

<http://researchcommons.waikato.ac.nz/>

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

Extended Identities

The Reducibility of Non-categorical Properties and
their Bearers

A thesis
submitted in fulfilment
of the requirements for the Degree
of
Doctor of Philosophy in Philosophy
at
The University of Waikato
by
Paul Donald Charles Hubble



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2020

Abstract

When we take apart big or complicated things—whether events, processes, systems, objects or states—we find smaller or simpler things. Still, some who purport that everything is physical also want to deny that complicated things *are* sets of simpler things, their relations, and interactions. But that view leads to some confounding puzzles. By contrast, I defend the view that complicated things are nothing over and above simpler things, their relations, and interactions. The latter are all physical and located in space and time, and so are complicated things, like minds and people, and other things defined by what they could or would do. That includes corkscrews and water soluble things. However, the complicated properties of these complicated things show that, very often, these things are either not quite *what* we think they are, or they are not wholly *where* we think they are.

Sometimes things aren't what we thought. There is no lumniferous aether or phlogiston. But often, complicated things elude identification with their sets of simpler things, relations and interactions, not because they don't belong in predictively successful models of our world, but because they aren't wholly *where* we tend to look. Their boundaries are wider and untidy. As it happens, our minds are like that, and so are we. Finally, the fact that the extended identity of things like us comes as a surprise means our self-engineering could be more self-aware, and should be more self-reflective.

Acknowledgements

For their curiosity, insight and encouragement, I thank all my teachers. At Ryerson University, I especially thank Don Snyder, Elizabeth Trott, Glenn Parsons, and John Caruana. At the Institute for Christian Studies, I especially thank Ron Kuipers and Lambert Zuidervaat.

For their tremendous and instrumental work throughout, I thank my supervisors, Justine Kingsbury, Dan Weijers, and Nick Munn. It has been a pleasure to work with such a team. I owe a special thanks to Justine, whose guiding wisdom and generosity with time and encouragement I cannot begin to measure.

For their effort and input, I thank Joe Ulatowski, and my examiners, Douglas Campbell and Rob Wilson.

For financial support, I thank the University of Waikato.

For their love and patient support, I thank my friends and family, but especially my wife, Fiona.

Scholarship is a truly communal effort.

Contents

| | | |
|----------|--|------------|
| 1 | Introduction | 1 |
| 1.1 | We Are What We Make? | 1 |
| 1.2 | Thesis Outline | 3 |
| 2 | The Difficulties with Extended Minds | 11 |
| 2.1 | The First Battle Line | 13 |
| 2.2 | The Varieties of Externalism | 23 |
| 2.3 | The Second Battle Line | 27 |
| 3 | Concepts of Dependence | 39 |
| 3.1 | Synchronic Dependence Relations Sketched | 40 |
| 3.2 | Bennett's Resemblance Class of Building Relations | 45 |
| 3.3 | Rebuilding | 48 |
| 4 | Puzzles of Synchronic Dependence | 63 |
| 4.1 | Molecules and Minds | 65 |
| 4.2 | Antireductionism and its Discontents | 66 |
| 4.3 | A Recipe for Problems | 70 |
| 4.4 | Prospects for Grounding and Just Enough Determination | 74 |
| 5 | Antireductionism and Multiple Realizability | 83 |
| 5.1 | The Argument for Antireductionism from Multiple Realizability | 85 |
| 5.2 | A Dilemma for <i>Multiple Realizability</i> , and its First Horn | 88 |
| 5.3 | The Second Horn and Similarity | 102 |
| 6 | Similarity and Reduction | 112 |
| 6.1 | Accounting for Similarity | 113 |
| 6.2 | Variation and Reduction | 122 |
| 6.3 | Variation and Kind Identity | 126 |
| 7 | Sloppy Sufficiency and Restricted Constitutionalism | 138 |
| 7.1 | Avoiding Overdetermination | 140 |
| 7.2 | The Sufficiency Error | 148 |

| | | |
|-----------|--|------------|
| 7.3 | The Ubiquity of the Sufficiency Error | 155 |
| 8 | Causal and Synchronic Dependence | 162 |
| 8.1 | Causal Taint | 164 |
| 8.2 | A Plurality of Building Relations | 172 |
| 8.3 | Parsimony Versus Pluralism | 179 |
| 9 | The Identity of Synchronic Dependence Relations | 184 |
| 9.1 | Characterizing Dependence Relation Monism | 186 |
| 9.2 | An Objection from Failed Asymmetry | 188 |
| 9.3 | An Objection from Failed Extensionality | 196 |
| 9.4 | An Objection from Differing Relata | 200 |
| 10 | The Metaphysics of Extended Minds | 211 |
| 10.1 | Revisiting the <i>Unrestricted Constitutionalism</i> Premise | 213 |
| 10.2 | Brains in Vats | 226 |
| 10.3 | The Boundaries of the Paradigmatically Mental | 234 |
| 10.4 | Biting a Little Bloat | 242 |
| 11 | Core Selves, Dynamic Selves | 246 |
| 11.1 | Dynamic Selves | 248 |
| 11.2 | Dynamic Bodies | 259 |
| 11.3 | Shared Tissues and Humane Futures | 269 |

Chapter 1

Introduction

1.1 We Are What We Make?

We call ourselves *Homo sapiens sapiens*. According to us, *thinking*, and the mind we credit with thinking, is our distinguishing feature. Art, technology and modern civilization are products of our special minds. These products are modifications of environments we usually consider external to our essential selves. We see ourselves moving freely through these environments as discrete individuals that take our minds with us wherever our brains go.

In my view, however, that's an incomplete concept of our minds and selves. As far as minds are like software, they do not run on brains alone, but on a broader suite of biological, technical, and social constituents. Our particular minds and selves are products of technology as much as they are producers of it.

In fact, technologies are much more like tissues than they are like parts of the environment. I have explored this comparison elsewhere (Hubble, 2009), and will not elaborate much further on it here. But briefly, it should first be understood that technology includes not only those novelties the word popularly brings to mind, but also those ubiquitous, ancient and now-mundane materials of human life stretching back beyond memory from more recent mind-altering creations like steam engines, to the Lighthouse of Alexandria, to stone tools for making, controlling and feeding fires; from Calvin Klein to furs; from concrete to adobe; from Gutenberg to papyrus and clay; from the daguerreotype to the paintings at Chauvet. Considered broadly, technologies are like tissues in

that they are living investments. They are adaptive assets, metabolic capital, structures whose function is to contribute to human organisms' homeostasis and reproduction. Life without technology is not an option for our species. As Barry Allen (2004) puts it,

Strip [modern sapiens] of tools, artifacts, material culture, and the result is not a naked ape thrown back upon its animal nature. It is an ecologically nonviable entity rapidly heading for extinction. A sheltering fabric of artifacts is an existential presupposition of *H. sapiens*, as much as fresh water. (p. 207)

Our extensive use of technologies is not just a side effect of an otherwise technology-free process of getting smarter. As Clark and Chalmers (1998) put it,

It may be that the biological brain has in fact evolved and matured in ways which factor in the reliable presence of a manipulable external environment. It certainly seems that evolution has favoured onboard capacities which are especially geared to parasitizing the local environment so as to reduce memory load, and even to transform the nature of the computational problems themselves. (p. 11)

Our present reliance on technology is but the deepest point we have yet reached in a channel that stretches back definitively through our ancestry. Predating our species, stone tools began a techno-biological arms race in which our species ultimately prevailed. In the course of hominin evolution, the brains that survived were ones well suited to such dynamic technological coupling. Our relatively static genome, rather than coding adaptations to the myriad vicissitudes of social and ecological environments, sows the seeds of a rich and flexible neural architecture that can plug into the whole memetically inherited cultural package, social and technical. In short, our brains have evolved in a technological body. That suggests our minds are technological, and that we ourselves are technological. We are partly what we make. We are *Homo technologicus sapiens*.

As I see it, two concerns motivate this comparison of technology with tissue and ultimately understanding technologies as constitutive of human minds and selves. One is the age-old question of how apparently law-abiding material stuff

can be so arranged as to have a rich mental life that feels like it intervenes in otherwise deterministic causal chains. I will not be providing a complete answer to that question, but I do think that answering the question of *what* material stuff is arranged into rich mental lives will make the remaining problem less confusing than it has seemed. Technology, considered as tissue, helps begin to explain the features of our minds and our selves better than do brain-centric conceptions of minds. However, the ultimate motivation for the internalization of technology into our concepts of self is a larger ethical concern, to which I shall return in the final chapter. The concern arises from the observation that Western culture attempts to live out a confusing and damaging paradox. On one hand, the concept of the brainy self as a discrete and autonomous rational agent dominates our treatment of roles, rights and responsibilities within society. But while we're big on personal freedom, most Westerners take for granted the technologically-extended mode of life we inherit, with its sprawling ecological footprint. Severely myopic but nonetheless human forces drive a steady march of novelty-production which is implicitly endorsed in its entirety as "progress". We remain blind to our captivity in this march even as (or perhaps partly because) we also fabricate superficial choices, like the choice between buying a black device or a white one. We tout ideals of liberty and self-expression, but excuse ourselves for being cogs in a machine when our ways of life do come into question. The concern, then, is that our influence spills well beyond our sense of responsibility. I suspect that is at least partly because we haven't adequately appreciated our boundaries as organisms and corresponding moral agents. In short, thinking we're smaller than we truly are leaves us taking less than full responsibility for our selves. If our very selves—minds and bodies—are bigger than we usually think, our agencies, our freedoms and responsibilities, are greater than we usually think.

1.2 Thesis Outline

Of course, the claim that we are bigger than we think needs to be argued for. However, such an undertaking turns out to be too big here. Instead, I concern myself mainly with two related underlying and yet more general issues. I am

concerned with the extent of the physical basis, constituents, or realizers of complex or “non-categorical” properties—such as a thing’s being of a certain kind or sort, having a certain disposition, having a certain function, or being such that it would or could behave a certain way in certain circumstances. I argue that anything on which an entity depends synchronically is constitutive of that entity. Additionally, the thesis is concerned with the physical reducibility or identity of such properties and their bearers. I argue that identity is the relation between an entity or property and all that it depends synchronically on.

Neither claim is uncontroversial. In favour of *Unrestricted Constitutionalism*, I will argue:

[Differentiability Condition] If constitution is best analyzed simply as synchronic ontological dependence, then anything on which an entity depends synchronically is constitutive of that entity.

[Non-Differentiability of Constituents] Constitution is best analyzed simply as synchronic ontological dependence.

So,

[Unrestricted Constitutionalism] Anything on which an entity depends synchronically is constitutive of that entity.

Additionally, I argue that,

[Reducibility] Identity is the relation between an entity or property and all that it depends synchronically on.

Here, note that I construe “entity” broadly, to include objects, states, processes, events, phenomena and experiences. And I take synchronic dependence to be the relation between an entity or property and those things, contemporaneous with it, without which it would not be what it is (I defend this view in more detail throughout).

To show why *Unrestricted Constitutionalism* at issue to begin with, in Chapter 2, I argue that externalism about minds is contentious, for one thing, because many theorists simply assume that the *Non-Differentiability of Constituents* is false. That is, they tend to assume that *not just anything* on which

an entity depends at some point in its duration is constitutive of that entity. They assume, rather, that among the physical conditions upon which an entity depends for its duration, some of those conditions are constituents, “proper parts,” or “core realizers,” and others are mere background conditions.

To show why reduction is at issue in debates over externalism, I argue in Chapter 2 that externalism about minds presupposes the scientific legitimacy of certain higher-level entities characterized by functional or dispositional properties. Following the way reductionists have typically eschewed functional kinds in order to combat multiple realization-based arguments against reduction, critics of the extended mind worry that whatever model of minds and selves allows us to imagine them as extended will not be a scientifically respectable one, rendering such theses trivial. By contrast, I will later argue that antireductionism is mistaken on different grounds.

Here’s how I will proceed to defend the underpinnings of my own externalist view and brand of reductionism that does not eliminate functional kind (nor other non-categorical properties). In Chapter 3, I introduce a variety of synchronic dependence relations—constitution, composition, and realization—and argue that they are closely related. Here, I join forces to an extent with Karen Bennett, who, across various works (2003, 2008, 2011, 2017), also connects issues of physicalism about minds to wider metaphysical issues about dependence. For starters, Bennett (2017) thinks dependence relations—such as causation, constitution, composition, and realization—are related to each other. We disagree somewhat about how and to what extent they are related, and whether synchronic dependence relations excluding causation are more closely related. I think that composition, constitution, and realization are identical—an issue I explore later.

Beyond sharing certain features, a further reason to think synchronic dependence relations are related is that, as I show in Chapter 4, very similar puzzles arise in each case in which an entity is claimed to depend on physical conditions—whether composed, constituted, or realized by them—and yet claimed not to be identical with those conditions. Specifically, “overdetermination” and “grounding” problems arise when we affirm physicalism but deny identity is the relation between an entity or property and all that it depends

synchronically on. These problems are *prima facie* reasons to accept reductive physicalism.

In Chapter 5 and Chapter 6, I defend a reductive view against the argument from multiple realizability. I account for ubiquitous multiple realizability as a transitive relation in a way that is amenable to reduction. I do so, crucially, without consigning functional kinds to the flames as unscientific, or “unprojectible.” The latter way of countering antireductionism would eliminate those broadly “functionalist” notions of kinds such as minds and selves according to which extended entities count among their members. A broadly reductive functionalism would be required for the inference from our partial constitution by technologies to our partial identity with technologies—which I’ve said is the ultimate possibility I want to clear ground to consider.

In Chapter 7, I continue to defend a reductive view, once again engaging with but departing from Bennett’s antireductive “compatibilism” that seeks to redress worries about overdetermination. I diagnose a widespread antireductive error in thinking about higher-level entities as distinct from physical conditions on the basis of the former’s having differing properties from the latter. The “sufficiency error” has two steps. First, we frequently hold fixed some set of physical conditions occupying an obvious or convenient spatiotemporal region, and we attribute properties to that set of physical conditions for which it is, strictly speaking, insufficient. Those conveniently bounded conditions are only sufficient under certain circumstances for the entity’s “non-categorical” properties—whether kind, sortal, modal, functional, or dispositional properties. In short, we first try to squeeze a property that is too big into a space that is too small. Following Bennett, I call this first step, “circumstantialism,” which would also deny *Unrestricted Constitutionalism*.

Circumstantialism is not, on its own, fatal. But it is trivial that a higher-level entity is distinct from a narrow set of physical conditions not sufficient for it. The second step in the sufficiency error is the further antireductive inference to the wholesale distinctness of higher-level entities from physical conditions. That inference is not valid, as I will show. Rather, if we hold fixed some spatiotemporal region or specific physical conditions, then the latter can be identified with all and only those properties for which it is sufficient

all on its own. On the other hand, if we take as an explanandum some non-categorical property, then the physical conditions fully sufficient to instantiate that property determine the spatiotemporal extent of the entity to which that property really belongs. I thus complete the argument for *Reducibility*, that identity is the relation between an entity or property and all that it depends synchronically on, whether constituents, parts, or realizers.

In Chapter 8, I argue against two ways of collapsing the distinction between constitutive and causal dependence. One is to deny that there is any distinction between them, and the other is to deny that such a distinction is theoretically useful. I argue that constitutive explanations are separable from etiological explanations and that it is useful to distinguish them. I grant, however, a point motivating the second denial. I grant that causal interactions are constitutive of temporally extended entities. But the question of which interactions are etiological and which are constitutive, although relative to the explanatory target, can nonetheless be sensibly answered. So the debate over externalism is not merely terminological.

Now, broadly following the mainstream EM debate, I have formulated my central argument for *Unrestricted Constitutionalism* in terms of constitution. In fact, as we'll see, externalists vacillate in their choice of synchronic dependence locutions, using "constitution," "composition," and "realization," among others. I survey these in Chapter 3, because antireductive arguments also get made in a similarly great variety of such terms. In Chapter 9, I argue that these terms are co-extensive, referring to relations that share all of their features and co-occur, and so, the relations themselves are identical. As predicates, the terms are ontologically innocent with respect to one another. I defend that view against some possible objections raised by Bennett (2011, 2017), who also considers but rejects a similar thesis about dependence relations more broadly. The upshot of my view is that it doesn't matter which synchronic dependence terminology is used to make externalist claims—a potentially surprising thesis that externalists seem to have assumed, but not defended. Together with the *Reducibility* claim, the interchangeability of these terms means that from our partial constitution or realization by technologies, we can infer our partial identity with them.

In Chapter 10, I argue that there is no clear competitor to the notion of constitution as captured by *Unrestricted Constitutionalism*. Specifically, I defend the claim that constitution is best analyzed simply as synchronic ontological dependence. For example, a hunk of marble may be said to constitute a statue in certain circumstances, but any viable notion of constitution applies equally to those circumstances; they also partly constitute the statue. Constitution as synchronic ontological dependence alone does not distinguish the intuitively most salient physical conditions from supposedly background ones.

I further defend this notion of constitution, and synchronic dependence generally—against an objection based on the idea that we could be brains in vats. This possibility can be taken to show that we could be wrong about everything, such that what explains offline and illusory experiences is the same as what explains apparently veridical experiences. So, brains explain experience generally. This line of thinking tends to promote circumstantialism (specifically, “internalism”) about the brain as constitutive, or as the core realizer, of our minds. But I argue that brains in vats do not show that we could be wrong about everything. I thus deny the analogy between offline/illusory experiences and veridical ones. I also argue that the notion of explanation in play in a valid rendering of the objection is not the relevant one.

