relates to Natural Religion. To which is added, some considerations on certain arguments therein advanced*. London: Joseph Masters
- Duncan, J., & Humphreys, G. W. (1989). *Visual search and stimulus similarity*. *Psychological Review*, V96, pp.433-458.
- Duncan, J., & Humphreys, G. (1992). *Beyond the search surface: visual search and attentional engagement*. *Journal of Experimental Psychology: Human Perception and Performance* V18:2, pp.578-88; discussion pp.589-93
- Ebert, J., and Wegner, D. (2010). *Time Warp: Authorship Shapes the Perceived Timing of Actions and Events*. *Consciousness and Cognition* V19, pp.481–489
- Eklund, M. (2008). *Deconstructing ontological vagueness*. *Canadian Journal of Philosophy* V38:1, pp.117-140
- Elder, C. (2008). *Against Universal Mereological Composition*. *Dialectica*, 62, pp. 433-54

- Engel, S., Glover, G., & Wandell, B. (1997). *Retinotopic organization in human visual cortex and the spatial precision of functional MRI*. *Cerebral Cortex* V7 :2, pp.181-192
- Eriksen, C., St. James, J. (1986). *Visual attention within and around the field of focal attention: a zoom lens model*. *Perceptual Psychophysics* V40, pp.225–240
- Evans, G. (1982). *The Varieties of Reference*. Ed. John MacDowell. Oxford University Press
- Feinberg, I. (1978). *Efference Copy and Corollary Discharge: Implications for Thinking and its Disorders*. *Schizophrenia Bulletin* V4, pp.636-640
- Fine, K. (1975). *Vagueness, Truth, and Logic*. *Synthese* V30:3-4, pp.265-230
- Fine, K. (1982). *Acts, Events, and Objects*. In *Language and Ontology: Proceedings of the 6th International Wittgenstein Symposium*, Holder-Pichler-Tempsky, pp.97-105
- Fine, K. (1999). *Things and their Parts*. *Midwest Studies in Philosophy* V23, pp.61-74
- Fine, K. (2001). *The Question of Realism*. *Philosophers' Imprint* V1:2, pp.1-30
- Fishman R. (1997). *Gordon Holmes, the cortical retina, and the wounds of war. The seventh Charles B. Snyder Lecture*. *Documenta Ophthalmologica* 93, pp.9-28
- Frankfurt, H. (1987). *Identification and Wholeheartedness*. in Schoeman, F. (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, Cambridge University Press, pp.159-176
- Freud, S. (1949). *The Ego and the Id*. Originally published 1923, trans. J. Riviere. London: The Hogarth Press Ltd
- Fried, I. et al. (1991) *Functional Organization of Human Supplementary Motor Cortex Studied by Electrical Stimulation*. *Journal of Neuroscience*, 11(11) 3656-3666
- Friedman, M. (1974). *Explanation and scientific understanding*. *Journal of Philosophy* V71, pp.5-19
- Frith, C. (2012). *Explaining Delusions of Control: The Comparator Model 20 Years On*. *Consciousness and Cognition* 21, pp.52–54
- Gallese, V., Keysers, C., & Rizzolatti, G. (2004). *A unifying view of the basis of social cognition*. *Trends in Cognitive Sciences*, 8:9, pp.396-403.
- Gaudry, J. (2008). <http://panexperientialism.blogspot.ca/2008/05/does-physicalism-entail-cosmopsychism.html>
- Gazzaniga, M., Bogen, J., and Sperry, R. (1962). *Some Functional Effects of Sectioning the Cerebral Commissures in Man*. *Proceedings of the National Academy of Sciences* 48:2, p. 1765
- Geach, P. (1967). *Identity*. *The Review of Metaphysics* 21:1, pp.3 - 12
- Gerrans, P. (2001). *Authorship and Ownership of Thoughts*. *Philosophy, Psychiatry, & Psychology*, V8:2/3, pp. 231-237
- Gibbard, A. (1975). *Contingent Identity*. *Journal of Philosophical Logic* V4, pp.187-222.
- Gilbert, M. (2002). *Collective guilt and collective guilt feelings*. *The Journal of Ethics*, 6: 115-143.
- Gillett, C. (2002). *The dimensions of realization: A critique of the standard view*. *Analysis*, 62, pp.316–323.
- Gillett, C., and Rives, B. (2005). *The Non-Existence of Determinables: Or, a World of Absolute Determinates as Default Hypothesis*. *Nous*, V39:3, pp.483-504
- Goff, P. (2006). *Experiences Don't Sum*. *Journal of Consciousness Studies*, 13, No. 10–11, 2006, pp. 53–61.
- Goff, P. (2009a). *Why Panpsychism doesn't Help us Explain Consciousness*. *Dialectica* 63:3, pp.289-311
- Goff, P. (2009b). *Can the panpsychist get round the combination problem?* In D. Skrbina (ed.) *Mind that Abides: Panpsychism in the New Millennium*. John Benjamins.
- Goff, P. (2010). *There is No Combination Problem*. in M. Blamauer (ed.) *The Mental as Fundamental. New Perspectives on Panpsychism*. Ontos
- Goff, P. (2011). *Property Dualists should be Panpsychists*. <http://consciousnessonline.files.wordpress.com/2011/02/property-dualists-should-be-panpsychists.pdf>

- Goff, P. (2012). *There is More than One Thing*. In Goff, P. (ed) *Spinoza on Monism*. Palgrave MacMillan, pp. 113-122
- Goff, P. (forthcoming). *Real Acquaintance and Physicalism*. In P. Coates and S. Coleman (eds.), *Phenomenal Qualities: Sense, Perception and Consciousness*. Oxford University Press.
- Goff, P. (forthcoming a). *The phenomenal bonding solution to the combination problem*. In G. Bruntrop and L. Jaskolla (Eds.) *Panpsychism*, Oxford University Press
- Goldman, A. (2008) *Simulating Minds*. Oxford University Press.