Having earlier defended functional kinds as projectible, I then argue that there is no *prima facie* reason that biotic–abiotic composites like human–tool couplings cannot be members of projectible kinds. Now, it’s true that we might sometimes eliminate certain equivocal or unprojectible functional kinds and corresponding properties from our theories. On the other hand, whenever apparently non-categorical properties—like those had by intentional states—figure in successful predictions, they should be taken seriously as objects of explanation. But we ought not to presuppose the boundaries of the physical conditions sufficient for them. I conclude Chapter 10 by arguing that the externalist should bite the bullet with respect to the apparent counter-intuitiveness of the view. The “bloat” objection is roughly that on this view minds might be constituted by anything and everything. While no such extreme conclusion follows from my view, I account for the counter-intuitiveness of my argument for externalism in terms of differing epistemic salience within the

ontological dependence bases of higher-level things.

I argue, albeit briefly, in Chapter 11 that there is a close relation between our minds and our selves as targets of praise and blameworthiness. Most plausibly, we *are* minds, which are identical with extensive physical conditions that include our bodies and technological carapaces. In every case of psychological continuity there is also circumstantial continuity constituting a person's being of a kind with their past and future selves. I argue that phenomenological evidence of our personal boundaries suggests that our bodies are dynamically extensible. The upshot is that our particular mental lives would fail to be duplicated if our biological duplicates were to be transplanted into different technical carapaces and wider technological circumstances.

Let me pause for a moment to reflect on the nature of this project. This is obviously a very different approach to the question of externalism than usual. Admittedly, the view that minds extend because it takes more than the thing to which we attribute non-categorical properties to instantiate those properties, might come as a sort of disappointment, an almost pyrrhic victory on behalf of externalism. Indeed, my liberal view of constitution *is* rather deflationary in the sense that it is a wholesale rejection of our intuitive restrictions on constitution and other synchronic dependence relations, which I lump together. But I think it's the right view, and defending it also goes a long way toward establishing a coherent physicalism about minds in the face of the putative mind-body problem, the apparent mystery of consciousness, and general puzzles about the relation between higher and lower-level objects of science. Moreover, settling conceptual arguments about what sort of things minds depend on allows us to move on to what might be more practical considerations that come fully into view.

Thus, in Chapter 11, I gesture at a few interesting and morally salient subsets of the complete constituents or total realizers of our minds and selves from the rest. Technologies are on the inside of each of these. Specifically, our technological parts are vast relative to our biological parts, in them we overlap with one another, and they are among the conditions of our experience that we can engineer and are responsible for, which has moral implications. Yet these are parts of ourselves of which we are often unconscious, and when we

are conscious of them, we are not conscious of them as parts of ourselves.

Wrapping Up

I am primarily occupied herein with the very notions of parthood, realization, constitution, and the like, and the reducibility of things to their parts. But in the end, I suggest that to be a person, on my view of the metaphysics, is to be woven into a radically joint venture which is fundamentally technological. Our particular mental lives would fail to transplant into bodies lacking these appendages; in transplantation, minds—not necessarily our own—would be made anew. Thus, tissue-like technologies, as life-supports, may be proper objects of moral concern. I suggest that the beauty of the fact that we can share huge parts of ourselves and our minds means that we could cultivate more life—more mental life—out of less stuff, and that less stuff of certain kinds would make for more flourishing all around. While it's not obvious that anything on which an entity depends synchronically is constitutive of that entity, in this chapter, my aim was to summarize why I think so, what is controversial about thinking so, and how I'm going to tackle those controversies in the coming chapters.

Chapter 2

The Difficulties with Extended Minds

Our minds are big. Not by having big ideas or contemplating the enormity of the universe. Rather, minds are big because, supposing minds are made of *something*, what minds are made of, and thus where they are to be found in physical space, extends beyond brains. Since my central concern for the better part of this thesis is with the question of whether that's true, the aim of this chapter is to understand what is controversial about thinking so. I argue that existing arguments for externalism appear to make plausible counterfactual dependence claims, but that they otherwise preach to the section of the choir that shares other key intuitions. These are intuitions about synchronic dependence relations which operate as implicit premises in inferences from the dependence of a mental state or cognitive processes on something extraneural to the latter's being part of that state or process. By and large, externalists haven't set themselves to the difficult task of explaining and defending the notion of constitution or other synchronic dependence relations they're working with. Nor have they defended the functionalism that the substantivity of externalism also requires. To the contrary, there is a trend toward casting vague aspersions in the direction of "metaphysics" while nonetheless being implicitly bound up with particular metaphysical commitments.

To show this, I discuss one particularly popular version of externalism, or the idea that our minds are bigger than our brains. That is the "extended

mind” (EM) hypothesis, as set out by Clark and Chalmers (1998).¹ I offer a new diagnosis of the dispute between this mainstream form of externalism and its critics; I argue that the deeper battle lines have not been accurately drawn by the major parties to the debate, and that their disagreements are special cases of divided intuitions over metaphysical issues that run through a broad swathe of philosophy. To show this, I discuss one particularly popular version of externalism, or the idea that our minds are bigger than our brains. That is the “extended mind” (EM) hypothesis, as set out by Clark and Chalmers (1998).² I offer a new diagnosis of the dispute between this mainstream form of externalism and its critics; I argue that the deeper battle lines have not been accurately drawn by the major parties to the debate, and that their disagreements are special cases of divided intuitions over metaphysical issues that run through a broad swathe of philosophy.

First, in Section 2.1, I highlight a deep divide between externalists and their critics over the general nature of higher-level entities and properties as objects of science. In Section 2.3, I highlight a related divide over the nature of higher-level entities and properties’ dependence on lower-level ones. Now, although I use debates over EM as a primary point of reference and a convenient way to probe those deeper issues, I pause in Section 2.2 to examine Hurley’s (2010) taxonomy of externalist positions in order to point out that the same issues are crucial to the question of externalism more broadly. As I indicated in Chapter 1, the same issues—the status of higher-level kinds as objects of science and their dependence on lower-level ones—lie at the heart of the question of *our* boundaries, which is why I want to bring them to the fore through the present discussion of externalism.

This is not intended as a comprehensive survey of the literature and the increasing interaction between the extended mind literature and emerging branches of cognitive science. The bulk of the thesis is spent taking a rather large step back to conceptual matters, and deals with the relevant literature on those matters. The basic idea to be put forward is that considerations of metaphysical sufficiency drive one to conclude that non-categorical prop-

¹ Every reference to Clark and Chalmers in this chapter is to Clark and Chalmers (1998).

² Every reference to Clark and Chalmers in this chapter is to Clark and Chalmers (1998).

erty bearers reduce to their properties, which reduce to their wide realizers. Extension, or embeddedness and intertwinedness, is the norm for entities bearing non-categorical properties, including mental states and cognitive systems, rather than the exception. The selective nature of the present survey is to be expected, given the much broader scope of the thesis as a whole. The point I wish to draw out is simply that those larger issues are at issue within the extended mind debate.

2.1 The First Battle Line

Clark and Chalmers use two examples to make their argument that minds extend beyond the skull. The first example is that of the process of rotating a Tetris game piece on the screen of a computer or game console to test for the best fit of the piece within the play area, a process that, if it was done in the head by mentally “rotating” the piece, would be considered a “cognitive” process. Clark and Chalmers suggest that the analogous and partly externalized process is still cognitive. The second example is that of a mental state, such as having a belief. A belief is a dispositional state that we say is about something, for example, a museum, either because of its causal history in relation to the museum, or because of the role that the belief would play, given certain (perhaps museum-related) inputs, in driving certain (perhaps museum-related) behaviours. For example, Inga desires to see a certain exhibit at the Museum of Modern Art, and the belief that the museum is on 53rd Street impels her there. Suppose now that a functionally equivalent disposition—to go to 53rd Street given the desire to see the same exhibit—is implemented by a somewhat different state of affairs. Otto, who has a poor memory for such details as addresses, writes the address in a notebook. Clark and Chalmers say that the notebook, in relation to Otto, partly constitutes his belief about the location of the museum, since it drives the same or very similar behaviour in the same or very similar circumstances that Inga’s “normal” belief does.

Clark and Chalmers discuss both cognitive processes and mental states. Now, no one seems to think the argument hinges on whether it is put in terms of cognitive processes, or mental states, or cognitive states, or mental processes,

or some combination thereof.³ In any case, then, extended cognitive processes or mental states, taken together with some plausible relation between these and *minds*, yields externalism—the idea that our minds are bigger than our brains. Candidate mind–cognitive relations might be that minds are i) identical with cognitive processes and/or mental states, or ii) wholly constituted by cognitive processes and/or mental states, or iii) partly constituted by cognitive processes and/or mental states. Now, I think it’s fair to assume, seemingly along with everyone else, that one way or another, whatever is true of cognitive processes and mental states is true of minds. So I also think the argument stands or falls independently of precisely which mind–cognitive-process relation is invoked.

But a bit more care must be taken here. First, there is a possibility of equivocating on one or more of the notions in play in such a way as to invalidate the inference from cognition to mind. Or, with clarified premises, there may be a fallacy of composition. To see the problem, suppose that consciousness is a definitive feature of minds. Suppose it’s part of what we mean by “mind.” Next note that we are confident that minds depend on brains. In spite of our usual mind–body problem—that there’s some sort of gap between neurological goings on and consciousness—neurological goings on seem like good candidates for a complete physical explanation of consciousness if there could be one. After all, the intuition goes, wherever you have a (healthy) brain, you have a mind, and wherever there is no brain—at least given the present state of AI—there is no mind. We imagine brains in vats to have mental lives and so we imagine that brains are sufficient and perhaps even nomologically necessary, or at least necessary in practice, for minds. As a result of this set of intuitions, there may

³ If there was a distinction between the cognitive and the mental (suppose there could be something cognitive where there is nothing mental, or vice versa), arguments for EM could be run in terms of just cognitive processes or mental states, and might fare differently. But the distinction is not at all clear. First, I take it that states are temporal parts of what are in reality dynamic processes, so I take it that mental states are parts of mental processes, and cognitive processes can be modeled as involving “cognitive” states, and that there would be nothing terribly unusual about speaking in terms of mental processes and cognitive states. In fact, some parties to the debate, including critics, switch freely among those terms (e.g. Adams and Aizawa, 2010, pp. 69–70). At any rate, second, no one appears to endorse extended cognitive processes while rejecting extended mental states, or vice versa.

be something fishy about involving non-living parts in whatever we take to be responsible for minds, ignoring, as intuitions often do, the fact that at chemical or atomic grain-level of description, brain parts are non-living, too. The worry, then, is that the EM theorist is arguing:

[**Extended Cognition**] There are cognitive processes (or mental states) that are partly constituted by extraneural entities.

[**Mind–Cognitive Relation**] Cognitive processes and mental states are constitutive of minds.

So,

[**Extended Mind**] Minds can be partly constituted by extraneural entities.

It may seem to the critic of the *Extended Mind* thesis that whatever notion of what it is to be a “cognitive” process or “mental” state that is inclusive of extended ones must pick out a different notion than the one in a premise the critic would accept relating the cognitive and mental to minds. That equivocation would render the inference from *Extended Cognition* to *Extended Mind* invalid. On the other hand, avoiding the equivocation and invalidity by injecting the broader notion of the cognitive and mental into the *Mind–Cognitive Relation* premise, the critic might think such a claim is strictly false—that minds are constituted by only the conscious *part* of the extended cognitive process or mental state. Finally, if the *Mind–Cognitive Relation* premise were to be clarified in that way and made plausible in the eyes of the critic, then the validity of the conclusion would then demand the further false generalization that “if M is constituted by part of $P+$ (P), then M is constituted by $P+$.” That is a fallacy of composition; it attributes a property (the relational property of constituting M) to a whole ($P+$), that belongs to one of its parts (P).

A similar point would apply even if we take some defining feature of minds to be a defining feature of the cognitive and mental themselves. An analogous objection migrates upstream to auxiliary arguments for extended processes and mental states. Here, an equivocation between 1) a notion of an extended state as a species of “belief”, and 2) “belief” as a species of the mental would

render the inference from the specific to the general invalid. Again, the critic might only want to allow that *part* of the extended process of believing or state of having a belief should be called cognitive or mental. Hence we see the critical response, characterized by Clark (2010a), that, “all Otto actually believes (in advance) is that the address is in the notebook” (p. 46). And, again, the inference from a mere *part of* P ’s being a G to P ’s being a G is not valid without being committed to a fallacy of composition.

In short, a critic of EM may not be inclined to accept *Extended Cognition* and the *Mind–Cognitive Relation* at the same time. In fact, the critic who assumes that brains are sufficient for minds will tend to think the line of reasoning for the premise he accepts tells against the other premise. The form of the objection is that what could make something cognitive or mental in a broader sense is different than what makes for minds as distinguished by some other feature.⁴ Whatever the distinguishing feature of minds, the point at present is twofold. First, the EM theorist must apply their theory of what makes something a belief and what makes belief mental consistently in order to avoid equivocating when claiming that extraneural entities can partly constitute cognitive processes or mental states. Second, and crucially, it is no good to avoid equivocation by changing *all* of the important terms to suit the inferences such that whatever is extended is unlike what the rest of us mean by “mind.”

So what notion of the cognitive and mental are Clark and Chalmers working with? It is an unabashedly functionalist notion of minds. Functionalism about minds takes them to be like other things characterized by what they do, such as cars or mousetraps. What does it take to be a car? Something like the ability to carry goods or passengers on roads. One might subdivide cars by specifying further functionality or interoperability. Rail cars for railroads, and toy cars for toy roads. Trucks carry more stuff. We can continue in this way, perhaps specifying the precise functionality of a kind down to a class of things

⁴ I’ll return later to the assumption that motivated the particular version of the objection as I introduced it, which took consciousness to be the distinguishing feature of minds. That was the assumption that brains are sufficient for consciousness. By Chapter 10, we’ll see why the externalist should deny as much.

that are indiscernible. But we don't usually do so. We, and other (non-human) cognitive agents, classify things more generously, and it seems both quite natural and practical to do so. For example, there are all sorts of reasonably good predictions one would miss the benefit of being able to make if one utterly failed to register any similarity between a Honda Civic and a Toyota Corolla. Likewise, functionalism about minds answers the question of what a mind is by reference to what minds do. Like cars, minds are identified by what they do. And like cars, not just anything counts. Functionality constrains design, and the more functional specificity, the more the constraint. Despite the rhetoric of early functionalists, neither minds nor reliable cars could be made of Swiss cheese, a point Block (1997) has called the "Disney Principle." But they could be made of anything that could get the relevant job or jobs done, the idea goes, and that might be a lot of different things. Unless part of what we mean by "cognitive" is "fully brain-based," then membership in the cognitive kind, or family of kinds, is at least in principle open to processes and states with extraneural entities. And we should not consider being brain-based a criterion for being "cognitive." The denial of the possibility that extraneural entities can constitute cognitive states or processes based on the requirement that cognition be brain-based, in the absence of further reasons for *that* requirement, would beg the question, as others have pointed out (see Menary, 2010; Clark, 2010b; Hurley, 2010). Rather, things like discrimination, awareness, flexible action, intelligent behaviour, goal-seeking, harm avoidance, reason, information processing, or perhaps consciousness, are what we seek to explain. We find brains at the center of all that, of course. But most of us would agree the latter is a fact we have discovered; it is no part of our concept of, say, "reasoning" that it should occur in brains alone—an omniscient and omnipresent creator is thought to reason, apparently without embodiment. Being reasonable, as is typical of functional properties, is identified by what being reasonable does, not by how it does so.

Clark and Chalmers thus presented their argument using what has become known as the "parity principle." The parity principle is the idea that a process or state should be considered "cognitive" or "mental," if some functionally equivalent process or state in the brain would be considered cognitive or mental.

Clark says Otto’s notebook is a “resource whose encodings, at appropriate run-time moments, inform Otto’s behavior in the way characteristic... of dispositional beliefs.” And, Clark claims, this is “all that matters” (2010b, p. 89). However, Clark’s is an appeal to the obviousness of functionalism, not a defense of it. And insofar as functionalism seems incompatible—in ways we’ll see in Chapter 4—with other plausible ideas (for example, a physicalist’s typical commitment to the causal closure of the physical world), functionalism about minds is not a foregone conclusion.

One way of rejecting functionalism about minds is to deny that functional kinds belong in a properly scientific ontology. Thus, in contrast to Clark’s functionalism, Adams and Aizawa (2010), EM’s “most outspoken critics” (Walter and Kyselo, 2009, p. 277) continue to defend their (2001) claims. One is that, “A first essential condition on the cognitive is that cognitive states must involve intrinsic, non-derived content” (2001, p. 48). Another is that, “The second necessary condition is a condition on the nature of processing” (2001, p. 51), and “The cognitive may, therefore, be assumed to be like other natural domains, namely, the cognitive must be discriminated on the basis of underlying causal processes” (2001, p. 52).

Aizawa and Adams (2005) spell out the derived/underived distinction involved in the first condition. Some representations, such as mental images of trees, “derive” their meaning from purely non-representational things. Others, such as written sentences about trees, “derive” their meaning from mental representations of trees, themselves underived. This condition is contentious and arguably doesn’t work as Adams and Aizawa intend it, for reasons put forward by others that I needn’t explore further (see Clark, 2005, 2010b, 2010a). But here, I want to point out that it’s really not clear how the un-derivedness criterion and the mechanistic criterion fit together. To see why, we must explore the second condition in more detail. First, note that the second condition is reminiscent of a point Kim (1992b), among others, has tried to make. Kim reasons,

instances of M [some putative mental kind] that are realized by the same physical base must be grouped under one kind, since *ex hypothesi* the physical base is a causal kind; and instances of M with different realization bases must be grouped under distinct

kinds, since, again *ex hypothesi*, these realization bases are distinct as causal kinds. Given that mental kinds are realized by diverse physical causal kinds, therefore, it follows that mental kinds are not causal kinds, and hence are disqualified as proper scientific kinds. Each mental kind is sundered into as many kinds as there are physical realization bases for it, and [psychology] as a science with disciplinary unity turns out to be an impossible project. (p. 18)

Kim attempts, unsuccessfully I think, to distinguish his (1992b) view from eliminativism. Eliminativism, or eliminative materialism, “is the radical claim that our ordinary, common-sense understanding of the mind is deeply wrong and that some or all of the mental states posited by common-sense do not actually exist” (Ramsey, 2003). Though I’ll evaluate Kim’s view in detail in Chapter 5, the sundering Kim refers to really does eliminate the sundered as a *scientific* kind, since Kim thinks it lacks projectibility. I thus take this line of thinking to be broadly eliminativist.

Notice that a view like Kim’s does not so much deny any of the premises of the arguments we’ve seen as it changes the subject to something putatively more amenable to rigorous scientific study. The issue changes from being about how to account for what we now consider to be cognitive states and processes, to whether the very things we’re trying to account for are really things that should be accounted for at all (and, if not, the rather vexed question of what we *should* be trying to explain). At any rate, to deny that intelligence, for example, is a kind at all is not, or not directly, to deny that a particular “intelligent” behaviour can partly depend on extraneural entities. No one is denying that Otto-and-his-notebook exists, and depends on the notebook. They deny, rather, that Otto-and-his-notebook is a member of a projectible kind.

Likewise, Adams and Aizawa (2001) suggest that extended-Otto’s disposition is not a belief because it ostensibly—at least at some grain-level of description—doesn’t work exactly the way brain-based beliefs work. Adams and Aizawa say,

Human memory displays a number of what appear to be law-like regularities, including primacy effects, recency effects, chunking effects, and others. Further, human memory capacities are task sensitive. Memories for images, faces, smells, and lists of words vary

in properties of formation and recall. Consider, then, the range of tools humans use as mnemonic aids. There are photo albums, Rolodexes, computer databases, strings around the finger, address books, sets of business cards, bulletin boards, date books, personal information managing software, palmtop computers, hand drawn maps, and lists of “things to do.” What are the chances of there being interesting regularities that cover humans interacting with all these sorts of tools? Slim to none, we speculate. There just isn’t going to be a science covering the motley collection of “memory” processes found in human tool use. (2001, p. 61)

If believing involves remembering, and brain-based remembering processes can be more finely individuated as a kind by their other effects, then believing can be sundered into subkinds. But it is odd that Adams and Aizawa imagine the relation of a would-be science of extended cognition to *a single* extended cognitive kind to be one-to-one. Rather, a “single” scientific discipline—unified in terms of method, grain-level of description and/or range of phenomena investigated—usually posits many kinds. Just as “memories for images, faces, smells, and lists of words” vary in properties, so might various human–tool couplings, each of which may be projectible. Something no one to my knowledge has wondered aloud is whether a particular set of oft-encountered human–tool couplings, taken as extended systems, might show *more* regularity as a kind than their non-extended sub-systems (biological humans). Given our rather great degree of neural plasticity, it would be unsurprising if more generalizations could be made about iPhone-users than about *Homo sapiens*, not just because all iPhone-users are partly constituted by *Homo sapiens*, but because iPhone-users are also partly constituted by mass-produced bits that are *very* alike as appendages. A similar line of reasoning would seem to apply to the states and processes of these systems. And the properties of these extended systems may show a remarkable degree of “homeostatic clustering” (see Boyd, 1999)—the tendency of instances of a self-reproducing kind to share multiple properties—since iPhone users don’t so much give birth to Achulean axe-makers as they do to more iPhone-users.

I said above that it’s not obvious how the two criteria set out by Adams and Aizawa (2001) are meant to work together, and now it should be clear why. It’s not as if one can define the class of mammals first by their being warm-blooded

and nursing, and then by their being divisible into marsupials, primates, canines, felines, rodents, ungulates, lagomorphs, equines, chiropterans, and the rest, as if to guarantee that any warm-blooded nursing animals we might later find would fail to be mammals if they failed to be members of a known mammalian subclass. Likewise, to require that all cognitive processes involve non-derived content, and to further *require* that “the cognitive must be discriminated on the basis of underlying causal processes” (2001, p. 52) looks ad hoc at best. The specific effects of human memory can’t be what makes beliefs *mental* and remembering *cognitive*, since beliefs are not the only mental states, and remembering is not the only cognitive process. The specificity that would rule out Otto’s belief as being of a kind with brain-based beliefs excludes the rest of cognition, too. Just as planning needn’t be functionally equivalent with remembering in order to count as cognitive, it is not clear why extended remembering would have to be functionally equivalent to “normal” remembering to count as cognitive. And of course, at some even finer grain-level of description, Inga’s remembering and mine don’t work *exactly* the same way either. So it’s not surprising that, as Menary (2010, p. 7) points out, other arguments for EM rely less on parity between extended and non-extended processes and states.

Yet “remembering” remains a clearly functional characterization of a kind, as does “problem solving,” etc., and perhaps even “involving non-derived content.” With or without perfect parity between extended and putatively non-extended processes or states, the notions of cognition and mentality in the argument for *Extended Mind* are functional.

I’ve set eliminativism and functionalism up as competing alternatives; we’ll see why in more detail soon enough. In Chapter 4, we’ll see what motivates the eliminativist line of thinking and, in Chapter 5, why eliminativism is neither a necessary nor a clearly viable road to take. Briefly, here’s how I see it. Functionalism and eliminativism are incompatible views about kinds and the appropriate objects of scientific enquiry. Either functional kinds can be projectible, or they cannot. If they can be projectible, then we must countenance at least some functional kinds as scientific kinds. I will argue in Chapter 5 that they are projectible and we must countenance functional kinds as scientific kinds. But if functional kinds can be scientific, and functional

kinds are basically open to anything that does the job, then there's no way to ensure in advance that a particular functional kind is limited to the existing card-carrying membership. Supposing we agree or can clearly stipulate what, for example, "intelligence" does in some paradigm cases, it is nomologically possible that something non-brain-based or not-fully-brain based is intelligent.

Although a defense of functional kinds is yet to come, suffice it here to draw a few preliminary points from the discussion so far. First, with the elimination of functional kinds goes our present cognitive science. Psychology and cognitive science, among other "special sciences above the level of physics," like biology, now trade in functional kinds (see Fodor, 1974). So assuming for now that we do indeed want to explain the likes of "remembering", we ought to accept that remembering is as remembering does, however it does so.

Second, the significance of EM hinges on the merit of functional kinds as scientific kinds. I briefly raised questions about each of Adams and Aizawa's criteria on the cognitive simply to show what is at issue in the debate over EM. Critics of EM don't have an easy task of stating what is common to all and only cognitive states, as if EM theorists had just misunderstood how to apply a kind term that is perfectly clear to everyone else. Of course, Adams and Aizawa (2010) acknowledge the lack of clarity in the concept of the cognitive and its extension. They say, "in all fairness to Clark and other extended mind theorists, it must be admitted that one of the shortcomings of contemporary cognitive psychology is that there is no well-established theory of just exactly what constitutes the cognitive" (p. 68). But neither do EM theorists have an obvious case. That's because, on the other hand, if proponents of EM intend their thesis to be empirically interesting and relevant to science, as they do, then claims about the instantiation of *F* *assume*—if they fail to state—that *F* is a member of a kind about which useful things can be said. It's not very interesting to claim that my pinky and the equator are parts of an *F*, namely, my-pinky-and-the-equator, that is of some kind, *G*, maybe fingery-things-plus-circumferences-perpendicular-to-rotational-axes, that nothing very useful can be said about, and it would be downright odd to worry about the constituents of instances of a kind like phlogiston that doesn't name anything we now take to be real. The significance of the EM hypothesis thus hinges on the

legitimacy of functional kinds as scientific kinds. Again, eliminativism isn't a direct counterargument; it's just that if functional kinds aren't scientific, the *Extended Mind* thesis lacks teeth. It appeals to a kind that is not useful to the overall project of discovering the world and dealing with it.

Finally, notice that the parties to the debate do not disagree on whether there are any extended processes because some empirical data is lacking. I take it that if the parties to the debate had before them a particle-by-particle, field-by-field, instant-to-instant account of the event of Otto's picking up his notebook and making his way to the museum, they would still disagree about the relation of the notebook to Otto's mind. While the parties to the debate disagree about the mark of the cognitive, Adams and Aizawa's (2001) mechanistic criterion highlights the possibility of disagreeing, not just about *which* standard makes something a member of a *given* kind, but about *what sort* of standards make something count as a member of *any* kind. In a closely-related discussion of the functionalist doctrine of multiple realizability, Bickle (2016, sec. 3) concludes:

What is at stake here should not be underemphasized: nothing less than one of the most influential arguments from late-20th century Anglo-American philosophy, one that impacts not only the philosophical mind-body problem but also the relationship between sciences addressing higher and lower levels of the universe's organization.

2.2 The Varieties of Externalism

I'm arguing that externalism requires further defense and hinges on larger philosophical issues, but I want to pause for a moment to reflect on the scope of "externalism" in that claim. I have made some fairly sweeping claims about "externalists," whom I defined at the outset of this chapter as those who think that cognitive processes or mental states can depend synchronically on (or can be constituted by) extraneural entities, or more than brains. This obviously includes a variety of thinkers, and indeed, I am trying to help myself to some general points about the state of a family of debates while nonetheless using Clark and Chalmers (1998) as my main point of departure. While it is the *locus*

classicus of the EM view, it is but one species of externalism. Presently, I'll address the "varieties of externalism," as famously outlined by Hurley (2010) in an essay by that name. I'll explore Hurley's taxonomy in order to explain why externalists of every variety need a plausible account of synchronic dependence and the merit of cognitive or otherwise functional kinds.

Arguments for externalism have historically been made in various ways and have sometimes made reference to the extendedness of the *vehicles* (e.g. Clark and Chalmers, 1998) and sometimes to the externality of the content of mental states (Burge, 1986; Putnam, 1975b). And some have referenced qualitative mental states, such as "enactivist" motor theories of perception (e.g. Noë, 2010; Thelen and Smith, 1996), and others (e.g. Clark and Chalmers, 1998) have referenced intentional states (like Otto's belief) or cognitive/behavioural capacities (playing Tetris). Hurley (2010) organizes these apparent varieties of externalism in a two-by-two matrix. She contrasts "what externalism" with "how externalism," and externalism about qualities with externalism about content. We thus have the following possible positions, informal expressions of which I take from Hurley.

[What-Content Externalism] "What-content explanations appeal to causal relations between an organism and its environment, understood to provide constitutive rather than merely causal explanation of mental content type, even though in many illusions and hallucinations, tokens of content types do not participate in the type of causal process that is nevertheless taken to explain, constitutively, their content" (Hurley, 2010, p. 106).

[Content-Enabling Externalism] "Most discussions of the 'extended mind' concern extended cognition—externalism about vehicles of intentional contents. They consider enabling explanations that cross internal–external boundaries, including body, environmental objects, or both" (Hurley, 2010, p. 127).

[What-Quality Externalism] "What the world is like can be part of what explains what experience is like" (Hurley, 2010, p. 145).

[Quality-Enabling Externalism] "What the world is like can be part of what enables us to experience what it is like" (Hurley, 2010, p. 143).

Hurley (2010) herself defends each of the forms of externalism she dis-

tinguishes. Moreover, her reasons for doing so often cross-cut those very distinctions, for example, when she denies that illusions and hallucinations “support what-quality internalism and enabling internalism,” saying that, “it’s no objection when explaining quality type that the type-explanatory external factors are absent in illusory cases, any more than it’s an objection when explaining content type” (p. 141). Hurley proposes distinctions between positions that then, by her own lights, stand or fall together.

On the other axis, Hurley’s distinction between what and how externalisms relies on a distinction between “enabling explanations” and other explanations one would think must be “non-enabling”. This is most obvious in her characterizations of *What-Quality Externalism* as compared to *Quality-Enabling Externalism*. Note, too, that the distinction cannot be a matter of differing grain-levels of description. My desire to eat fruit may be completely explained by my brain or completely explained by my brain, other tissues, fruit, etc. These are substantively different explanations (if one is true, then the other is false) at roughly the same grain-levels of description. And supposing the brain-only explanation is true, that would not make an explanation at the grain-level of neurons false. Nowhere does Hurley (2010) cash out what she means by “explanation,” but it appears to be something in the vicinity of what others have variously called “metaphysical explanation”, “ontological dependence,” “supervenience,” or “grounding”—very roughly, the relation that obtains between a thing and all the stuff (potentially describable at various grain-levels) that it takes to get the thing in question going and/or to keep it going. And, at times, Hurley (2010) again seems to tie the fates of *what* and *how* explanations together, saying that if “what enables qualitative experience is ongoing embodied interactions of the brain with environment” (p. 141), that is, if *Quality-Enabling Externalism* is true, then, “the extended what-quality explanation and the corresponding extended quality-enabling explanation would converge in the natural world” (p. 142). The implication is that *Quality-Enabling Externalism* entails *What-Quality Externalism*. It is certainly not clear that anyone wants to hold those positions apart.

Further, broadly citing Robert Wilson (2004), Hurley also says more generally, “how-externalism may require what-externalism” (p. 105). If anything,

however, the what/how distinction *appears* to get some traction when one contrasts the historical claims Hurley categorizes as *What-Content Externalism* and *Content-Enabling Externalism*. That is, at face value, there is a difference between the “semantic” externalism of Putnam (1975b) and Burge (1986), and the “active” externalism of Clark and Chalmers, since there is an obvious distinction to be made between Otto’s notebook and the museum his putative belief is about. Semantic externalism is the idea that the museum is constitutive of beliefs about it. It is silent on the status of Otto’s notebook. By contrast, active externalism claims Otto’s notebook is a constituent of his belief about the museum. It is silent as to any dependence relation the belief may have to the museum. It certainly seems as if the notebook and the museum, even if they are both determinative of the belief in some sense, are determinative of it in different senses; Clark and Chalmers explicitly distinguish their active externalism about “vehicles” from semantic externalism about “content”. So one possibility is to include the notebook in the enabling explanation of Otto’s belief, as Clark and Chalmers do, but still not extend the enabling explanation as far as the museum. One might claim instead that the museum stands in a causal relation to extended Otto. So it might then be tempting to think that Hurley means to divide enabling explanations from non-enabling explanations in the same way that we might divide metaphysically explanatory, ontological dependence, supervenience, or grounding relations into causal and constitutive species. Roughly, causal explanations are those which specify the bits we’d wiggle to influence something prior to its occurrence. Roughly, constitutive explanations are those which specify the bits we’d wiggle to alter something during its occurrence.

But Hurley explicitly says that “the what–how distinction doesn’t align cleanly with either a causal–constitutive distinction or an external–internal distinction” (p. 106). Now the trouble is, it’s hard to see how one can have it both ways. What-explanations can’t be distinguished from enabling explanations unless by the causal/constitutive distinction she takes to be a separate issue. One can’t maintain that features of the environment constitutively explain mental content or quality types, while excluding those features from the enabling explanation, without consigning the meaning of “enabling” to

further obscurity. In other words, if the what/how distinction doesn't align with a causal/constitutive distinction, then it's entirely unclear how one could be a what-externalist without also being an enabling externalist. Instead, on closer inspection, the difference between an account that takes the museum to constitute Otto's belief, and an account that doesn't, is not a difference between non-enabling explanations and enabling ones, but between enabling explanations involving rival notions of constitution.

2.3 The Second Battle Line

We can now draw the second battle line, although it might cut across a corner of the externalist camp. I'll first deal with the matter of labeling because it will help to see the terrain. Positions that disagree over *What-Content Externalism* because of differing notions of constitution, though not best characterized as "what" and "how" externalisms, might better be characterized as "unrestricted" externalism and "restricted" externalism, following their views about constitution, particularly whether it is restricted or not. Define unrestricted externalists as those who accept *What-Content Externalism* because they accept, among other things, as I do, that,

[Unrestricted Constitutionalism] Anything on which an entity depends synchronically is constitutive of that entity.

Then, one way of denying *What-Content Externalism* is to deny *Unrestricted Constitutionalism*. Call that restricted constitutionalism. Here, then, is how externalisms might be divided by notions of constitution. Recall that to arrive at the *Extended Cognition* claim requires further argument. First, externalisms differ in the examples of processes or states and extraneural entities to which they appeal. They make claims of the form: "if not for x (an extraneural entity), then not y (a mental state or process)", where x and y temporally overlap, and fill these claims in differently. On their own, these differences might be shallow, except for precisely how the argument from the dependence claim to *Extended Cognition* is cashed out. Going in for *Unrestricted Constitutionalism*, *Extended Cognition* would straightforwardly follow from the truth of a cognitive process or mental state's synchronic dependence on an extraneural entity. On

the other hand, one might think that constituents belong to a narrower subset of things on which a process or state depends synchronically—a subset that meets some additional criteria. And this notion of constitution may correspond more closely to our ordinary one. So one might implicitly deny *Unrestricted Constitutionalism* by supplying a notion of constitution according to which the *particular extraneural x s to which one's example appeals* are constitutive of y .