- Goodman, N. (1951). *The Structure of Appearance*, 1st edition. Cambridge, MA: Harvard University Press.
- Gopnik, A. (2000). *Explanation as orgasm and the drive for causal understanding: The evolution, function and phenomenology of the theory-formation system*. In F. Keil and R. Wilson (eds.) *Cognition and Explanation*. Cambridge, Massachusetts: MIT Press. pp.299-323
- Grahek, N. (2007). *Feeling Pain and Being in Pain* (2nd Edition). MIT Press.
- Grimm, S. (2009). *Reliability and the Sense of Understanding*. In De Regt, H., Leonelli, S., and Eigner, K. (eds.) *Scientific Understanding: Philosophical Perspectives*. University of Pittsburgh Press, pp.83-100
- Grossberg, S., & Mignolla, E. (1985). *Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading*. *Psychological Review* V92, pp.173–211
- Gurwitsch, A. (1964). *The Field of Consciousness*. Pittsburgh: Duquesne University Press
- Hare, C. (2007). *Self-Bias, Time-Bias, and the Metaphysics of Self and Time*. *The Journal of Philosophy*, Volume 104:7, pp.350-373
- Harman, G. (1990). *The Intrinsic Quality of Experience*. In *Philosophical Perspectives* 4, edited by J. Tomberlin: Ridgeview Publishing Co.
- Haugeland, B. (2002). *Authentic Intentionality*. In M. Scheutz (ed.), *Computationalism: New Directions*. MIT Press, pp.159-175
- Heeger, D., Boynton, G., Demb, J, Seideman, E. & Newsome, W. (1999) *Motion opponency in visual cortex*. *Journal of Neuroscience* V19, pp.7162-7174
- Helle, R. (2013). *The Stoic Case for Cosmic Holism and Cosmopsychism in De Natura Deorum II*. Master's Thesis, University of Oslo
- Heller, M. (2000). *Temporal Overlap is Not Coincidence*. *Monist* 83:3, p.362-380
- Hellie, B. (2005). *Noise and Perceptual Indiscriminability*. *Mind* 114: 481–508.
- Hellie, B. (2010). *An Externalist's Guide to Inner Experience*. in Bence Nanay, ed., *Perceiving the World*. Oxford: Oxford University Press.
- Hellie, B. (2011). *There It Is*. *Philosophical Issues*, 21:110–164.
- Hellie, B. (2013). *Against Egalitarianism*. *Analysis*, 73:304–320.
- Hellie, B. (forthcoming-a). *How We Do*. In N. Charlow & M. Chrisman (eds.), *Deontic Modals*. Oxford University Press.
- Hellie, B. (forthcoming-b). *Love in the Time of Cholera*. In B. Brogaard (ed.), *Does Perception Have Content?* Oxford University Press. Accessed 2014 <http://individual.utoronto.ca/benj/love.pdf>
- Helmholtz, H. v. (1866). *Handbuch der Physiologischen Optik*. Leipzig: Voss.
- Hempel, C. and Oppenheim, P. (1948). *Studies in the Logic of Explanation*. *Philosophy of Science*, Vol. 15, No. 2, pp. 135-175
- Hempel, C. (1965). *Aspects of Scientific Explanation*, New York: Free Press.
- Horgan, T. (1993). *From Supervenience to Supradupervenience: Meeting the Demands of a Material World*. *Mind* V102.408, pp.555-586

- Horgan, T. and Tienson, J. (2002). *The Intentionality Of Phenomenology And The Phenomenology Of Intentionality*. In D. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press.
- Horgan, T., and Potrc, M. (2012). *Existence Monism Trumps Priority Monism*. In P. Goff (ed.) *Spinoza on Monism*, Palgrave Macmillan, pp.51-76
- Holmes, G. (1944). *The organization of the visual cortex in man*. Proceedings of the Royal Society, Series B, V132, pp.348-361
- Huebner, B. (2011). *Genuinely Collective Emotions*. European Journal for Philosophy of Science, V1:1, pp.89-118
- Hurley, K. (1998). *Consciousness in Action*. Harvard University Press.
- Hume, D. (1888). *Treatise of Human Nature*, ed. by L. A. Selby-Bigge. Oxford: Clarendon Press. (Original work 1739. Partly reprinted in Perry 1975.)
- Husserl, E. (1970). *Logical Investigations*, trans J.N. Findlay, London: Routledge and Kegan Paul
- Husserl, E. (1982). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*, First Book, trans F. Kersten, Dordrecht: Kluwer Academic Publishers
- Husserl, E. (1989). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*, Second Book, trans R. Rojcewicz and A. Schuwer, Kersten, Dordrecht: Kluwer Academic Publishers
- Husserl, E. (2001). *Analyses Concerning Active and Passive Synthesis: Lectures on Transcendental Logic*. Lectures given between 1920 and 1926, trans. A. Steinbock, Springer
- Jackson, F. (1993). *Block's Challenge*. In *Ontology, Causality and Mind: Essays in Honour of D M Armstrong*. New York: Cambridge University Press
- Jackson, Frank. 1998. *From Metaphysics to Ethics*. Oxford University Press.
- Jackson, F. (2003). *From H₂O to Water: The Relevance to a priori Passage*. In Lillehammer, H and Rodriguez-Pereyra, G., eds. *Real Metaphysics*. London: Routledge pp.84-97
- James, W. (1890). *The Principles of Psychology*. Harvard University Press.
- James, W. (1909). *A Pluralistic Universe*. Longman, Greens, & Co.
- James, W. (1916). *Talks To Teachers On Psychology; And To Students On Some Of Life's Ideals*. Henry Holt and Company
- Jaskolla, L. and Buck, A. (2012). *Does panexperientialistic holism solve the combination problem?* *Journal of Consciousness Studies*, V19:9-10, pp. 190-199
- Jenkins, C. S. (2011). *Is metaphysical dependence irreflexive?* *The Monist* V94, pp.267-276.