To my knowledge, however, no one has supplied such a notion. The second thing to clarify, then, is what parts of the terrain just labelled are actually occupied. Semantic externalism and enactivism almost certainly need *Unrestricted Constitutionalism*, or something like it. This, I take it, is what analogously leads Wilson (2001) to make the general remark that what he calls “wide realization” “makes most direct metaphysical sense of the widespread recognition that a range of mental properties are not individualistic... and a view that externalists should readily agree with” (p. 6, my emphasis). But it is difficult to know whether, for example, the active externalism of Clark and Chalmers, while clearly a form of externalism, subscribes to unrestricted or restricted constitutionalism. Interestingly, Clark and Chalmers do seem to supply additional criteria for constituents of cognitive systems. Summarizing the original (1998) position, Clark (2010c) offers, “a rough-and-ready set of additional criteria to be met by nonbiological candidates for inclusion into an individual’s cognitive system.” These are that such a candidate “be reliably available and typically invoked”, “be more or less automatically endorsed”, “be easily accessible as and when required”, and has been “consciously endorsed at some point in the past and indeed is there as a consequence of this endorsement” (p. 79). These (1998) criteria have spawned their own discussions (e.g. Sterelny, 2004), about the functional parity of notebooks and biological memories in terms of trustworthiness, and so forth. However, these kinds of discussions are relevant to deciding the extent of cognitive constituents only if restricted constitutionalism is viable. I’m ultimately going to argue that it isn’t, that it is unmotivated and sloppy.

More importantly at this juncture, it’s not clear that the above criteria have been consistently intended as a positive stance on restricted constitutionalism. For one thing, in the space of a couple of essays in a single volume (Menary,

2010), let alone in his larger set of writings, Clark uses such locutions as the following. The mental or cognitive is said to be “grounded in” (p. 43), to “supervene on” (p. 45), to be “composed of” (p. 48), or to “depend on” (p. 95) nonbiological stuff. Going the other way, this stuff is variously referred to as “physical vehicles of content and of cognitive processes” (p. 46), “aspects of an integrated whole” (p. 52), “candidates for inclusion into an individual’s cognitive system” (p. 47), “co-opted into playing a real cognitive role” (p. 53), “partially constitutive of mind,” and most often said to be “part of” some “implementation” (p. 52), “physical realization” (p. 84), or “supervenience base” (p. 86). And Clark is by no means the only one to drift among these locutions. This drift is perhaps understandable. We’ll see in Chapters 3, 4 and 9 that those relations which have received the most attention in the analytic metaphysics literature have received mostly separate treatment and have spawned their own literatures. While the apparent complexity and contentious state of the debates on these relations makes the lack of clarity within the EM literature understandable, the conclusion EM theorists want nonetheless requires a sensible notion or notions of what it is for one thing to depend synchronically on others. My own simple view of constitution, and indeed my deflationary view that synchronic dependence relations are identical, seeks to rectify just that. But as it stands, the terminological drift in the literature makes it difficult to see what technical notion is in play.

Moreover, while restricted and *Unrestricted Constitutionalism* are conflicting,⁵ it’s possible that theorists vacillate between them. For example, Clark

⁵ It might seem like the restricted and unrestricted notions of constitution could co-exist if theorists simply made their claims more precisely in terms of either. Wilson (2007), for example, argues that we can account for the truth of the claims, “This water is nothing more than the molecules that constitute it,” and, “This statue is something over and above the marble that constitutes it,” in terms of differing notions of constitution. Wilson calls the notion operant in the former claim “compositional” constitution, while “ampliative” constitution is the operant notion in the latter. Ultimately, I’ll argue that there’s a problem with the ampliative notion of constitution in the sense that it takes for granted the difference between constituents and the rest of the physical conditions on which the constituted thing depends, when that very boundary may be questioned, as it is in the present context. I’ll suggest in Chapter 10 that these notions of constitution are, in fact, in tension for that reason. For now, however, *Unrestricted Constitutionalism* can be understood as the claim

(2010c) offers enactivist arguments for externalism as well, and concludes by saying,

Confronted by the kaleidoscope of cases encountered in the previous chapters, the proper response is to see mind and intelligence themselves as mechanically realized by complex, shifting mixtures of energetic and dynamic coupling, internal and external forms of representation and computation, epistemically potent forms of bodily action, and the canny exploitation of a variety of extrabodily props, aids, and scaffolding. Minds like ours emerge from this colorful flux as surprisingly seamless wholes: adaptively potent mashups extruded from a dizzying motley of heterogeneous elements and processes.

As much as I share his overall perspective, what's missing from Clark's ambitious project is philosophical clarity about realizers, mixtures, couplings, exploitations, scaffolds, emergence, and wholes. The dizzying motley of terminology makes it difficult to pin down the concept or concepts of dependence in play so as to assess the conclusion that minds depend constitutively on extra-neural entities. It turns out that Clark and others' terminological liberties are justifiable, it's just that no one has yet provided that justification. In any case, the immediate point is that insofar as we are asked to countenance somewhat extraordinary or counterintuitive objects—extended states and processes—it is not obvious that externalists' notions of constitution are like our ordinary or intuitive notion of one thing's being part of or constituting another. So the battle line between unrestricted and restricted notions of constitution might cut across the externalist camp, but it's not clear whether there are any restricted externalists.

Might the internalist camp be divided by the same line? It would seem so, but there is a similar problem: it's not clear whether there are any unrestricted internalists. Because critics of externalism like Adams and Aizawa disagree about what makes a process or state cognitive or mental, it remains unclear

that “anything on which an entity depends synchronically is compositionally constitutive of that entity,” and restricted constitutionalism can be understood as the claim that “not everything on which an entity depends synchronically is compositionally constitutive of that entity.”

whether *they* think that the states or processes appealed to by externalists like Clark and Chalmers are partly constituted by extraneural entities, and just aren't cognitive, or whether they think the states and processes in question aren't constituted by anything extraneural. One might expect Adams and Aizawa to deny *Unrestricted Constitutionalism* and opt for restricted constitutionalism, first because it accords with common intuition, but moreover because they accuse EM theorists of committing a fallacy they call the "coupling-constitution" fallacy. Adams and Aizawa describe the fallacy as taking cases "in which some object or process is coupled in some fashion to some cognitive agent," and inferring, "that the object or process constitutes part of the agent's cognitive apparatus or cognitive processing" (2010, p. 68).

But, crucially, criteria for delineating restricted constituents from mere circumstances upon which a thing synchronically depends are orthogonal to the delineation of what is cognitive and non-cognitive. So Adams and Aizawa's marks of the cognitive—non-derived content and mechanistic homogeneity—cannot be read as committing them one way or another as to unrestricted or restricted constitutionalism. Let me put this another way. On the one hand, there is the question of what makes an entity cognitive. On the other hand, there is the question of what (if anything) would make some, but not all, of the physical conditions on which an entity depends synchronically *constituents* of that entity. And Adams and Aizawa do not provide criteria of the latter sort.

There is a further problem that suggests the "coupling-constitution" fallacy may be a misnomer. Note that in describing the fallacy as taking cases "in which some object or process is coupled in some fashion to some cognitive agent," and making the further invalid inference, Adams and Aizawa take the uncoupled agent to already be independently "cognitive." By contrast, EM theorists take a whole process to be cognitive in virtue of the causal role that process plays, and thus take Otto-and-his-notebook to be a cognitive system, not so much because Otto is a cognitive system, but because Otto-and-his-notebook function as one. So EM theorists are quite simply not guilty of the inference pattern Adams and Aizawa accuse them of. In fact, no one to my knowledge disputes Adams and Aizawa's point that something's being in causal interaction with another thing does not make the first a part of the second.

Yet, despite Clark and others' insistence that EM theorists make no inference of the sort they are accused of, Adams and Aizawa have remained unmoved. What explains this impasse? A clue, I think, is that Adams and Aizawa say they "have been impressed by the extent to which [inadequate attention to the mark of the cognitive] appears to have been involved in so many process externalists' succumbing to one or another version of the coupling-constitution fallacy" (2010, p. 78). They say,

So, if the fact that an object or process X is coupled to a cognitive agent does not entail that X is part of the cognitive agent's cognitive apparatus, what does? The nature of X , of course. One needs a theory of what makes a process a cognitive process rather than a noncognitive process. (2010, p. 68)

Now, notice here that X first stands for an *external* object or process—a mere part of a system or process that EM theorists claim is cognitive—while in the last instance X might be taken to stand for a whole process whose nature makes it cognitive or not. Clark understandably misses this equivocation when he replies that,

Adams and Aizawa seem to think that some objects or processes, in virtue of their own nature... are, as we shall now put it, *candidate parts* (for inclusion in a cognitive process), whereas other objects or processes, still in virtue of their own nature, are not. This, I think, must be the way to give sense to that otherwise baffling question "is some X cognitive?" (2010b, p. 85)

Again, while Adams and Aizawa's marks of the cognitive—as involving non-derived content and as mechanistically homogeneous—are problematic, even accepting them would do little to clarify what, if anything, about parts themselves would make them candidates as parts of cognitive systems (and why that should rule out extraneural entities). But Adams and Aizawa do not take their criteria to do this work. They claim,

the skull does not constitute a theoretically significant boundary for cognitive science. More specifically... being inside the brain cannot be the mark of the cognitive. This seems to us to be true and obvious. The bounds of cognition must be determined by finding

the mark of the cognitive, then seeing what sorts of processes in the world have the mark. Following this method, we see that, as a matter of contingent fact, the cognitive processes we find in the real world all happen to be brain bound. It appears to be just a contingent empirical fact that cognitive processes are not transcorporeal processes. (2001, p. 46)

They defend this view consistently.⁶ Since they are mostly clear⁷ on the specific point that the mark of the cognitive tells us what processes, as wholes, are cognitive, *X* in the second instance in the equivocal passage above must be understood as referring to a whole, of which the first *X* is merely a candidate part. Since their question in that passage is about a whole cognitive process

⁶ For a sample, Adams and Aizawa say,

We do not propose that non-biological structures are incapable of supporting non-derived content. Our view is that, as a matter of contingent empirical fact, cognitive processing typically occurs within brains, even though it is possible for it to extend. (2008, p. 47)

The point is not that the electronic processes in the computer could not be so organized as to constitute a cognitive process; it is merely that as a matter of contingent empirical fact they are not. (2008, p. 143)

⁷ Adams and Aizawa say,

Our view has always been that, as a matter of contingent empirical fact, pencils, paper, eyeglasses, Rolodexes, and so forth happen not to be parts of any currently existing cognitive economy. It is, in principle, possible that, say, a pencil or notebook could be so employed as to be a contributor to a cognitive process. This would happen if the pencil and notebook were, in whole or in part, to satisfy conditions on the possession of non-derived content and participate in the kinds of causal relations that we theorize are distinctive of cognitive processes. This is just to say that portions of the body and tools in the extracorporeal environment could be parts of cognitive processes, if only they were to bear the mark of the cognitive. (2008, pp. 128–129)

I say they are “mostly” clear that the mark of the cognitive applies to cognitive wholes rather than to parts of cognitive wholes because that’s the only way pencils might in principle be among such parts. Obviously pencils on their own will not satisfy their criteria and still be pencils. So there is something confused when Adams and Aizawa go on to say that in order to be parts of cognitive systems, pencils themselves would need to bear the mark of the cognitive. But I needn’t sort out their confusion. Beyond diagnosing the source of an impasse in the EM debate as it has so far played out, what I want to pursue in the main text is the question of what *does* remain problematic about the EM thesis.

rather than a part of one, in fact, Adams and Aizawa don't ask the "baffling question" Clark attributes to them about the candidacy of parts in cognitive systems. In fairness to Clark, Adams and Aizawa have not clearly articulated their overall view, and even made remarks that further confuse matters.⁸ Anyway, if the question of what makes something a cognitive process and what makes something part of a cognitive process are indeed separate in the sense that the criteria for being cognitive apply not to parts but to wholes, and the criteria for parthood might be synchronic dependence, plus, perhaps, attachment or the like, then how should we understand the connection between the mark of the cognitive and the alleged fallacy? I submit that the reason Adams and Aizawa spot a fallacy in connection with a supposed inattention to the mark of the cognitive is simply that Adams and Aizawa, for other reasons, do not accept that extended systems, as wholes, count as cognitive. They already and only count as cognitive what are in EM theorists' view subsystems, sub-states, or sub-processes of cognitive ones. So they understand EM theorists as making a whole "cognitive" in virtue of having (on their view) a properly

⁸ At the outset to Adams and Aizawa (2010), the authors quip,

Question: Why did the pencil think that $2 + 2 = 4$? Clark's answer: Because it was coupled to the mathematician. That about sums up what is wrong with Clark's extended mind hypothesis. Clark apparently thinks that the nature of the processes internal to a pencil, Rolodex, computer, cell phone, piece of string, or whatever, has nothing to do with whether that thing carries out cognitive processing. (p. 67)

The first caricature of Clark's view suggests a reading of "whether that thing carries out cognitive processing" as its doing so *all on its own*, in which case, it totally misrepresents the EM theorist's view, and leads Clark to respond,

the thrust of Adams and Aizawa's rhetoric is, again, and again, to draw attention to the evident absence of cognition *in the putative part* as a way of 'showing' that coupling... cannot play the kind of role it plays in the standard arguments for cognitive extension. (2010b, p. 83)

And Clark says, it would be "crazy to think that a pencil might think" (2010b, p. 83). On the other hand, if "whether that thing carries out cognitive processing" is understood as its being involved in cognitive processing as a part, Adams and Aizawa would be on target, but inconsistent with the reasonably clear aspect of their own view already noted. That is, while they've told us what makes a thing cognitive, they haven't told us what makes something part of something cognitive.

cognitive part: a brain, brain state, or brain process. But in that case—that is, if Adams and Aizawa are right about which processes are cognitive—EM theorists would be guilty of a fallacy of composition: attributing to the whole a property of one of its parts. What might have been a minor equivocation on *X*, combined with liberties taken in certain caricatures of Clark’s view (just noted in Footnote 8), together with the likely case that Adams and Aizawa may have misdescribed the inference that bothers them, has meant that defenders of EM have failed to appreciate what might be a genuine concern, as we saw in Section 2.1.

The upshot of my alternative diagnosis of this dispute as it relates to the *Unrestricted Constitutionalism* premise is that it is not entirely clear whether any parties to the debate over EM dispute the claim that anything on which an entity depends synchronically is constitutive of that entity. Rather, Adams and Aizawa appear to deny that anything *cognitive* depends at some time in its duration on extraneural entities because, by their lights, the processes which do aren’t wholly cognitive, and the part which is cognitive (the brain) does not so depend. Again, while Adams and Aizawa name the fallacy they detect a “coupling-constitution fallacy,” the fact that EM theorists plainly do not make an inference of the form Adams and Aizawa suggest, together with the fact that they detect the fallacy in connection with an “inadequate attention to the mark of the cognitive” (2010, p. 78), suggests that their misdiagnosis doesn’t amount to a separate objection to EM, but flows from their objection to the scientific seriousness of functional kinds that would include as members both Inga’s normal beliefs and Otto’s notebook-involving “beliefs.”

So it is clear that there is a disagreement about what makes something count as cognitive, as a special case of a more general disagreement about what makes any thing the kind of thing that it is. It is less clear whether there is any disagreement over constitution. Nonetheless, as I signalled above, insofar as externalists’ notion(s) of constitution (or similar) implies rather extraordinary entities, such as human–notebook couplings, it is not obvious that these are consistent with ordinary, restricted intuitions about constitution. As I said above, it is not clear that there are any restricted externalists, but there are probably unrestricted externalists. To put the same point another way, if

anything on which an entity depends synchronically is constitutive of that entity, then *Extended Cognition* follows, no matter what the extraneural entity is, as long as a mental state or cognitive process depends synchronically on it. By contrast, internalists plausibly need to deny *Unrestricted Constitutionalism*. At least, it looks much more difficult to both affirm it and be an internalist, as that would involve claiming that neural goings on are always completely sufficient for mental life.

I highlight this particular battle line, and yet I overlay it cautiously onto debates as they exist. That may seem like an odd position to be in. From one perspective on the question of *Unrestricted Constitutionalism*—the intuitive rejection of it—it might appear that I’m misidentifying the issues. But if I’m right that the intuition is faulty and that restricted constitutionalism is unmotivated and sloppy, then existing debates have failed to clearly identify the issues. So I’m willing to go out on a limb. I take comfort in the fact that the question—though it comes in different guises—of what makes one thing constitute another runs through a whole range of issues in philosophy. There is a significant debate to be had over *Unrestricted Constitutionalism*, and what has in fact happened is that externalists have largely ignored certain technicalities, outsourcing work on concepts of dependence as if to borrow clarity from another part of philosophy where there is in fact further debate to resolve—debate that crucially impacts the question of the physical extent of minds and our very selves.

Wrapping Up

The goal of this chapter was to understand what is controversial about externalism—the idea that minds are bigger than brains in the sense that cognitive processes and mental states can frequently be constituted in part by extraneural entities. I’ve argued that some key underlying assumptions in arguments for externalism require further elaboration and defense. As exemplified by the mainstream debate over EM, debates about externalism are continuous with larger issues in philosophy. This, I think, is how Clark has come to observe that,

The notion of the extended mind draws strong reactions. Many feel it is patently false. These same people tend to feel that the mind is simply and obviously just the activity of the brain. Others regard it as patently true, and they tend to be those who identify the mind with an essentially socially and environmentally embedded principle of informed agency.... My own feeling is that we have not yet reached the philosophical or scientific bottom of this debate. (2010a, p. 64)

Indeed, what one says about EM will be informed by what one says about minds, full stop. I began by suggesting that consciousness is a defining feature of minds as typically conceived, which, paired with certain intuitions about what worldly machinery could explain our minds, produces discomfort with the EM thesis, particularly as offered by Clark and Chalmers. The discomfort is as serious and lasting as the intuition that consciousness has a merely neural basis, an intuition we should ultimately reject. But the overall point here is that the sort of worry that keeps EM from being an obvious thesis is generated by disagreement about what could explain minds and thus how to understand mind–physical world relations. Divided intuitions over EM are symptoms of underlying disagreements over the relation or relations of objects of the special sciences to the objects of fundamental physics.

In Section 2.1, I showed that a major fault line in the debate over EM runs along proponents’ and critics’ differing notions of kindhood. Again, when Clark likens Otto’s notebook to dispositional beliefs and claims this likeness is “all that matters” (2010b, p. 89), he recapitulates functionalism rather than defends it. A wider variety of functionalist externalisms cite different extraneural entities and *could* be committed to different notions of constitution. But that does not make for a “what” versus “how” difference between externalisms. Rather, it makes for a possible disagreement about what belongs in a *how* explanation, that is, a possible disagreement over the *Unrestricted Constitutionalism* premise or what should replace it. In Section 2.3, I showed that a medley of synchronic dependence relations or expressions are used when considering how to “account for” (to use one such expression) our minds, and it is not clear that these terms are being used consistently or in well-motivated ways. I’m ultimately going to defend a notion of constitution as synchronic ontological dependence according

to which what makes things like notebooks part of certain cognitive processes also makes things like museums part of those processes, and according to which the truth of the latter doesn't change with a change in the terminology with which the synchronic ontological dependence relation is described. I argue that synchronic dependence relations are identical, that functional kinds can be projectible, and that wholes are nonetheless identical to their complete parts, constituents, or total realizers. I argue that externalism about minds follows.

I agree that we have not reached the bottom of the debate. Yet I suspect we already have what empirical evidence we need to decide the issue when it is framed in a coherent way. What we disagree on are large chunks of the conceptual schemes in which the evidence is understood. What we should say about the constituents of minds has everything to do with how to think about higher-level kinds in general. An account of higher-level kinds is the first step toward having a plausible theory of minds. After all, minds must fit into an overall model of what exists if they are to be kept as a kind in a scientific ontology. For my part, it is unsurprising that underlying issues with externalism intersect with a great many other areas of philosophy. What makes this a useful contribution to the philosophy of mind is that I explicitly tie the question of extended minds to these wider issues in philosophy. It is these issues that need to be addressed before a final assessment of externalism.

Chapter 3

Concepts of Dependence

The philosophy of mind frequently, if implicitly, relies on appeals to various dependence relations thought to hold, or thought to fail to hold, between mental states and other objects or events in the world. These primarily include claims about identity, supervenience, realization, constitution, and composition, all of which are usually held to be synchronic. These terms invoke dependencies that, for example, persons might have on their bodies, minds might have on brains, value might have on paper bills or works of art, and macro properties of objects might have on their atomic or chemical compositions. Synchronic dependence relations are usually held to be non-causal. Whereas causes precede their effects, the molecules of an apple don't cause it to be that apple at a later time, nor does my body-at-a-time *cause* me-at-that-time to exist, as if at some time later than it. So entities are contemporaneous, that is, synchronous, with their constituents, realizers, supervenience bases, or component parts.

But as we saw in Section 2.3, philosophers don't always use consistent language or provide formal definitions of their terms. To further complicate matters, formal definitions of dependence relations are each the subject of philosophical debates in their own right. And, their separate treatments in the literature explains the difficulty in untangling these relations. But, as Karen Bennett (2011) notes, their overlapping use in *application* belies a family resemblance. Somewhat surprisingly then, no systematic comparison between these relations seems to have been made until the recent work of Bennett (2011, 2017). Taking her as work as a starting point, I begin with a brief descriptive survey of a few of the synchronic dependence relations most invoked

in the extended mind literature and most relevant to the philosophy of mind in general. Section 3.2 then summarizes Bennett’s view of what she calls “building” relations, including those relations about to be discussed, and a few others. Suspecting that “no one else has even claimed that the various building relations form a conceptually unified family” (2011, p. 95), she formulates what she believes they have in common. I agree that building relations have something in common, but I characterize the resemblance differently in Section 3.3. Then, after seeing in Chapters 4 to 6 how issues surrounding dependence relations are tied up with questions about the nature of higher-level entities—and ultimately with formulating a plausible physicalism—my characterization will allow me in Chapters 8 and 9 to develop a more detailed taxonomy of dependence relations and defend the view that higher-level things are identical with everything they depend synchronically on, and that synchronic dependence relations are identical with one another. That allows us to infer things about our identity from our “constituents,” “realizers,” “parts,” “supervenience bases,” or “implementations”—these being one and the same.

3.1 Synchronic Dependence Relations Sketched

3.1.1 Supervenience

Supervenience is a relation of covariance roughly captured by the slogan “there cannot be an *A*-difference without a *B*-difference.” (McLaughlin and Bennett, 2014) Supervenience is (trivially) reflexive, meaning that there “cannot be an *A*-difference without an *A*-difference.” And supervenience is transitive; if *A* supervenes upon *B*, and *B* supervenes upon *C*, *A* supervenes upon *C* (McLaughlin and Bennett, 2014). At least since Kim’s (1984) “Concepts of Supervenience,” a feature of the concept of supervenience as covariance has been that it is non-symmetric, meaning that supervenience can sometimes be asymmetric, and sometimes symmetric. It is asymmetric when, for example, the mental supervenes upon the physical. That is, for every mental difference there is a physical difference, but not vice versa. Two worlds might differ in physical properties without differing in mental ones, say, if one of Saturn’s rings had one more molecule of ammonia than it does (Kim, 1987). On the

other hand, there appear to be cases of symmetrical supervenience: the volume of a sphere covaries with and supervenes on the surface area of a sphere, and vice versa (McLaughlin and Bennett, 2014). Notice that bare supervenience formulations, conforming to the slogan, “there cannot be an *A*-difference without a *B*-difference,” (McLaughlin and Bennett, 2014) do not specify the modal force of the dependence between *A* and *B*, whether logical, metaphysical, nomological, or physical. What Kim (1984) calls “weak” supervenience, of the form “there is no *A*-difference without a *B*-difference,” does not even specify that the dependence relation is a necessary one (as opposed to a matter of happenstance), and are therefore compatible with strong emergence or even dualism if they are set to work capturing mind-body relations. Finally, bare supervenience formulations do not capture the nature of the dependence between *A* and *B*, whether the relation is that between a cause and its effect, that between a thing and itself or its interrelated properties, or some other kind of dependence. However, as Bennett (2011) puts it, “most actual supervenience claims in philosophy were and are *intended* as building claims” (p. 85), which are, on Bennett’s view, asymmetric. Although I won’t discuss supervenience much herein, it is this use to which philosophers have put supervenience in formulating physicalism as a general claim about how things are physically built that puts it in the dependence relation ballpark. Note, too, that any dependence relation is also one of supervenience in its most general sense, that is, if *A* depends on *B*, *A* also supervenes on *B* by covarying with *B*.

3.1.2 Constitution

Constitution seems to be one of the less clear relations in analytical metaphysics. According to Ylikoski (2013), constitution only began to receive much careful attention in the decade previous. Note, first, that the literature diverges by using constitution in two senses divided by a seemingly important axis. That axis is what Bennett calls the “unification” axis:

as I have characterized them, some of the relations seem to involve a certain kind of wrapping up into one, and others do not. That is, some unify, or gather together a multiplicity, in a way that others do not. This is reflected in a difference in logical form: some of the relations are many–one, and others are not. (2017, p. 17)

Bennett herself treats constitution as “one–one,” meaning that its relata are both singular. For example, a piece of marble might constitute a statue (see Baker, 1997, 1999; Wasserman, 2004, 2017; J. Wilson, 2009; R. A. Wilson, 2009). Such literature, on the topic known as “material constitution,” uses constitution fairly consistently in this way, although even it may occasionally slip into talk of “constituent parts” in the plural (see Wasserman, 2017; R. A. Wilson, 2009). But there are other (admittedly overlapping) bodies of literature that use constitution explicitly in the plural sense. What is sometimes known as the “new mechanistic” approach in the philosophy of science and mind (see Craver, 2014; Baumgartner and Casini, 2016; Gillett, 2013; Kirchhoff, 2015; Ylikoski, 2013) consistently uses constituents in the plural. Finally, the extended mind literature is famous for claiming (or denying) that things beyond the brain can partially constitute cognitive processes (see Clark and Chalmers, 1998; Adams and Aizawa, 2001, 2008; Clark, 2010c; Kirchhoff, 2015; Menary, 2010). Partial constitution, of course, implies plural constituents, and thus a many–one relation between constituents and the constituted.

There is a consensus among all of the above accounts, however, that constitution is a synchronic relation. Constitution is also generally considered to be an asymmetric dependence of some “higher-level” entity on some other “lower-level” entity or set of entities. Levels, in this context, are usually understood as ontological “joints of nature,” for example, the atomic, the chemical, the biological, or mechanical, and so on “up,” perhaps, to the psychological, the economic and the sociological.

3.1.3 Realization

Realization has been employed to capture an asymmetric dependence that is non-causal, not unlike constitution, but usually held to obtain between properties rather than whole entities. The primary significance of the idea of realization is a feature of the realized individual—its *multiple realizability*. Realization has been employed by functionalists to emphasize the possibility that differing individuals can play the same role. When things are defined by what they do, multiple, indeed, very different, things could fit the same specification. The classic example is that of a mousetrap, which can come in

a variety of forms and materials while still being a mousetrap in virtue of its ability to trap mice. More interesting for theoretical mind–body relations, and the thought that made functionalism the dominant such theory, is the possibility that the same functional or computational state might be implemented by different hardware states (see Putnam, 1975a).

3.1.4 Composition

Composition is the usual relation of parts to wholes in a sense that has historically, if controversially, been inclusive of a variety of relata. A few examples of part–whole relations include: “The remote control is part of the stereo system”; “The left half is your part of the cake”; “The contents of this bag is only part of what I bought”; “The outermost points are part of the perimeter”; “The first act was the best part of the play”; “The goalie is part of the team” (Varzi, 2016, sec. 1). In its most general sense, composition is considered to be reflexive, meaning that every individual is part of itself. Composition is antisymmetric, meaning that if x is part of y and y is part of x , then x is y . Otherwise, if x and y are distinct, and x is part of y , y cannot be part of x . And composition is transitive; if x is part of y and y is part of z , x is part of z .

On the other hand, what is known as *proper* composition, and, correspondingly, proper parthood, can be defined as the relation between parts and wholes such that nothing is part of itself. Proper composition, then, is irreflexive, asymmetric, and still transitive. This will be an important difference. Bennett (2017) takes proper composition, but not reflexive composition, to count as a building relation in her sense.

Even the somewhat more restricted notion of proper composition as just sketched is very broad, and there are of course many questions and issues that arise, for example, about whether the same relation is really had by acts to plays, components to mechanisms, points to lines, fractions to wholes, and roles to organizations. With such an array of relata, some of dubious ontological status, one might wonder whether to expect composition to always be an ontological dependence relation. But I need not settle that. All that matters is that any real individual ontologically depends on its parts.

With such an array of relata, one might also wonder if the composition relation is always synchronic. Yet, even the parts of the play together span the same duration as the play. So composition is non-causal. The parts of the play do not together cause the play at a later time, as causes precede effects, rather, the duration of an individual contemporaneous with the span of its temporal parts. Nor does the first act of a play cause the play, though it is among the causes of later acts. The first act cannot cause the play because they begin at the same moment.

3.1.5 Identity

Identity, or sameness, is usually thought of in two ways. First, there is what is known as numerical identity; an individual is numerically identical only to itself. Secondly, things can be the same in the sense of having one or more shared properties. This is known as qualitative identity, and it comes in degrees.

I mention identity for two reasons. The second reason will become clear in Section 4.2. But first, identity seems related to putative synchronic dependence relations because it might seem like one such relation. Imagine that a batch of pure water “just is” some number of H_2O molecules. It doesn’t matter for present purposes that this is actually contentious. Suppose that the water and the set of molecules have all the same properties. We naturally think that the water depends on the H_2O . Like other dependencies, this is supported counterfactually: delete the H_2O , and you have deleted the water. Delete some of the H_2O , and you have changed *that* particular batch of water into another. The two batches are not identical, for example, because they do not have the same mass. But the opposite is also true: delete the water, and you have deleted the H_2O . That much, we should expect. If A is B , then B is A , and every relation in which A stands to B , B also stands to A . But that is just to say that everything stands in relation to itself just as it stands in relation to itself, which is to say very little; identity is trivially reflexive. On the one hand, to say an individual’s existence or character depends on its existence or character is much less informative than to say what else it might depend on, for example, its ancestors, or its parts. On the other hand, the dependence of individuals on their parts raises the question of whether complete composition

suffices for identity. Suppose that it does. Indeed, one might define complete composition as the minimally complete set of physical conditions sufficient to necessitate the individual. Individuals, after all, often have parts, and for at least some individuals, like a particular mass of water, it seems plausible that once you have all the parts, there you have the individual. Then, to say that an individual's existence and character depends on the existence and character of its parts might well be explanatory: the relation between water and H_2O is indeed an empirical discovery that explains water's reactivity with lithium and that can be used to make predictions about the solubility of a hypothetical molecule in water. Still, if it is plausible to think, as we shall see Bennett does, that being asymmetric is essential to building relations, then the identity relation, being symmetric, cannot be a building relation.

3.2 Bennett's Resemblance Class of Building Relations

I am now in a position to summarize Bennett's (2017)¹ view as to how certain relations, including some of those I've sketched, form a resemblance class, before motivating, in this chapter, a similar approach that culminates in a rather different view, that synchronic dependence relations are identical, which I defend in Chapter 9.

Bennett characterizes what she calls "building" relations, and I will follow her in using that term for the class of relations with the features she specifies. First of all, since resemblance classes are "cheap,"—meaning that they are not extra ontological commitments—Bennett (2017) thinks a class should be theoretically useful and "best answer to," in this case, "our ordinary understanding of what it is for one phenomenon to account for another" (p. 31-32). Bennett argues,

three features are individually necessary and jointly sufficient for a relation to count as a building relation. All building relations are

- i. directed, in that they are antisymmetric and irreflexive,
- ii. necessitating, roughly in that builders necessitate what they build, and

¹ Every reference to Bennett in the remainder of this chapter is to Bennett (2017).

- iii. generative, in that the builders generate or produce what they build. Built entities exist or obtain because that which builds them does. (2017, p. 32)

Bennett’s overall argument, then, is that each of these features are part of our ordinary understanding of what it is for one phenomenon to account for another, and that the class of relations marked by these features is theoretically useful.

3.2.1 Building is Directed

By “directed,” Bennett means asymmetric,² so that “that nothing builds itself, and no two things mutually build each other” (2017, p. 33). Note that identity is excluded from the family here. Bennett goes on to deflect a couple of arguments for the possibility of symmetric and reflexive “building”, and provides positive reasons to accept asymmetry by appealing to “our pretheoretic concept of building” (p. 39), paradigmatic philosophical uses of dependence and necessitation locutions, as well as by appealing to theoretical utility. Bennett argues that it is theoretically useful to use building concepts to explain relative fundamentality (or ontological priority), according to which, “all it is for one thing to be more fundamental than another is for certain patterns of building to obtain between them”(p. 40). And the only way building can do this further work entails that it be asymmetric, since the relation *more fundamental than* is asymmetric.

3.2.2 Building is Necessitating

Bennett thinks that building relations all involve necessitation. Bennett’s position is that building cannot be indeterministic, and goes to some length to defend what she calls “building determinism,” the view that, “If something is built, something [or set of things] necessitates it” (p. 51). This is “a claim about full building, not partial,” because “no one thinks that mere partial

² Asymmetric means antisymmetric and irreflexive. Bennett sometimes opts for pairing the latter two terms to make the irreflexivity explicit, though it is part of the notion of asymmetry.

builders necessitate what they build” (p. 49). However, Bennett thinks it is an interesting question whether what she calls “full” builders necessitate what they build. So Bennett says, “I do not claim that each building relation B is such that if x fully B ’s y , it necessitates y ” (p. 49). Bennett thinks one could plausibly hold that “only something ‘larger’ than x , some $x + C$, guarantees y , even though x alone counts as the full building base [of y]” (p. 49). C could be some further set of facts, enabling conditions, or circumstances. As we’ll see in more detail in Chapter 4 and Chapter 7, formal characterizations of “constitution” and “realization,” as one–one relations distinct from identity, explicitly include something like C . So there are two ways to be a building determinist. One is to be a building “necessitarian,” holding that,

[**Necessitarianism**] “For all x and y and all building relations B , if x fully B ’s y , then, necessarily, $x \rightarrow y$.” (p. 52)

The other way to be a building determinist is to be a building “circumstantialist,” and hold that,

[**Circumstantialism**] “For all x and y and all building relations B , if x fully B ’s y , then, $\Box[(x + C) \rightarrow y]$.” (p. 52)

Here, C is some “to-be-specified set of background circumstances” (p. 52), with the further condition that C not fully include y or anything that fully builds y , blocking “cheap” cases wherein anything necessitates y in C , because C itself is fully sufficient for y . Bennett offers a few reasons to prefer the circumstantialist picture, but thinks nothing much of importance turns on which version of building determinism one accepts, because building involves necessitation on both accounts. Notice, however, that the *Unrestricted Constitutionalism* on which my main argument is premised is a species of *Necessitarianism*. And, as we’ll see in the next several chapters, the coherence of physicalism about higher-level entities and properties—and with it, the EM debate—turns on seeing that *Circumstantialism* involves unmotivated and misleading intuitions about the boundaries of entities.