- Johansson, I. (2004). *On the Transitivity of the Parthood Relations*. In H. Hochberg and K. Mulligan (eds.) *Relations and Predicates*, Frankfurt: Ontos, pp. 161-181.
- Johnston, M. (2007). *Humans Beings revisited: My body is not an animal*. In *Oxford studies in metaphysics. V:3*, ed. D. Zimmerman, Oxford University Press, pp.33-74
- Johnston, W., Dark. V. (1986) *Selective attention*. *Annual Review of Psychology* V37, pp.43–75
- Kant, I. (1999) *Critique of Pure Reason*. Originally Published 1781, trans. P. Guyer and A. Wood, Cambridge University Press
- Kelly, S. (2003). *Edmund Husserl on Phenomenology*. in Robert C. Solomon (ed.) *Blackwell Guide to Continental Philosophy*, Oxford: Blackwell Publishers, pp.112-142
- Kelly, S. (2004). *On seeing things in Merleau-Ponty*. In: T. Carmon (Ed.), *The Cambridge companion to Merleau-Ponty*. Cambridge: Cambridge University Press.
- Kellman, P., & Shipley, T. (1991). *A theory of visual interpolation in object perception*. *Cognitive Psychology* V23, pp.141–221
- Kim, J. (1998). *Mind in a Physical World*. MIT Press

- Kim, J. (1999). *Making Sense of Emergence*. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Vol. 95, No. 1/2, Reduction and Emergence "The Thirty-Third Oberlin Colloquium in Philosophy", pp. 3-36
- Kim, S. (2011). *Multiple realization and evidence*. *Philosophical Psychology* V24:6, pp.739–749
- Kind, A. (2001). *Putting the Image back in Imagination*. *Philosophy and Phenomenological Research* V62(1), pp.85-109.
- Knobe, J., and J. Prinz. 2008. *Intuitions about consciousness: experimental studies*. *Phenomenology and Cognitive Science* 7: 67–83.
- Korsgaard, C. (1989). *Personal Identity and the Unity of Agency: A Kantian Response to Parfit*. *Philosophy and Public Affairs*, V18:2, pp.109-123 In J. Dancy (ed.) *Reading Parfit*, Wiley
- Korsgaard, C. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press
- Kriegel, U. (2004). *Consciousness and Self-Consciousness*. *Monist* 87, pp.182-205
- Kriegel, U. (2009). *Subjective Consciousness: A Self-Representational Theory*. Oxford University Press
- Kripke, S. (1980). *Naming and Necessity*. Cambridge: Harvard University Press
- Lamy, F. (1701). *De la connoissance de soi-meme*. 2nd ed. Chez Nicolas le Clerc. (Trans in Schachter 2002)
- Lane, T. and Liang, C. (2011). *Self-Consciousness and Immunity*. *Journal of Philosophy* 108 (2), pp.78-99
- Langland-Hassan, P. (2013). *Craniopagus Twins and the Possibility of Introspective Misidentification*. <http://consciousnessonline.files.wordpress.com/2013/02/langland-hassan-co5.pdf>
- Leibniz, G. W. (1967). *The Leibniz-Arnauld Correspondence*. Trans. and ed. by H.T. Mason. Manchester: Manchester University Press
- Leibniz, G. W. (1981). *New Essays on Human Understanding*. Trans. Eds. P. Remnant and J. Bennett, Cambridge: Cambridge University Press
- Leibniz, G.W. (2012). *The Discourse on Metaphysics*, in *The Discourse on Metaphysics and Other Writings*, Latta and Montgomery (trans.), revised by Peter Loftson (ed), Broadview Press, pp.57-102 (originally published 1686)
- Lennon, T., and Stainton, J. (eds.) (2008). *The Achilles of Rationalist Psychology*. *Studies In The History Of Philosophy Of Mind*, V7, Springer
- Leopold, A. (1949). *A Sand County Almanac*. New York: Oxford University Press
- Levine, J. (1983). *Materialism and Qualia: The Explanatory Gap*. *Pacific Philosophy Quarterly* V64, pp.354-361
- Lewis, D. (1971). *Counterparts of persons and their bodies*. *Journal of Philosophy* V68, pp.203-211.
- Lewis, D. (1976). *Survival and Identity*. in *The Identities of Persons*, A. Rorty (ed.), Berkeley: California
- Lewis, D. (1979). *Attitudes De Dicto and De Se*. *The Philosophical Review*, V88:4, pp. 513-543
- Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Blackwell
- Lewis, D. (1988). *Vague Identity: Evans Misunderstood*. *Analysis*, 48:3, pp.128-130
- Lewis, D. (1991). *Parts of Classes*. Oxford: Basil Blackwell
- Lewis, D. (1993). *Many, But Almost One*. in *Ontology, Causality, and Mind: Essays in Honour of D. M. Armstrong*, eds. John Bacon, Keith Campbell, and Lloyd Reinhardt, Cambridge University Press, pp.23-45
- Lewis, D. (2003). *Things qua Truthmakers*. In H. Lillehammer & G. Rodriguez-Pereyra (eds.), *Real Metaphysics: Essays in honor of D. H. Mellor*. Routledge, pp.25-38
- Lewtas, P. (2013). *What is it like to be a Quark?* *Journal of Consciousness Studies*, V20:9-10, pp.39-64
- Lipton, P. (2009). *Understanding without Explanation*. In De Regt, H., Leonelli, S., and Eigner, K. (eds.) *Scientific Understanding: Philosophical Perspectives*. University of Pittsburgh Press, pp.43-64
- Loar, B. (1990). *Phenomenal States*. *Philosophical Perspectives* V4, pp.81-108
- Locke, J. (1836). *An Essay Concerning Human Understanding*. (Originally published 1689), T. Tegg and Son.