3.2.3 Building is Generative

Finally, Bennett argues that all building relations are “generative,” that is, they license a certain kind of talk. She offers the principle that,

For all building relations B , and all x and y , x ’s B -ing y makes true certain explanatory and generative claims. For example, if a builds b , then b exists (obtains, is instantiated. . .) because a does, b exists (obtains, is instantiated) in virtue of a , a makes b exist (obtain, be instantiated), and so forth. (p. 58)

Bennett says, “I take all these phrases—‘in virtue of’, ‘because’, ‘make’, ‘make it the case’, ‘explain’—to be getting at the same idea. The idea in all cases is that builders *generate* the built” (p. 58). She thinks the generative clause is needed because a relation can be asymmetric and necessitating without being a building relation. I’ll have more to say about that idea in Section 3.3.2.

3.3 Rebuilding

With a picture of Bennett’s view in hand, I want to follow-up on some issues I flagged while summarizing it. First, I want to discuss transitivity. Then I’ll discuss generativity. At the same time, I’ll also flag issues with necessitation and directedness to be picked up in Section 9.2. Then I’ll present my view of how building relations are unified.

3.3.1 Transitivity

Interestingly, Bennett doesn’t think it is crucial that building relations be transitive. Transitivity in the case of building would mean that if a builds b , and if b builds c , then a builds c . Bennett thinks transitivity is obviously false of the relation between a and the set $\{a\}$, given that the same relation holds between $\{a\}$ and $\{\{a\}\}$, because, she claims, a is not a member of $\{\{a\}\}$. Bennett also notes that transitivity is mildly controversial in the case of grounding,³ and in the case of composition. However, *is a member of*

³ Bennett (2017) cites Schaffer (2012) as denying the transitivity of grounding in contrast to most others who think grounding is transitive. But although Schaffer offers counterexamples

and equivalent uses of *part of* might be intransitive because of additional criteria for membership *beyond* strict dependence, such as the condition that members be of a particular kind. So, for instance, in the case of composition, Varzi (2016) suggests that my left pinky is not “part” of the team of which I am a “part.” What is crucial to notice is that there’s more than ontological dependence packed into such a use of “part.” Likewise, although a is not a member of $\{\{a\}\}$, in the usual sense involving membership criteria other than or in addition to dependence, if a set-forms $\{a\}$, where “set-forms” means asymmetrically necessitates and generates, whether partly or completely, then a also set-forms $\{\{a\}\}$, partly or completely. If it sounds mistaken to say that the team depends on my left pinky, that must be either because the team does not depend on my membership, or because my membership does not depend on my left pinky, or both.

Further, one would think the connection Bennett posits between building and relative fundamentality, together with the transitivity of *more fundamental than* would entail transitivity as much as the same connection and the asymmetry of relative fundamentality entail the asymmetry of building relations. However, Bennett claims that the intransitivity of building is compatible with the transitivity of relative fundamentality. She explains,

Suppose that there is some nontransitive building relation B such that aBb , bBc , but $\sim aBc$. [A deflationary account of relative fundamentality in terms of building] entails that a is more fundamental than b , and b is more fundamental than c . The transitivity of *more fundamental than* entails that a is also more fundamental than c . But that does not conflict with the assumption that a does not build c : it simply has to be the case that one thing can be more fundamental than another despite not standing in a building relation to it. And that is clearly possible. Sodium ions are more fundamental than benzene rings, but benzene rings are not even partially built out of sodium ions—sodium is simply not involved. A similar point holds for tokens of types that in general are building related. Hydrogen atoms and oxygen atoms are more fundamental than water molecules, but that does not entail that a particular

to the transitivity of grounding, I take him as ultimately *defending* it, once grounding is reflected in contrastives (or, I think, adequately fragilized counterfactuals).

water molecule in Ithaca is built out of hydrogen and oxygen atoms located in Phoenix. (p. 47)

Actually, the point about sodium and benzene is the same as the point about tokens of types that are in general building related. If sodium ions *are* more fundamental than benzene rings, it is because atomic kinds of things build molecular kinds of things. But it's hard to see how any of this works as an example of intransitivity of building and transitivity of fundamentality. Suppose your pinky finger is not a member of the All Blacks. If your pinky finger is relatively more fundamental than Mr. Rogers, on Bennett's account, that must be because pinky finger kinds of things (tissues) often build the kind of thing Mr. Rogers is (an organism). And if Mr. Rogers is relatively more fundamental than the All Blacks, on Bennett's account, that is because Mr. Rogers's kind of thing (people) often build the kind of thing the All Blacks is (a sports team). If relative fundamentality reduces without remainder to building, it follows that the transitivity of relative fundamentality reduces to the transitivity of building. Your pinky finger's being relatively more fundamental than the All Blacks is just a case of the general fact that tissues partly build teams of people.

So I take dependence to be transitive: if x depends on y , and if y depends on z , then x depends on z .

3.3.2 Generativity Reconsidered

Recall from the end of Section 3.2 that Bennett adds the generativity clause to her characterization of building because non-building relations can be asymmetric and necessitating. Something that I have until now left aside is that Bennett considers "set formation," and "grounding" as candidate building relations. Why haven't I discussed them? Well, I will discuss them here, and in Chapter 9, to the extent that Bennett appeals to them as examples. But since there is a huge literature on grounding that does not directly overlap with the EM debate, grounding is mostly beyond my scope herein, and I must leave it to others to decide whether the considerations in this part have any bearing on the concept of grounding; I suspect they will, given a broad notion of grounding as synchronic ontological dependence between any sorts of things. On the

other hand, if grounding is taken strictly as a relation among facts, it might be plausible to deny that grounding resembles synchronic dependence relations used in scientific explanations, which posit dependence relations between things other than facts, like neurons and minds. Similarly with set formation, I'll suggest that on a certain definition of what it takes to form a set, set formation is not a relation of interest herein. And that might be enough to avoid the example that compels Bennett to add the generativity clause. So, let's hear the example. First, Bennett says, "Socrates and {Socrates} mutually necessitate each other, even though the building only goes one way; the necessitation is symmetric but the building is not" (p. 56). She continues,

There is an antisymmetric and irreflexive necessitation relation that holds between Socrates and {Socrates} that intuitively is a building relation—namely, set formation. But there is also an antisymmetric and irreflexive necessitation relation that holds between Socrates and {Socrates} that intuitively is *not* a building relation—namely, set *membership*. The account of building I have presented thus far cannot distinguish between these. It does not settle which of set membership and set formation is a building relation; it does not settle, as it were, the direction of building. So the account thus far must be supplemented. . . . (pp. 56-57)

I take the idea to be that Socrates' asymmetric and necessitating membership in {Socrates} is the "inverse" relation (as Bennett says just prior to the above example) of Socrates' asymmetric and necessitating set-forming of {Socrates}. Socrates is a member of {Socrates}, but not vice versa. Or, if you prefer, {Socrates} has Socrates as a member, but not vice versa. Then, I take the idea above to be that, $\Box(\{Socrates\} \rightarrow Socrates)$. That is, necessarily, if {Socrates} exists, then Socrates exists. And yet we wouldn't say Socrates exists *in virtue of* {Socrates}. Just the opposite. But notice that the claim being denied talks about existence. So, on the other hand, perhaps we *would* say that Socrates "is a member" of {Socrates} *in virtue of* {Socrates}. So I'm just not sure the example shows that being asymmetric and necessitating are not adequate to make a relation a building relation.

However, suppose it does show as much. Then something more might be needed to pick out building relations. But generativity, as Bennett formulates

it, can't distinguish between the cases in a principled way. Recall that both Bennett and I set out to analyze certain concepts and ask whether they have anything in common. Bennett introduces her project by saying,

we talk about building all the time, and have been doing so as long as we have been doing philosophy at all. We do it whenever we claim that some entity or phenomenon can be accounted for in other terms, and also whenever we deny that it can be. Either way, we are making claims about what builds—or fails to build—what.
(p. 2)

If there's a further characteristic uniting the class of building relations, I don't see how it can be the characteristic of licensing building talk, unless the point is that there's just one building relation and all these turns of phrases mean the same thing. But Bennett, as we'll see in Chapter 9, denies as much. So it is opaque to me how analyzing building relations can both illuminate our generative talk while being illuminated by it. Moreover, I think there's a way to accommodate the intuition that $\{\text{Socrates}\}$ doesn't build Socrates without adding a further criterion for building relations. Recall from Section 3.2.2 that, "no one thinks that mere partial builders necessitate what they build" (p. 49). I agree; only full builders necessitate what they build. That sets up the following dilemma. If, on the one hand, set formation is understood such that in a world with just one entity, A , there exists $\{A\}$, then, *ex hypothesi*, $\{A\}$ just is A . It cannot be anything else. The building relation can't bring about anything that the builders do not; Bennett herself argues (ch. 7) that building relations are themselves built by the relatum in the first place of the relation. In that case, it would be wrong to say that A exists because $\{A\}$ does to whatever extent it's wrong to say that Socrates exists because Socrates does. If, on the other hand, we want to say, *non-trivially*, that $\{A\}$ exists because A exists, we are employing an ontologically-laden notion of set formation itself. That is, if $\{A\}$ is not just A , it must be something more, and A alone is insufficient for $\{A\}$. In that case, in addition to depending on Socrates, the existence of $\{\text{Socrates}\}$ also depends the empty set—some collector or cognizer or curly braces—so that A and the empty set together build $\{A\}$. So, Socrates does not, all on his own, necessitate $\{\text{Socrates}\}$, and the example fails.

My view, then, is that we don't need a seemingly circular appeal to licensing of "in virtue of," etc, to analyze uses of those very turns of phrase, since, I claim, and will proceed to defend in Section 3.3.3, what all such locutions have in common is simply that they express dependences that can be rendered as true counterfactuals. {Socrates} doesn't build Socrates since it is *not* necessarily true that if {Socrates} did not exist, Socrates would not exist. Rather, it is true that if Socrates did not exist, neither would {Socrates}. The dependence here is asymmetric.

In fact, what I think the example shows is that partial building relations are asymmetric, but not necessitating, and full building relations are necessitating, but not asymmetric. I return to this point in Section 9.2.

3.3.3 Building and Dependence

The claim I just made, that all these relations express dependences that can be rendered counterfactually, is contentious. So a discussion about the connection between building and dependence is in order. Recall that Bennett argues that building involves necessitation. She also thinks that necessitation is *usually* reflected in what she calls, following Lewis (2000), "patterns of counterfactual dependence." However, just because some *a* builds some *b*, Bennett thinks it isn't *always* true that if *a* had not occurred or been instantiated, *b* would not have occurred or been instantiated. She cites the possibility of preemption and overdetermination as cases that don't support true counterfactuals. Causal preemption, for example, occurs when Suzy throws a stone at a glass bottle, breaking it, and Billy throws an equally well-aimed stone at the same glass bottle a split second later. Then the claim that the breaking of the bottle depended on Suzy's throw looks false, because the counterfactual claim that had she not thrown the stone, the bottle would not have broken, is false—Billy's would have done the job (see Lewis, 2000).⁴ Causal overdetermination is supposed to occur in cases where some effect is caused by many events independently sufficient for that effect, such as a death caused by a firing squad. Analogously, synchronic overdetermination might occur, for example,

⁴ Every reference to Lewis in this chapter is to Lewis (2000).

when some general fact, like, “something exists,” is all too well grounded; it is grounded by the fact that Bennett exists, and that her sunglasses do. And Bennett thinks the problem of the many⁵ shows there is “an awful lot of compositional overdetermination.” So Bennett says,

assuming the falsity of mereological essentialism, many composites can exist in the absence of a few of their parts. And many nonfundamental properties—even fairly determinate ones—are such that their instantiation can tolerate small differences at the microlevel. I could probably be in exactly the same qualitative state of pain even if one or two fewer neurons were firing. (ch. 3, note 18)

Mereological essentialism is the view that the numerical identity of wholes is fragile, so that for any individual, *that* individual does not survive any change of its parts (see Chisholm, 1973). In calling wholes “fragile,” I am borrowing Lewis’s turn of phrase, which we’re about to see, with respect to events and putting it to a corresponding use with respect to parts. I do this to highlight a general principle analogous to mereological essentialism; let’s call it,

[Building Essentialism] The numeric identity of all entities is fragile, such that for any individual, *that* very same individual does not survive any change to its building base.

Now, one way to resist the examples of preemption and overdetermination, while still assuming the falsity of *Building Essentialism*, comes through in a footnote of Bennett’s. She makes the point that many counterfactual claims in cases of building *do* come out true. One reason to think they do Bennett adapts from Lewis, who says,

When asked to suppose counterfactually that C does not occur, we do not really look for the very closest possible world where C’s conditions of occurrence are not quite satisfied. Rather, we imagine that C is completely and cleanly excised from history, leaving behind no fragment or approximation of itself (p. 190).

⁵ The problem of the many consists in the incompatibility of my intuition that there is just one desk at which I now sit, with the thought that no particular subset of its parts have any greater claim to being a desk than do many other similar subsets.

Bennett calls this “a ban on replacement evaluations of the relevant sort of counterfactual. They are inappropriate in building contexts as well as in causal ones.” And, she says, this ban “blocks a variety of putative counterexamples to the claim that built entities counterfactually depend on that which builds them” (ch. 3, note 18). It’s true that such a ban narrows the range of cases where dependence is not reflected in true counterfactuals. Taking compositional overdetermination as an example, a ban on replacement readings works to maintain the truth of the counterfactual. So if Bennett’s pain, *M*, depends on some neural state, *P*, then the counterfactual, if *P* had not occurred, *M* would not have occurred, is true as long as we read the counterfactual as involving *P* cleanly excised from history. The idea is that we ought not to imagine something similar to *P* in its stead.

So the ban would repair a counterfactual rendering of compositional dependence even in cases of compositional overdetermination. But what about preemption and firing squads? The story gets more complicated here, because as I read Lewis, though he toys with the idea, he actually imposes no ban on replacement evaluations of counterfactuals, while resisting the apparent counterexamples to the counterfactual analysis of causation. His solution is more subtle. Lewis’s strategy is to make the generic counterfactual, “if *C* had not occurred, *E* would not have occurred,” more *fragile* in one way or another. The first way would be to espouse something like *Building Essentialism*. In that case:

There is an obvious solution to cases of late cutting [or late preemption]. Without Suzy’s preempting rock, the bottle would still have shattered, thanks to Billy’s preempted rock; but this would have been a different shattering. It would, for instance, have happened a little later. The effect that actually occurred depended on Suzy’s throw, but not on Billy’s. (p. 185)

But Lewis is ambivalent here, noting that, “supposing that the event itself is fragile... would fly in the face of much of our ordinary talk” (p. 186). A putative alternative that Lewis modifies from Paul (1998) is to, “take a tailor-made fragile proposition about that event and its time,” so that, “The negation of that fragile proposition is the consequent of our causal counterfactual” (p. 185). For example, we could change the counterfactual to read, “without *C*, *E* would

not have occurred at all, or would have occurred at a time different from the time that it actually did occur, or would have occurred in a manner different from the manner in which it actually did occur” (p. 187). That gets the right result, that Suzy’s rock was the cause of the shattering. But Lewis thinks we can still do better, since we are ambivalent about the distinction between *E*’s not occurring at all and its occurring differently. Instead, whether or not we take events to be fragile, their “alterations” are, by definition:

Let an alteration of event *E* be either a very fragile version of *E* or else a very fragile alternative event that is similar to *E*, but numerically different from *E*. If you think *E* is itself very fragile, you will think that all its alterations (except for the actual alteration) are alternatives. If you think *E* is not at all fragile, you will think that all its alterations are versions. You might think that some are alternatives and others are versions. Or you might refuse to have any opinion one way or the other, and that is the policy I favor.
(p. 188)

Then, we refer to *alterations* in our counterfactual, and get the right result. “Suzy’s throw caused the shattering of the bottle and Billy’s preempted throw did not because, without Suzy’s throw, the alteration of the shattering which actually did occur would not have occurred, and a different alteration would have occurred instead” (p. 188).

Now, all this time, we’ve been imagining that *C* does not occur at all. But what if some alteration of *C* occurs? The apparent problem is that, on a fragile view of *C* itself (about which Lewis is trying to be neutral), some minor change to *C* counts as *C*’s not occurring, and in that case the counterfactual might not have the expected truth value. Lewis worries that, “if *almost* – *C* occurred instead of *C*, very likely the effects of almost-*C* would be almost the same as the effects of *C*” (p. 190). But the eventuation of almost the same effect doesn’t make a difference to the truth of a counterfactual that refers to the effect’s actual fragile “alteration.” So the problem—a counterfactual that expresses intuitive dependence coming out unexpectedly false—only arises if some *E* was multiply causable, such that almost-*C* had the *very same* effect as *C*. Note, but otherwise leave aside for the moment, that *this* is where it is tempting to think a ban on replacement of *C* with almost-*C* could do some

work. But I think Lewis can be read as saying, *usually* or *implicitly*, “When asked to suppose counterfactually that *C* does not occur... we imagine that *C* is completely and cleanly excised from history, leaving behind no fragment or approximation of itself” (p. 190). I think *usually* or *implicitly* is the right interpretation, because earlier, although Lewis says, “One repair would be to rewrite our counterfactual analysis, or add a gloss on its interpretation, in order to make this explicit,” he goes immediately on to say, crucially,

But there is another remedy. We could look at a range of alterations of *C*, not just one. As on the side of effects, we need not say which, if any, of these are versions of *C* and which, if any, are alternatives to *C*. These alterations may include some in which *C* is completely excised, but *we need not require this*. They may include some that are almost but not quite *C*, but we say nothing that restricts us to the closest possible alterations. Then we look at the pattern of counterfactual dependence of alterations of the effect upon alterations of the cause. (p. 188, my emphasis)

The upshot is that Lewis’s repair isn’t a ban on replacements; it encourages them. The causes of an event are those that make differences to it, and that influence *does* make true counterfactuals of the right, more precise, sort. Now, I think that any fragilization strategy actually amounts to reading dependence counterfactuals according to something like the principle of *Building Essentialism* about events, one way or another. Lewis recommends expressing dependence in counterfactuals that, unlike our ordinary ways of speaking about events, refer to fragile alterations about which *Building Essentialism* is true. That looks suspiciously like insisting that events really are fragile, despite our ordinary ways of speaking. That is why the “mixed” fragilization strategy works in the case of almost-*C*. If almost-*C* counts as *C*’s not occurring, that is a fragilization of *C* which yields the right result as long as *E* is also taken to be fragile, and provided that *E* itself, the fragile thing, is not multiply causable by *C* and almost-*C*. If an individual fragile event could be caused by distinct fragile individual causes, then referring to alterations in the antecedent of the counterfactual won’t make the required repair anyway. I’ll say more on this point shortly. Here, the point is that there is no deep difference between permitting ourselves to suppose that almost-*C* counts as *C*’s “not occurring” or

counts as a “version” of C . The point is simply that we ought not to read some cause—for example, a particular neurological state—as fragile while reading an effect—for example, an evasive action—in an ordinary or liberal way. If we make that mistake, we’ll get what are obviously causal relations failing to be reflected in true counterfactuals tested by this double-standard. “If C had not occurred, E would not have occurred,” looks false in case, for example, an alternative neurological state counts as a non-occurrence of the event whose causative role in a behaviour is being tested, while an ever-so-slightly different evasive action is taken *not* to count as the behaviour’s non-occurrence, that is, it is counted as occurring. We must be even-handed in the fragilization.

We can adapt this strategy to cases of overdetermination in non-causal building generally or synchronic dependence. That is, if a builds b , then, as long as a and b are taken to be the actual alterations or fragile individuals, the following counterfactual is true: if a had not been instantiated, then b would not have been instantiated. Again, this fragilization strategy endorses something like *Building Essentialism* about either fragile “alterations” or fragile individuals. Supposing individuals are not fragile, some of our ordinary counterfactuals might come out false. So we need counterfactuals that refer to fragile “alterations.” But that, I have said, is much like supposing individuals really are fragile, unlike our ordinary ways of speaking. And that is just to say that what we’re really doing when we think we’re referring to an individual in our ordinary way of speaking is referring to a kind (I return to this point briefly below, and in Section 11.1). So, contra Bennett, I’m not inclined to *assume* the falsity of mereological essentialism or *Building Essentialism*. I’m similarly not inclined to trust intuitions about the “exact sameness” of two pains involving some neurons on one occasion, and involving a few less on another occasion. Either the first pain did not depend on exactly *those* neurons (among other things), or the second pain is not strictly the same (though they are both kinds of pain, probably the same kind of kind of pain, and so on).

Now, there’s still a problem with the fragilization strategy in the case of causation, as I noted above, and it is a problem for testing synchronic dependence, too. I said that the fragilization strategy works except if a particular effect can be caused in multiple ways. And, as I’ll argue in Section 6.1,

I think there *are* what I call “switches” in nature such that certain effects really are multiply causable. Suppose “if exactly C had not occurred, exactly E would not have occurred,” may be false even though, for example, a particular neurological state is intuitively a cause of a particular evasive action. It seems like a minor neurological difference might not make a difference to the action, though it makes minor differences elsewhere in the causal network—for example, it can be detected by a sufficiently accurate brain scanner. So there are cases where almost- C occurred instead of exactly C , and an obviously causal relation fails to be reflected in a certain counterfactual, even though we are even-handed in our fragilization and reading both a cause and an effect alike as fragile events or alterations of events. To repair *that*, we would have to, contra Lewis’ attempt at neutrality, insist that almost- C *cannot* count as not- C . That is, we would indeed have to go back to imagining C cleanly and completely excised from history to get the right result.

Correspondingly, one might press an intuition like Bennett’s about having the same pain despite a neural difference, thinking, surely, there can be multiple realizability just as there can be multiple potential causes of an effect. Indeed, I’m going to argue in more detail in Section 6.1 that the former is conceivable because of the latter. The familiar idea of multiple realizability is that certain alterations to, for example, a neural state, might not change the mental state that it realizes. Perhaps a “version” of pain involving some rearranged neurons with the same functional profile can be said to be “the same” pain. Of course, again, the fact that the rearrangement counts as an alteration entails that it is a difference that makes a difference *somewhere* in the world. Still, we almost never try to refer to *all* of a cause’s effects. When we variously say things of the form, “if C had not occurred, E would not have occurred,” E doesn’t stand for the entirety of events in the forward lightcone of C . If it did, then any change to C could well count as C ’s not occurring, since exactly the same world would not result, and the counterfactual would come out true, no ban on replacement evaluations required. But such dependence claims as, “if it had not rained yesterday, something around here would have been different,” are not particularly useful. In the case of synchronic dependence, however, we can go more usefully in both of two ways. We might test the dependence of *just*

such an entity—as individuated by all of its properties—on one of its parts by removing that part. A change in any property affirms the dependence. The part (say, a hydrogen atom) was indeed a part of just that entity (a fat molecule) with just those properties (being saturated, having such and such mass, etc.). But we also characterize things according to just some of their properties. Similar neural states can have the same effect (play the same role) within a *particular* downstream branch of a causal network, despite necessarily having differing total effects in the entirety of their forward lightcones. Suppose, quite plausibly, that to test the dependence of a pain on a neural state, we can modify or entirely prevent the neural state, but we can only monitor some branches of the downstream causal network (behaviour, etc) for effects characteristic of the pain. Suppose now that we modify the neural state a little, but detect no change in the pain or behaviour. We say “the” pain occurred anyway. But in that case, the consequent refers not to a fragile alteration, but to a non-fragile kind. A counterfactual where a replacement is allowed to count as satisfying the antecedent can be false in virtue of the multiple realizability of that kind. So when testing the dependence of a putative “entity” characterized by just one or a few properties—individuated by just one or a few effects—we must once again impose a ban on replacement evaluations of the causal counterfactual. An objection to the counterfactual analysis on the basis of multiple realizability fails because it still involves a double-standard, though less obviously. The objection is that it can be false that, “if exactly P (a neural event) had not occurred, M would not have occurred.” For example, the “same” pain, it might be claimed, could have been realized by slightly different neural states or events. But this objection counts P —the set of parts or physical conditions being tested for dependence—as a fragile alteration. Replaced by a similar set, P is taken not to occur. However, to count M as occurring anyway is to take a mere subset of the full effects of P —those functions that characterize pain—as *characteristic of P itself*, which is to take P as a *non-fragile* thing (a kind, in fact). Rather, we can only legitimately test the dependence of a non-fragile pain (M) on a non-fragile neural event (P), which is to say, we ban replacement evaluations of counterfactuals in which non-fragile “entities”—characterized by few properties and effects—figure in the consequent, so that the antecedent

isn't satisfied on the cheap.

The point can be summarized this way. *Building Essentialism* is actually compatible with the multiple realizability of properties and kinds. The way to read a counterfactual expressing synchronic dependence is determined by what is found in the consequent of the counterfactual, whether an individual or a kind. To test the dependence of M on P where M is a kind (e.g. pain generally), the antecedent of the counterfactual must be considered satisfied only if “not P ” involves no fragment or approximation of P (e.g. a neural state). The upshot is that building and dependence go hand-in-hand and the counterfactual analysis of dependence remains plausible. What it is to be a builder, a cause, or a part is to make a difference to the built, the effect, or the whole. Difference-making is reflected in true counterfactuals of the right sort, and dependence is thus testable, at least in principle, if we could design the right experiments.

Wrapping Up

In this chapter, I've sketched a variety of dependence concepts used especially in physicalist, or would-be physicalist explanatory projects, including the debate over what constitutes our minds, realizes our mental states, or what our cognitive systems have as parts. I've summarized a set of features—directedness (asymmetry), necessitation, and generativity—that Bennett thinks are common to a group of relations that overlaps with the ones I'm most interested in. I've questioned whether generativity does any work that the other features don't, and I've flagged concerns about asymmetry and necessitation to be picked up in Section 9.2. I agree with Bennett that building involves necessitation and that building is deterministic. But since only full builders (or perhaps full builders in the circumstances), necessitate what they build, and since the building claims we encounter most often are just partial building claims, they don't obviously involve necessitation. And although full building is necessitating, it looks symmetric. Thus, I've also introduced but left underdeveloped the claim that full building is identity: that claim is yet to be motivated in Chapter 4, defended against worries about multiple realizability in Chapter 5 and Chapter 6, and

in Chapter 7 against the idea that dependent things have differing properties that what they depend on. Rather than being picked out by directedness, necessitation, and generativity, I think building relations express dependence which can be rendered in true counterfactuals. I'm hereafter going to refer interchangeably to building and dependence relations. I think "dependence" neatly captures the resemblance class of relations Bennett calls "building" relations, and it already refers, in a way that is amenable to both ordinary and philosophical use, to what unites the variety of ways of saying that one thing depends on another.

Chapter 4

Puzzles of Synchronic Dependence

In Chapter 3, I discussed the way in which Bennett (2011, 2017) thinks building relations are unified. In partial contrast, I have given a somewhat different characterization of only the most common synchronic dependence relations employed by philosophers of mind, but also by philosophers of biology, science, and explanation. Synchronic dependence relations get used both in specific explanations or proposed explanations, as well as in general to formulate a variety of materialisms, physicalisms or naturalisms, or to deny them, with respect to the objects of a given domain of inquiry, as well as to formulate ways in which these domains themselves relate. I take it that the reason constitution, composition, and realization get employed in these contexts is that there are certain shared features among them. They are all counterfactual-supporting synchronic dependence relations. On the other hand, there are many treatments of each such putative relation, there is imperfect consensus whenever there is any at all, and there are puzzles at every turn. And that might seem like a reason to deny any sort of stronger unity among them. However, in this chapter, I argue that synchronic dependence claims, with one other key ingredient, give rise to puzzles that share the same basic recipe. Whenever some entity (object, event, state, property, or phenomenon) is claimed to be composed of, constituted by, or realized by some physical base that is considered to be sufficient for it, and also claimed to be distinct from that base, there is born 1) a grounding-like problem—an unanswered in-virtue-of-what question—and,

2) an overdetermination problem due to ontological overpopulation combined with the physicalist's commitment to the causal completeness of the world. I suggest that the shared recipe for the twin problems, and the prospect of similar solutions, are reasons to believe synchronic ontological dependence relations are closely related. Finally, since these problems arise in the context of each such relation, the antireductionist cannot appeal to any putative relation itself for solutions. Although it will take further discussion in Chapters 5 to 10 to demonstrate, the EM debate does not avoid questions about mind–physical world identity or non-identity by couching itself in one synchronic dependence locution as opposed to another. The EM thesis stands or falls with the more complete theory of a mind–physical world relation it entails or is entailed by. And, at last, no amount of constitution or realization of our minds by technologies would amount to the view that we are partly what we make, unless partial constitution and partial realization are partial identity.

Before considering, in Sections 4.2 and 4.3, the general puzzles that arise with respect to our concepts of entities, their properties, natures, and dependence on other things, it behooves us to review a handful of examples that deploy those concepts. I do this in Section 4.1, but with one caveat. Readers who have given the topics some thought are likely to think, *hang on, there are no problems here; we should simply reject this or that premise or way of speaking*. Indeed, that's exactly what we should do. Still, I want to make the immediate point that there is a common recipe for some very similar problems. I'll argue in Section 4.4 that this point is further substantiated to the extent that these problems have the same solution: we should reject a particular sort of recurring move or claim. So, of course I think there is a mistake being made in the recipes for these problems. Note, too, that the fact that there must be a solution to these problems does not mean there are no such problems. After all, part of the problem, or perhaps the underlying problem, is that philosophers disagree on the best solution. They disagree on where the recipe goes wrong. So, before anticipating one's preferred antidote to the unpalatable mixture, bear with me while I commit some mistakes in order to say what it is. If I'm right about the common recipe, there's just one mistake. And if I'm right about what the mistake is, to be discussed in the coming chapters, then we should think that

identity is the relation between an entity or property and all that it depends synchronically on, and, relatedly that anything on which an entity depends synchronically is constitutive of that entity.

4.1 Molecules and Minds

Consider Otto. Otto has a body, and we might say that Otto's body is composed of some tissue, composed in turn of molecules. Otto has a mind, and we might say that Otto's brain realizes his mind. And Otto is a person, we might say that person is constituted by either his body, his mind, their life histories, or some combination thereof. So, in our toy ontology, we have a person, called Otto, a mind, a body, a brain, and tissues. Now, let's consider various alterations to these. A rough sketch of a range of possibilities will suffice.

First alteration: one day, Otto eats a little extra, and gains a little weight, more molecules composing a little extra tissue. Later, Otto does some exercise, and loses some weight—in fact, he weighs a little less after exercising than he did before overeating. Still, there is Otto, there is his mind, there is his body, each respectively constituted by, realized by, and composed by an altered fusion of molecules.

Second alteration: with the advent of artificial intelligence and with the help of super-duper neuroscientists, Otto is able to undergo a procedure that uploads his knowledge, memories, preferences, and goals into a nano-technologically advanced bionic android capable of discriminating sights, sounds, smells, tastes, and touches. It is capable of planning and executing complex movements, and dynamically adjusting its responses according to changes in the environment. The bionic duplicate of Otto speaks Spanish in a smooth, low baritone, and improvises in a jazz band. The bionic duplicate of Otto passes Turing tests, makes friends, has worked his way through a few normal jobs, and is dating his most recent boss's daughter. Many of us are inclined to think the bionic body realizes a mind, and some even think these constitute a person.

Third alteration: as it turns out, in the upload process, biological Otto's brain is damaged such that he stares unflinchingly ever thereafter in whatever direction he faces; he does not move of his own accord, or respond to sights,

smells, sounds, tastes, or touches. In ordinary language, many of us will say Otto's mind is "gone." But, Otto's body, his damaged brain, and his tissues persist. And Otto, the person, is right there, too. At least, his family loves him, or tries to, and most of us will say he has the right to dignified treatment.

Final alterations: one day, Otto dies. And, one day, bionic Otto, chips fried by some electromagnetic impulse, is beyond repair. What does not go simultaneously out of existence, however, are the molecules composing the tissues and the nano-technological materials.

So we have a body that survives, in the first alteration, gains and losses of molecules and tissues. We have a mind that survives, through the second alteration, replacement of its realizer parts with different kinds of realizing parts. Through the third alteration, we have a person that survives drastic reduction in mental capacities, or loss of them altogether. And we have molecules that remain through the fourth alteration, despite not composing a living body, realizing a mind, or constituting a person.

Now, it matters less that some of us would be inclined to resist a number of the purported intuitions above (I certainly am) than it matters that ideas like these are available and defensible. Most importantly, the intuitions in play are arguably *conflicting*, as I'll explain below. But first, there's another ingredient in the mix to discuss.

4.2 Antireductionism and its Discontents

A hallmark question arises wherever we find synchronic dependence relations. The question is whether the dependent entity has all and only the same properties as the physical conditions on which it synchronically depends, and is thus identical with them, or is "something more," or perhaps just something *else*. Now, there is an obvious sense in which something can depend synchronically on something else and not be identical to it, that is, when one thing *partly* depends on another. I presumably depend on my right arm, today at least, but I am not my right arm. But consider complete dependence. Reductionism, roughly, is the view that when *a* synchronically depends on *b*, such that *b* is sufficient for *a*, *a* just is *b*. Broadly "antireductive" views, on the other hand, hold that

even when an *a* synchronically depends completely on some *b*, or some *b*-in-the-circumstances, *a* can have properties that *b* does not, so that *a* and *b* are not identical. Taking the higher-level entity to be something *more* is “emergentism.” Taking the higher-level entity to be something *else* can be broadly called “antireductive physicalism.” Now, notice that the key to having many of the above intuitions is to think for some reason that persons are not reducible to bodies or that minds are not reducible to brains, for example, because they have differing persistence conditions or modal properties—pertaining to what an entity is possibly or necessarily—like the ability to be differently realized. Similarly, the idea goes, bodies are not identical with molecules because bodies are wholes that survive changes of parts in everyday metabolism, let alone surgery. And yet there is clearly some dependence that persons have on their bodies, minds have on brains, and bodies and brains have on molecules.

So, whence the conflict among our ideas about the synchronic dependence of nonetheless distinct things? Let’s consider a few traditional problems.

The usual grounding problem (see Burke, 1992; Bennett, 2004; Heller, Heller, and Mark, 1990; Zimmerman, 1995) can be exemplified this way. Suppose Otto’s body constitutes a person, Otto. Under normal conditions, Otto’s body for a given duration seems fully sufficient for Otto: bringing it into being brings Otto into being. Yet Otto’s body can undergo changes that Otto does not, and vice versa. Otto’s body is still there when he dies, but his body no longer constitutes a person. These “various ways a thing *was*, *will*, *would*, *could*, or *must* be” (Wasserman, 2017) have been called “non-categorical properties.” So, by the law of indiscernibility of identicals, Otto cannot be identical with his body, since Otto and his body differ in their non-categorical properties. Otto and his body seem to be spatiotemporally *coincident* entities. The grounding problem is the question of what makes the difference between an emergent property or irreducible entity and its physical conditions, if not something among those physical conditions. What grounds their differing non-categorical properties? As Bennett (2004) notes, one doesn’t avoid the problem by claiming that something’s kind is grounded in its persistence conditions, any more than one can avoid the problem by claiming that its persistence conditions ground its kind.