- Lockwood, M. (1989). *Mind, Brain, and the Quantum*. Oxford: Oxford University Press
- Lockwood, M. (1993). *The Grain Problem*. In Robinson, H. (ed.) *Objections to Physicalism*. Oxford: Oxford University Press.
- Lotze, Microcosmus: An Essay Concerning Man and His Relation to the World. (See Vol. I, trans, by Elizabeth Hamilton and E. E. Constance Jones [New York: Charles Scribner's Sons, 1894], p. 158.)
- Lovelock, J. (2000). *Gaia: A New Look at Life on Earth*. Oxford: Oxford University Press.
- Lowe, E. J. (1996). *Subjects of Experience*. Cambridge University Press
- Lycan, G. (1979). *A new lilliputian argument against machine functionalism*. *Philosophical Studies* 35, pp.279-287
- Lynn, M., Berger, C., Riddle, T, and Morsella, E. (2010). Mind control? Creating illusory intentions through a phony brain-computer interface. *Consciousness and Cognition* V:19, pp.1007-1012
- Madden, R. (2012). The Naive Topology of the Conscious Subject. *Nous* 100:0, pp.1-15
- Margulis, L. (1998). *Symbiotic Planet: A New Look at Evolution*. London: Weidenfeld & Nicolson.
- Markosian, E. (1998). *Brutal Composition*. *Philosophical Studies* 92:3, pp.211-249
- Masrour, F. (forthcoming). *Unity of Consciousness: Advertisement for a Leibnizian View*. In *Sensory Integration and the Unity of Consciousness*, eds. C. Hill and D. Bennett. Cambridge: MIT Press.
- Matey, J. (2013). *Representing the Impossible*. *Philosophical Psychology*, V26:2, pp.188-206
- Matthews, F. (2003). *For Love of Matter: a Contemporary Panpsychism*. SUNY Press, Albany NY
- Maxwell, G. (1979). *Rigid Designators and Mind-Brain Identity*. In Savage, C. (ed.) *Perception and Cognition: Minnesota Studies in the Philosophy of Science* Vol. 9. Minneapolis: Minnesota University Press, pp.365-403
- McGinn, C. (1989). *Can we Solve the Mind-Body Problem?* *Mind* V98, pp.349-66
- McLaughlin, B. (1992). *The Rise and Fall of British Emergentism*. In A. Beckermann, H. Flohr & J. Kim (eds.), *Emergence or Reduction?: Prospects for Nonreductive Physicalism*. De Gruyter
- McQueen, K. (2013). *A Priori Entailment and the Metaphysics of Science*. Doctoral Dissertation, Australian National University. Accessed February 2014:
<http://philrsss.anu.edu.au/sites/default/files/PENULTIMATE%20DRAFT%20%28for%20submission%20with%20margins%29.pdf>
- McQueen, K. (2014a). *A Priori Entailment and Mass Additivity*. Accessed February 2014:
<http://philrsss.anu.edu.au/sites/default/files/A%20Priori%20Entailment%20and%20Mass%20Additivity%20-%20WEB%20%26%20SW.pdf>
- McQueen, K. (2014b). *Reductionism, Levels, and Metal Rods*. Accessed February 2014:
<http://philrsss.anu.edu.au/sites/default/files/Reductionism%20levels%20and%20metal%20rods.pdf>
- McTaggart, J. (1988). *The Nature of Existence*. V1, ed. C. D. Broad. Cambridge: Cambridge University Press
- Melloni, L., Molina, C., Pena, M., Torres, D., Singer, W., and Rodeiguez, E. (2007). *Synchronization of Neural Activity across Cortical Areas Correlates with Conscious Perception*. *The Journal of Neuroscience*, V27:11, pp. 2858-2865
- Mendelssohn, M. (1767/1789/2002). *Phaedon or, The Death of Socrates*. Originally published 1767, trans. C. Cullen, 1789, Schiller Institute Inc. http://www.schillerinstitute.org/transl/mend_phadn_cullen.html (accessed July 2013)
- Merleau-Ponty, M. (1962). *The Phenomenology of Perception*. Originally published 1945, trans. Colin Smith. Routledge and Kegan Paul
- Merricks, T. (2001). *Objects and Persons*. Oxford: Oxford University Press
- Merricks, T. (2005). *Composition and Vagueness*. *Mind* 114, pp.615-637
- Michotte, A., Thines, G., & Crabbe, G. (1964/1991). *Amodal completion of perceptual structures*. In: G. Thines, A. Costall, & G. Butterworth (Eds.), *Michotte's experimental phenomenology of perception* (pp. 140-167). Hillsdale, NJ: Erlbaum.

- Mijuskovic, B. L. (1984). *The Achilles of Rationalist Arguments: The Simplicity, Unity, and Identity of Thought and Soul From the Cambridge Platonists to Kant: A Study in the History of an Argument*. Martinus Nijhoff.
- Mill, J. S. (1843). *A System of Logic, Ratiocinative and Inductive*. 8th Edition Published 1882, Harper & Brother, New York.
- Miller, H., Gipson, C., Vaughan, A., Rayburn-Reeves, R., and Zentall, T. (2009). *Object permanence in dogs: invisible displacement in a rotation task*. *Psychonomic Bulletin and Review*, V16:1, pp.150-155
- Mizrahi, V. (2009). *Is Colour Composition Phenomenal?* in D. Skusevich and P. Matikas (eds), *Color Perception: Physiology, Processes and Analysis*, Nova Science Publishers, pp. 185-202.
- Mole, C. (2010). *Attention is Cognitive Unison: An Essay in Philosophical Psychology*. Oxford: Oxford University Press
- Molnar, G. (2003). *Powers*. Oxford: Oxford University Press.
- Moore, J., and Haggard, P., (2008) Awareness of action: Inference and prediction. *Consciousness and Cognition* 17 136–144
- Mørch, H. H. (2013). *Comments On Sebastian: What Panpsychists Should Reject: On The Incompatibility Of Panpsychism And Organizational Invariantism*.
<http://consciousnessonline.files.wordpress.com/2013/02/hassel-comments-on-sebastian.pdf>
- Mørch, H. H. (2014). *Panpsychism and Causation: A New Argument and a Solution to the Combination Problem*. Doctoral Dissertation, University of Oslo
- Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation*.
- Nagel, E. (1963) *Wholes, sums, and organic unities*, in: D. Lerner (Ed.) *Parts and Wholes* (New York, The Free Press).
- Nagel, T. (1971). *Brain Bisection and the Unity of Consciousness*. *Synthese* 22 pp.396-413.
- Nagel, T. (1974). *What is It Like to Be a Bat?* *Philosophical Review* V83 pp.435-450
- Nagel, T. (1986). *The View From Nowhere*. Oxford University Press.
- Nagel, T. (2004). *Panpsychism*. In D.S.Clarke, *Panpsychism: Past and Recent Selected Readings*, State University of New York Press, originally published Cambridge University Press, 1979
- Nanay, B. (2010). *Perception and Imagination: Amodal Perception as Mental Imagery*. *Philosophical Studies*, V50, pp.239-254
- Nieder, A. (2002). *Seeing More than Meets the Eyes: Processing of Illusory Contours in Animals*. *Journal of Comparative Physiology A*, V188, pp.249-260
- Nielsen, T. (1963). *Volition: Towards a New Experimental Approach*. *Scandinavian Journal of Psychology* V4, pp.225-230
- Noë, A. (2005). *Real presence*. *Philosophical Topics*, 33, pp.235–264
- Noonan, H. (2003). *Personal Identity*, 2nd ed. First edition 1989. Routledge
- Noonan, H. (2010). *The Thinking Animal Problem and Personal Pronoun Revisionism*. *Analysis* 70:1, pp.93-98
- Nozick, R. (2003). *Personal Identity Through Time*. in *Personal Identity*, eds. Martin and Barresi, Blackwell, pp.92-115, originally in *Philosophical Explanations*, Belknap, 1981, pp.29-48, 50-51, 58-61, and 69
- Olson, E. (1997). *The Human Animal: Personal Identity without Psychology*. Oxford University Press.
- Olson, E. (2003). *An Argument for Animalism*. In *Personal Identity*, (eds) Martin and Barresi, Blackwell, p.318-335
- Parfit, D. (1971). *Personal Identity*. *Philosophical Review* 80, pp.3-27.
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press
- Parfit, D. (1999). *Experiences, Subjects, and, Conceptual Schemes*. *Philosophical Topics* 26, pp.217-270
- Paul, L. A. (2002). *Logical Parts*. *Nous* V36:4, pp.578-596

- Perky, C. (1910) *An Experimental Study of Imagination*. American Journal of Psychology V21, pp. 422–452.
- Perry, J. (1972). *Can the Self Divide? The Journal of Philosophy*, V69, pp.463-488
- Perry, J. (1979). *The Problem of the Essential Indexical*. *Nous* 13, pp.3-21.
- Pettit, P. and List, C. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press
- Phillips, I. (2014). *The temporal structure of experience*. In V. Arstila and D. Lloyd (Eds.), *Subjective Time: the Philosophy, Psychology, and Neuroscience of Temporality*, pp. 139–158. Cambridge, MA: MIT Press.
- Plato (2000). *The Republic*. (Originally published 360BCE). Tom Griffith (trans.), G. R. F. Ferrari (ed.), Cambridge University Press
- Ploner, M., Freund, H. J., and Schnitzler, A. (1999). *Pain Affect without Pain Sensation in a Patient with a Postcentral Lesion*. *Pain* 81: 211-14.
- Plotinus (1956). *Enneads*. Trans. and eds. Stephen Mackenna and B. S. Page, Faber and Faber limited, <http://archive.org/stream/plotinustheennea033190mbp#page/n11/mode/2up> (accessed July 2013)
- Polger, T. (2008). *Two confusions concerning multiple realization*. *Philosophy of Science* V75, pp.537–547.
- Powell, C. T. (1990). *Kant's Theory of Self-Consciousness*. Oxford: Oxford University Press
- Prete, F. (2004). *Complex Worlds from Simpler Nervous Systems* (ed.). MIT Press
- Prinz, J. (2011). *Is Attention Necessary and Sufficient for Consciousness?* In C. Mole, D. Smithies and W. Wu (eds.), *Attention: Philosophical and Psychological Essays*. Oxford University Press
- Prinz, J. (2012). *The Conscious Brain: How Attention Engenders Experience*. Oxford University Press
- Proclus (1963). *The Elements of Theology*. Trans. and ed. E. R. Dodds. Oxford, Clarendon Press.
- Putnam, H. (1965/2003). *The Nature of Mental States*. In O'Connor, T., Robb, D., and Heil, J., (eds) *Philosophy of Mind: Contemporary Readings*. Routledge, pp.223-235
- Raffman, D. (1994). *Vagueness without Paradox*. *The Philosophical Review*, V103:1, 1994, pp. 41-74
- Raffman, D. (2012). *Indiscriminability and phenomenal continua*. *Philosophical Perspectives* 26 (1):309-322
- Raffman, D. (forthcoming). *Unruly Words: A Study of Vague Language*. ?? Press
- Ramachandran, V. S. (1995). *Filling in the gaps in logic: Reply to Durgin et al.* *Perception* V24, pp.841–843
- Ramachandran, V.S & Rogers-Ramachandran, D. (1996). *Synaesthesia in Phantom Limbs Induced with Mirrors*. *Proceedings of the Biological Sciences*, V263:1369, pp. 377-386
- Raymont, P. (2005). *Conscious Unity* <http://philpapers.org/archive/RAYCU.1.pdf>
- Recanati, F. (2012). *Immunity to error through misidentification: what it is and where it comes from*. In S. Prosser and F. Recanati (eds), *Immunity to Error through Misidentification*, Cambridge University Press pp.80-121
- Roelofs, L. (2014a). *Phenomenal Blending and the Palette Problem*. *Thought*, V3:1, pp.59-70
- Roelofs, L. (2014b). *What are the Dimensions of the Conscious Field?* *Journal of Consciousness Studies*, Volume 21, Numbers 7-8, pp. 88-104
- Rosen, G. (2010). *Metaphysical dependence: grounding and reduction*. In Hale and Hoffman, *Modality: Metaphysics, Logic, and Epistemology*. Oxford: Oxford University Press. pp.109-136.
- Rosenberg, G. (1998). *The Boundary Problem for Phenomenal Individuals*. In S. Hameroff, A. Kaszniak and A. Scott (eds.), *Toward a Science of Consciousness: The First Tucson Discussions and Debates (Complex Adaptive Systems)*. MIT Press.
- Rosenberg, G. (2004). *A Place for Consciousness: Probing the Deep Structure of the Natural World*. Oxford: Oxford University Press.
- Rosenthal, D. (1986). *Two concepts of consciousness*. *Philosophical Studies*, 49: 329–359

- Rosenthal, D. (2007). *Phenomenological overflow and cognitive access*. Behavioral and Brain Sciences, V30:5-6, pp.522-523
- Rozemond, M. (2008). *The Achilles Argument and the Nature of Matter in the Clarke Collins Correspondence*. In Lennon, T., and Stainton, J. (eds), *The Achilles of Rationalist Psychology. Studies In The History Of Philosophy Of Mind*, V7, Springer
- Russell, B. (1903). *The Principles of Mathematics*. Cambridge University Press
- Russell, B. (1919). *Introduction to Mathematical Philosophy*, London: Routledge
- Russell, B. (1985). *The Philosophy of Logical Atomism*. In D. Pears (ed.) *The Philosophy of Logical Atomism*. LaSalle, IL: Open Court, pp.35–155
- Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson.
- Sartre, J. (1956). *Being and Nothingness: An Essay in Phenomenological Ontology*. Originally published 1943, trans. Hazel E. Barnes, New York: Philosophical Library
- Sato, A. & Yasuda, A. (2005). *Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership*. *Cognition* 94:3, pp.241-255
- Schachter, J. (2002). *Pierre Bayle, Matter, and the Unity of Consciousness*. *Canadian Journal of Philosophy*, V32:2, pp. 241-265
- Schaffer, J. (2010). *Monism: The priority of the whole*. *Philosophical Review* 119, pp.31-76.
- Schaffer, J. (2012). *Why the World has Parts*. In P. Goff (ed.) *Spinoza on Monism*, Palgrave Macmillan, pp.77-91
- Schechter, E. (2010). *Two Unities of Consciousness*. *European Journal of Philosophy*, V21:2, pp.1-25
- Schiffer, S. (1987). *Remnants of Meaning*. Cambridge: Cambridge University Press.
- Schilder, P., and Stengel, E. (1931). *Asymbolia for pain*. *Archives of Neurology & Psychiatry*, 25(3): 598-600.
- Schwitzgebel, E. (2014). *If Materialism is True, the United States is Probably Conscious*. <http://www.faculty.ucr.edu/~eschwitz/SchwitzPapers/USAconscious-140130a.pdf>
- Seager, W. (1990). 'The Logic of Lost Lingers', *Journal of Philosophical Logic*, 19, pp.407-428.
- Seager, W. (1995). *Consciousness, Information and Panpsychism*. *Journal of Consciousness Studies*, V2-3, pp.272-88
- Seager, W. (2001). *The Constructed and the Secret Self*. In Andrew Brook and Richard DeVidi *Self-Reference and Self-Knowledge*. pp.247-268
- Seager, W. (2010). *Panpsychism, Aggregation and Combinatorial Infusion*. *Mind and Matter*, V8:2, pp. 167-184
- Seager, W. (forthcoming). *Panpsychist Infusion*. In *Panpsychism*, (eds) Bruntrup, G. and Jaskolla, L., Oxford University Press
- Searle, J. (1990). *Collective Intentions and Actions*. in *Intentions in Communication*, ed. by Philip R. Cohen, Jerry Morgan, and Martha E. Pollack. Cambridge, Massachusetts: MIT Press
- Searle, J. (1980/2003). *Minds, Brains, and Programs*. In O'Connor, T., Robb, D., and Heil, J., (eds) *Philosophy of Mind: Contemporary Readings*. Routledge, pp.345-365
- Searle, J. (2000). *Consciousness, Annual Review of Neuroscience*, V23, pp.557–578
- Sebastien, M. (2013). *What Panpsychists should Reject: On the Incompatibility of Panpsychism and the Principle of Organisational Invariance*. <http://mindingthebrain.files.wordpress.com/2013/02/ppoi.pdf>
- Segal, S. (1972). *Assimilation of a Stimulus in the Construction of an Image: The Perky Effect Revisited*. In P.W. Sheehan (Ed.), *The Function and Nature of Imagery*, New York & London: Academic Press, pp. 203-230
- Selfridge, O. (1959). *Pandemonium: A Paradigm for Learning*. Symposium on the Mechanization of Thought Processes, HM Stationery Office
- Sellars, W. (1963). *Science, Perception and Reality*. London: Routledge and Kegan Paul

- Setiya, K. (2008). *Practical Knowledge*. *Ethics* V118:3, pp.388-409.
- Shani, I. (2010). *Mind Stuffed with Red Herrings: Why William James' Critique of the Mind-Stuff Theory Does not Substantiate a Combination Problem for Panpsychism*. *Acta Analytica*, 25:4, pp.413–434.
- Shoemaker, S. (1968). *Self-Reference and Self-Awareness*. *Journal of Philosophy* 65, pp.555-567.
- Shoemaker, S. (1970). *Persons and Their Pasts*. *American Philosophical Quarterly* 7:4, pp.269-85.
- Shoemaker, S. (1984). *Personal Identity*. B. Blackwell.
- Shoemaker, S. (1994). *Self-Knowledge and "Inner Sense": Lecture II: The Broad Perceptual Model*. *Philosophy and Phenomenological Research*, Vol. 54, No. 2, pp. 271-290
- Shoemaker, S. (1997). *Parfit on Identity*. In 'Reading Parfit', ed. Jonathan Dancy, Blackwell Oxford, pp. 135-148
- Shoemaker, S. (2000). *Realization and mental causation*. In B. Eleovich (Ed.), *Proceedings of the 20th world congress, vol. 9: Philosophy of mind* (pp. 23–31). Bowling Green: Philosophy Documentation Center.
- Shoemaker, S. (2003a). *Self, Body, and Coincidence*. *Proceedings of the Aristotelian Society* 63, pp.287-306.
- Shoemaker, S. (2003b). *Consciousness and Co-Consciousness*. In Axel Cleeremans (ed.), *The Unity of Consciousness*. Oxford University Press.
- Sider, T. (1997). *Four-Dimensionalism*. *Philosophical Review* 106:2, pp.197-231.
- Sider, T. (2001). *Maximality and Intrinsic Properties*. *Philosophy and Phenomenological Research* 63:2, pp.357-364.
- Sider, T. (2007). *Parthood*. *Philosophical Review* V116, pp.51–91
- Sider, T. (2011). *Writing the Book of the World*. Oxford University Press
- Sider, T. (2013). *Against Parthood*. In K. Bennett and D. Zimmerman (eds.) *Oxford Studies in Metaphysics*, volume 8. Oxford: OUP
- Siegel, S. (2006). *Which Properties are Represented in Perception?* In Szabo Gendler & Hawthorne (eds.) *Perceptual Experience*, Oxford : Clarendon Press, pp.481-503
- Siewert, C. (forthcoming). *Speaking up for Consciousness*. Forthcoming in *Current Controversies in Philosophy of Mind*, Routledge. <http://charlessiewert.com/attachments/article/66/SpeakingUpFinal.pdf>
- Simons, P. (1987). *Parts: A Study in Ontology*. Clarendon Press
- Skrbina, D. (2005). *Panpsychism in the West*. MIT Press
- Smart, J. J. C. (1981). *Physicalism and Emergence*. *Neuroscience* V6, pp.109-113
- Smithies, D. (2011). *Attention is Rational-Access Consciousness*. in C. Mole, D. Smithies and W. Wu (eds.), *Attention: Philosophical and Psychological Essays*. Oxford University Press
- Sperry, R. (1964). *Brain Bisection and Mechanisms of Consciousness*. In John C. Eccles (ed.), *Brain and Conscious Experience*. Springer-Verlag
- Spinoza, B. (1994). *Ethics Demonstrated in Geometrical Order*, in E. Curley (ed. and trans.), *A Spinoza Reader*, Princeton University Press, pp.85-265 (originally published 1677)
- Sprigge, T. (1983). *The Vindication of Absolute Idealism*. Edinburgh: Edinburgh University Press
- Stephens, G. L., and Graham, G. (2000). *When Self-Consciousness Breaks: Aliens Voices and Inserted Thoughts*. The MIT Press.
- Stoljar, D. (2001). *Two Conceptions of the Physical*. *Philosophy and Phenomenological Research*, Vol. 62:2, pp.253-281.
- Stoljar, D. (2006). *Comments on Galen Strawson - Realistic Monism: Why Physicalism Entails Panpsychism*. *Journal of Consciousness Studies* 13:10-11, pp.170-176.
- Strawson, G. (2006). *Realistic Monism - Why Physicalism Entails Panpsychism*. *Journal of Consciousness Studies* 13:10-11, pp.3-31.

- Strawson, G. (2006). *Realistic Monism: Why Physicalism Entails Panpsychism*. In A. Freeman (ed.), *Consciousness and its Place in Nature: does physicalism entail panpsychism?* Imprint Academic
- Strawson, G. (2009). *Selves: An Essay in Revisionary Metaphysics*. Oxford University Press
- Strawson, G. (2013). *Self-Intimation*. Phenomenology and the Cognitive Sciences, October 2013, published online: <http://link.springer.com/article/10.1007%2Fs11097-013-9339-6/fulltext.html>
- Strawson, P. F. (1974). *Imagination and Perception*. In *Freedom and Resentment and Other Essays*, London: Methuen, pp.45-65
- Swinburne, R. (1984). *Personal Identity: The Dualist View*. Oxford: Blackwell
- Swinburne, R. (1986). *The Evolution of the Soul*. Oxford University Press. Revised edition 1997.
- Synofzik, M., Vosgerau, G., and Newen, A. (2008). *Beyond the comparator model: A multifactorial two-step account of agency*. *Consciousness and Cognition* V17, pp.219–239
- Théoret, H., Kobayashi, M., Ganis, G., Di Capua, P. and Pascual-Leone, A. (2002) *Repetitive transcranial magnetic stimulation of human area MT/V5 disrupts perception and storage of the motion aftereffect*. *Neuropsychologia* V40:13, pp.2280-2287
- Titchener, E. B. (1909). *A textbook of psychology*. New York: Macmillan. Accessed August 26 2013: http://archive.org/stream/textbookofpsycho00edwa/textbookofpsycho00edwa_djvu.txt
- Tononi, G. and Balduzzi, D. (2009). *Qualia: The Geometry of Integrated Information*. *Computational Biology* V5:8
- Tononi, G. (2012). *Integrated information theory of consciousness: an updated account*. *Archives Italiennes de Biologie*. 150:2-3, pp.56-90
- Tononi, G., and Koch, C. (2014). *Consciousness: Here, There, but Not Everywhere*. <http://arxiv.org/ftp/arxiv/papers/1405/1405.7089.pdf>
- Treisman, A., and Gelade, G. (1980). *A feature-integration theory of attention*. *Cognitive Psychology*, V12:1, pp.97–136.
- Treisman, A., and Sato, S. (1990). *Conjunction Search Revisited*. *Journal of Experimental Psychology: Human Perception and Performance*, V16:3, pp.459-478
- Treue, S., and Maunsell, J. (1999). Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *Journal of Neuroscience*, V19:17, pp.7591-7602
- Trout, J. (2002). *Scientific Explanation and the Sense of Understanding*. *Philosophy of Science* 69:2, pp.212-233
- Tye, M. 1992. *Visual qualia and visual content*. In *The Contents of Experience*, edited by T. Crane. Cambridge: Cambridge University Press.
- Tye, M. (2003). *Consciousness and Persons: Unity and Identity*. MIT Press
- Tye, M. (2009). *A New Look at the Speckled Hen*. *Analysis*, V69:2: pp.258-263.
- Unger, P. (1979). *There Are No Ordinary Things*. *Synthese* 41:2, pp.117 - 154.
- Unger, P. (1980). *The Problem of the Many*. *Midwest Studies in Philosophy* 5:1, pp.411-468
- Unger, P. (1990). *Identity, Consciousness, and Value*. Oxford: Oxford University Press.
- Unger, P. (2004). *Mental Problems of the Many*. In *Oxford Studies in Metaphysics*, Vol. 1, ed. Dean Zimmerman. Oxford: Clarendon Press.
- Unger, P. (2006). *All the Power in the World*. Oxford University Press
- Van Cleve, J. (1990). *Mind-Dust or Magic? Panpsychism Versus Emergence*. *Philosophical Perspectives* V4, pp.215-226.
- Van Inwagen, P. (1990). *Material Beings*. Cornell University Press
- Varzi, A. (2006). *A Note on the Transitivity of Parthood*. *Applied Ontology*, V1:2, pp.141-146.
- Watzl, S. (2010). *The Significance of Attention*. Doctoral Dissertation, Columbia University

- Watzl, S. (2011). *Attention as Structuring of the Stream of Consciousness*. In C. Mole, D. Smithies & W. Wu (eds.), *Attention: Philosophical and Psychological Essays*. Oxford University Press
- Wegner, D. (2002). *The Illusion of Conscious Will*. MIT Press.
- Wiggins, D. (1967). *Identity and Spatio-Temporal Continuity*. Oxford University Press
- Wilson, A. (2009). *Disposition-manifestations and Reference-frames*. *Dialectica* 63, pp.591–601.
- Wilson, M. (1999). *Objects, Ideas, and “Minds”*: Comments on Spinoza’s Theory of Mind, in *Ideas and Mechanism: Essays on Early Modern Philosophy*, Princeton University Press, p. 126-140,
- Wilson, J. (1999). *How Superduper does Physicalist Superdupervenience need to be?* *Philosophical Quarterly* V49:194, pp.33-52
- Wilson, J. (2002). *Causal Powers, Forces, and Superdupervenience*. *Grazer Philosophische Studien* V63, pp.53-78
- Wilson, J. (2005). *Supervenience-based Formulations of Physicalism*. *Nous* V39:3, pp.426–459
- Wilson, J. (2009). *The Causal Argument against Component Forces*. *Dialectica* 63, pp.525–554.
- Wilson, J. (2010). *Metaphysical Emergence: Weak and Strong*. in *Metaphysics and Science*, Stephen Mumford and Matthew Tugby, eds. Accessed online September 2013: <http://individual.utoronto.ca/jmwilson/MEWaS.pdf>
- Wilson, J. (2012). *Fundamental Determinables*. *Philosophers' Imprint* V12:4, pp.
- Wilson, J. (2013-a). *Nonlinearity and Metaphysical Emergence*. In Mumford and Tugby (eds.), *Metaphysics and Science*, Oxford University Press, pp.201-229
- Wilson, J. (2013-b). *No Work for a Theory of Grounding*. *Inquiry*, pp.1-45 Access online 2013 <http://individual.utoronto.ca/jmwilson/NWfaTG.doc>
- Wilson, J. (2013-c). *A Determinable-Based Account of Metaphysical Indeterminacy*. *Inquiry* V56:4, pp.359–385.
- Wilson, J. (forthcoming). *Metaphysical Emergence*.
- Wilson, J. and Biggs, S. (2013). *Abductive Two-Dimensionalism: A New Route to the A Priori Identification of Necessary Truths*. Accessed September 2013: <http://individual.utoronto.ca/jmwilson/A2D.docx>
- Wittgenstein, L. (1958). *The Blue and Brown Books*. New York: Oxford
- Wu, W. (2011). *Attention as Selection for Action*. In C. Mole, D. Smithies & W. Wu (eds.), *Attention: Philosophical and Psychological Essays*. Oxford University Press
- Zihl, J., von Cramon, D., and Mai, N. (1983). *Selective disturbance of movement vision after bilateral brain damage*. *Brain* V106, pp.313-340
- Zuboff, A. (1990). *One Self: The Logic of Experience*. *Inquiry* V33:1, pp.39-68.
- Zuboff, A. (1982). *The Story of a Brain*. In D. R. Hofstadter and D. C. Dennett (eds.) *The Mind's I: Fantasies and Reflections on Self and Soul*. New York: Basic Books, pp.202-220