A related problem, the problem of the many, consists in the incompatibility of my strong intuition that there is just one desk at which I now stand, with the quite natural thought that no particular subset of its parts have any greater claim to being a desk than do many other similar subsets. For example, certain edges of my desk are rough cut, and there are little splinters that don't seem particularly core to my desk. Even now, I can pull some off without much change, if any, to my desk. After all, my desk with a certain splinter and my desk without it are perfectly functional desks. Even supposing the splinter stays where it is, the thing without the splinter is right there, too. It is perfectly real; after all, it is attached to the splinter. And it is a desk. In fact, there are a great many overlapping groupings of wooden bits that have equal claim to being desks. But this seems like an undesirable result, perhaps more so in the case of bodies. Consider our strong intuition that human Otto has just one body, with the quite natural thought that no particular subset of his body parts have any greater claim to being that body than do many other similar subsets. For example, certain bits of skin and sweat could be excluded, and certain bits of protein could be included, and the body would still be a body—the body has the property of being such that it would still be that body. Suppose the largest fusion of molecules or tissues sufficient for that body includes a certain fingernail. But a fusion without the fingernail is right there, perfectly real; after all, it is attached to the fingernail. So it seems there are multiple sets of cells equally functional as bodies.

Another connected problem is what I'll call the problem of regularity among functional kinds. Multiple realizability has long been used to motivate antireductionism with respect to mind–body relations. The story goes that certain neural processes realize pain in humans, quite different neural processes realize pain in octopi, and very different processes could realize pain in martians or androids. As in the second alteration imagined at the outset, Otto's mind, it is supposed, might be realized by different mechanisms with the same causal profile over some significant range of inputs and outputs. Moreover, the usual story (see, for example, Fodor, 1981) is that it doesn't even matter whether silicone transistors ever *in fact* realize pain; it only matters that it is possible in principle for them to do so. It is possible in principle, the story goes, because

functional kinds are individuated by effects that can be brought about in various ways. My mental state now, then, has the property of being such that it might have been realized differently, and hence, by the law of indiscernibility of identicals, is not identical with any particular physical bases, which have no such property. Although the resulting issue has been picked up by others in the philosophy of mind (see Shapiro, 2000; Polger and Shapiro, 2016), Kim (1992b) was perhaps the first to argue in detail that we shouldn't expect there to be any induction-supporting, "projectible," or law-like regularities among multiply realizable functional kinds. Natural kinds, Kim argues, fall under the kinds that they do in virtue of their causal powers. He argues physicalists should believe that causal powers of higher-level entities are inherited from their lower-level physical bases. Different physical bases will differ in their causal powers, so they will not fall under the same natural kind, and there will be no interesting (ie, "non-analytic") truths about all instances of, for example, "pain," if "pain" can be realized by neurons or by silicone-based transistors. We'll see more of this problem in Chapters 5 and 6.

The last example brings us nicely to what is known as "the exclusion problem," which arises most often in the context of mind–physical condition antireductionism. Consider the apparent spatiotemporal coincidence, but non-identity, of Otto's love for his most recent boss's daughter with the realizers supposedly sufficient for that love. As Wilson (2001) reminds us, "the sufficiency thesis is implicit in standard conceptions of realization" (p. 4). That is, realization is understood as a relation of determination. "Metaphysical sufficiency would seem to have the strength to underwrite an account of mental *causation*", thus making "realization a metaphysically robust relation," and yet "problematic because, so-construed, physical realizers states, themselves being physical, seem to have no room for distinctly *mental* causation" (pp. 2-3). The trouble, then, is that if a mental state and its physical realizers are distinct, which of these *two* entities causes Otto to send flowers? Normally, we say the love does. But, quite clearly, with his bionic brain removed, no such behaviour occurs, so it is difficult to say that Otto's brain doesn't also cause them. But if Otto's brain does the work, it seems that his mental states never cause anything (see also Prior, Pargetter, and Jackson, 1982). Antireductionism, as

a metaphysical thesis which claims that the brain state and the mental state are ontologically distinct, can't also say they are the *same cause*, but if they *both* cause some wincing, such wincing would be causally “overdetermined.” Causal overdetermination violates either of two other plausible notions about the world, namely that it is causally complete (or “closed”), and that effects don't usually have more than one sufficient cause or set of causes. Bennett describes the problem nicely, “As it is often put, the physical does all the work, and there is nothing left for the mental to do” (2003, p. 471). We'll see more on this problem in Chapter 7.

4.3 A Recipe for Problems

The idea that these issues are related is not without precedent. Kim, for example, has wielded worries about exclusion to produce reductionist arguments in a whole array of contexts (see Kim, 1997, 1999, 2000, 2006, 2012). And Bennett (2004, 2017) has made some cursory remarks about the similarity among these problems (though perhaps not the problem of regularity among functional kinds). Bennett (2004), while discussing the usual grounding problem, says,

[The problem of the many] concerns the idea that the world contains any number of objects with only minutely differing *spatial* boundaries. The [grounding] problem here concerns the idea that the world contains any number of objects with only minutely differing *modal* boundaries. I therefore dub it “the modal problem of the many.” (2004, p. 360)

Although this dubbing is appropriate, we're about to see that the usual problem of the many is *also* a modal problem of the many. Finally, Bennett notes that “the problem of the many is a cousin of the exclusion problem” (2017, p. 75). She thinks of the problem of the many as demonstrating compositional overdetermination, and that compositional overdetermination *resembles* causal overdetermination. It would be strange if behaviours were double-caused by mental states and brain states, and,

it would be just as puzzling if, given one complete ground G_I of some fact or thing A , we were forced to also countenance a second

complete ground G_2 . After all, G_1 is by itself enough, at least in the circumstances, to guarantee that A obtains or exists; given G_1 , what work is left for G_2 to do? (2017, p. 75)

Thus Bennett gestures at, but doesn't quite spell out, how these problems are related. So I'm going to spell it out, because I think the problems are closer than cousins, and I think they have the same solution, though I'll come to that later.

Take the following sets of claims considered in parallel:

| | A | B | C |
|---|--|-----------------------------------|--|
| 1 | x is composed of $(n)y$. | x is realized by y . | x is constituted by y . |
| 2 | $(n \pm 1)y$ would be sufficient for x . | z would be sufficient for x . | y' would be sufficient for y , but not for x . |
| 3 | $x \neq (n)y$. | $x \neq y$. | $x \neq y$. |

The A-set of claims roughly sets up the problem of the many. The B-set of claims roughly sets up of the problem of regularity among functional kinds. The C-set of claims roughly sets up puzzles of material constitution, and the usual grounding problem. Below I'll say what further physicalist commitments turn these into puzzles, but first, I want to note the similarities among these antireductive components of each setup.

First, notice the similar forms of A1, B1, and C1. It should be understood as implicit in each claim that y or the y s are sufficient to synchronically instantiate x . That's the first key ingredient: x 's synchronic dependence on some fully sufficient base. Without the sufficiency claim, none of the problems get off the ground at all, because if y is just *part* of what it takes to make an x , then there's no problem with the non-identity of x and y .

Second, A2, B2, and C2 all involve claims about non-categorical properties. That is, they all involve claims about the "various ways a thing *was*, *will*, *would*, *could*, or *must be*" (Wasserman, 2017). Bennett (2004) calls these "sortalish" properties. She says,

I will use the label "sortalish properties" to refer indiscriminately to

- (i) persistence conditions, particularly modal properties like being essentially shaped about like so,
- (ii) kind or sortal properties, and
- (iii) properties that things have partially in virtue of their instantiation of properties in categories (i) or (ii). (2004, p. 341)

A2 and B2 involve x 's sortalish properties, and C2 involves y 's sortalish properties. One of the ideas motivating the problem of the many is that it can be indeterminate whether such and such a bit of stuff is part of some entity, and that a thing with or without certain parts *would* still be the thing that it is. This is another denial of mereological essentialism—the idea we saw in Section 3.3.3 that parts are essential to their wholes (Chisholm, 1973). The set up to the problem invokes the sortalish properties of even ordinary objects. Likewise, the idea of multiple realizability is that some other kind of physical base *would* be sufficient to instantiate a given mental state, event, or property. The idea of constitution is that one entity can constitute another, but *would* not do so if it was altered in a certain way or found in some other circumstances, even while still existing as the entity it is.

Finally, the sortalish properties in the second claims entail the failed identity in A3, B3, and C3. This distinctness is the second key ingredient. I should note at this point that antireductionists have not *typically* invoked the sortalish properties of things like desks in order to make claims about their distinctness from bits of wood and so forth. However, by my lights, if antireductive physicalism was true—though it isn't—it would be true about tables as well as minds. Indeed, as we'll see in Section 4.4, one strand of reductionism has used these problems as motivation to reject the reality of such entities and their properties.

Two things result, in all three cases, regardless of their formulations in terms of composition, constitution, or realization. The first is what I'm calling a grounding-like problem for each. Rereading A1 in light of A3, we might ask, what grounds the fact that some y s compose an x ?¹ In the case of multiply realizable but irreducible functional kinds, we can ask, what grounds the

¹ I take this grounding-like problem to be very similar to the question of “special composition,” due to van Inwagen (1995), which is the question of when some plurality of things compose some single thing.

supposed sameness (and projectibility) of multiply realizable functional kinds? In the case of constitution, the grounding question is familiar: what grounds the difference between supposedly distinct yet spatiotemporally coincident entities? Notice, too, that there is a spatiotemporal coincidence of putatively distinct realizer and realized entities. And notice that there is spatiotemporal overlap in the case of the many possible slightly varying fusions that would compose ordinary objects. Thus, in addition to the grounding-like problems, the second result of each setup is an exclusion problem. When y gives rise to x , they are supposed to be distinct, but real and causally efficacious. This overpopulation is problematic when added to the physicalist's commitment to the causal completeness of the world and the plausible thesis that effects don't usually have more than one sufficient cause or set of causes. I noted above that Bennett thinks the grounding and exclusion problems are cousins, because compositional overdetermination *resembles* causal overdetermination. I suggested that they are twins, because they are produced by the same basic recipe: some y that is fully sufficient for some synchronic x , but not identical to it. In fact what makes compositional overdetermination strange is that the overpopulation would result in causal overdetermination. If there was an actual plenitude of mostly coincident fusions of tissue building a plenitude of Ottos when there appears to be one, then they would, oddly enough, overdetermine my judgement that there is just one Otto. So, too, with the realization of Otto's mind by his brain. If his brain is sufficient for his mind, and yet not identical with it, and yet both are real and causally efficacious, they overdetermine their effects. So, too, with the bodies and the persons they constitute.

Now, the fact that A2 and B2 involve claims about x 's sortalish properties while the C2 involves claims about y 's sortalish properties suggests more room in the logical space of problem setups than the three I gave. I suspect that a puzzle can be set up for each putative way of altering some physical base (by addition, subtraction, or partial or complete kind-replacement) such that some higher-level entity persists, and one for each way a higher-level entity might not persist while its physical base does. My hypothesis is that there will be a grounding-like problem and an exclusion problem that arises for any case in which we affirm the dependence of some thing on a fully sufficient base, yet

deny the identity of the two.

4.4 Prospects for Grounding and Just Enough Determination

Further evidence of the similarity among concepts of dependence is the way that relations, relata, problems, and solutions are interwoven in the literature. This interweaving, among other things, is part of what I take Bennett to mean when she says, “we use the various terms like ‘compose’, ‘realize’, ‘arise from’, or ‘get out of’, in so many mixed up motley ways, with so many kinds of relata. The hard part is untangling the concepts, not seeing that they are somehow related” (2017, p. 30). Take the Stanford Encyclopedia of Philosophy, which comprises articles on relations, such as “Material Constitution” (Wasserman, 2017), on relata, such as “Ordinary Objects” (Korman, 2016), on problems, such as “The Problem of the Many” (Weatherson, 2016), and on purported solutions, such as “Relative Identity” (Deutsch, 2017), to mention just a few. It is not difficult to see that each of these are inroads to much the same terrain—they intersect everywhere and reference many of the same important issues and contributions. For example, in “Ordinary Objects,” Korman (2016) discusses the prospects of being permissive or eliminativist with respect to ordinary objects, and the issues motivating those solutions with respect to ordinary objects are exactly the ones we’ve seen: Korman (2016) canvases the problem of the many, puzzles of material constitution and grounding, worries about overdetermination, as well as others which can be understood as involving grounding-like problems.

Earlier, I said what mattered is that there are ubiquitous and conflicting intuitions available about dependence and distinctness. Of course, these inconsistencies need to be remedied, and I’ll offer some remarks here before wrapping up. However, rather than survey any solutions in detail, I want to draw out two points. First, I suggest that any given line of solution to one setup of problems is a potential solution to other setups. Second, I suggest that because grounding and overdetermination problems, and their possible solutions, cross-cut concepts of dependence, appealing to the relations themselves won’t fix anything. And while there are certainly differences between total sets of

philosophical commitments, because of the orthogonality of such differences to differences in synchronic dependence terminology, synchronic dependence concepts themselves do not entail or even clearly mark any such differences.

Now, except for explicitly linking realization and its puzzles with other synchronic dependence relations and their puzzles, I am by no means alone in thinking that similar puzzles recommend similar solutions. One *prima facie* reason to think so is that many of those philosophers offering such solutions have intended them to solve more than one puzzle. To mention just a few examples, Geach's (1980) notion of relative identity, which denies the indiscernibility of identicals, has been applied to both the problem of the many (see Weatherston, 2016) and to puzzles of material constitution (see Wasserman, 2017). Similarly, constitution theorists like Johnston (1992) have intended their accounts to offer solutions to both the problem of the many and to puzzles of material constitution. Van Inwagen (1995) opens his broadly eliminativist account by saying, "The metaphysically puzzling features of material objects [including organisms] are connected in deep and essential ways with metaphysically puzzling features of the constitution of material objects by their parts" (p. 18), positioning his view as a solution to those puzzles. Chisholm (1973) considers "a philosophical puzzle pertaining to the concepts of whole and part," and then says, "The proper solution, I believe, will throw light upon some of the most important questions of metaphysics" (p. 581). I take it that this list could be continued at length. As Bennett (2017) remarks, "The familiar problem about causal exclusion does arise for determination more broadly, and the same panoply of solutions is available" (p. 76). While Bennett is reasonable to take that as obvious, I will flesh this out just a little further.

Let me sketch a few very broad lines of solution.² First, one might deny the

² Because my treatment is so broad, I offer the following caveat. There is a well-blended spectrum of views and resulting uncertainty as to how a particular view should be characterized, for example, whether as a form of deflationism or eliminativism. Chisholm (1973) positions himself as deflationary insofar as he claims not to be rejecting any of the apparently conflicting intuitions that give rise to a particular puzzle, but claims to be showing them not to be conflicting when properly clarified. On the other hand, Wasserman (2017, sec. 4) says that Chisholm's is "an 'eliminativist' view insofar as it denies the existence of mereologically ductile objects." The rough taxonomy I'll offer attempts to be exhaustive over a wide domain,

distinctness that leads to the puzzles to begin with. There are two broad ways of denying that some dependent entity is distinct from the physical conditions upon which it synchronically depends, for any given concept of dependence.

Broadly eliminativist positions embrace the unintuitive claim that there is no such x related to y in that way. For example, Unger (1979) famously denies the existence of ordinary objects, and some eliminativists hold that composition does not occur (for defenders of this view, see Korman, 2016, n. 3). Note, here, a connection with eliminativism in the philosophy of mind. Eliminative materialists, as I noted in Section 2.1, hold that our ordinary psychological terms fail to carve the world at its joints, and that for the purpose of explanation such terms should be—and, in the process of scientific investigation will be—replaced by references to, for example, neurological processes or kinds thereof (see Ramsey, 2003). The problems that suggest such a similar solution are the same. Roughly, in the case of composition, if it is indeterminate whether an object is composed by one of many possible similar fusions of parts, then, strictly speaking, there is no such object. Roughly, in the case of realization, if “pain” is not identical to its various realizers, because any of them are sufficient and none of them necessary for the realization of pain, then, strictly speaking, there is no such natural kind as “pain” that can figure in any useful generalizations.

Broadly ontologically deflationary views argue that our talk of the synchronic dependence of higher-level entities needn’t entail any further “ontological commitments.” I take ontologically deflationary views to be reductive insofar as they deny the *ontological* distinctness of the relata in cases of composition, constitution, and realization (despite variously allowing epistemic, pragmatic, explanatory, or predicative distinctness). For example, composition is a relation that arguably holds between different grain-levels of description of one and the same world. Lewis (1991) argues that composition is identity and that mereology is ontologically innocent, and many have followed him (see Wallace, 2011b, 2011a; Cotnoir and Baxter, 2014; Smid, 2015). Constitution, too, has

at the manifest expense of detail. If I’ve done my job well, every particular brand of “ism” should fall under the broader class for which I’ve borrowed the same label, but at the end of the day, I’ll happily relinquish any particular label.

been subject to deflation in terms of identity (see Noonan, 1993; Biro, 2017). And the thought that realized entities are not reducible to their realizers because different entities could fulfill the same role has been challenged, especially by Kim (1992b), Shapiro (2000), and Polger and Shapiro (2016), who consider their positions to be more deflationary than eliminativist.

To deny distinctness in either way is to be some sort of reductionist. Reductionists do not face the grounding and overdetermination problems, but may face other problems. For example, one might leverage overdetermination worries in service of the claim that mental and aesthetic properties are physical ones, only to accept the burden of showing how. And reductionists have to account for variability and multiple realizability—a project for Chapter 6.

How does antireductive physicalism fare? Earlier, I suggested a possible distinction between emergentism and antireductive physicalism. I hold this distinction among antireductionisms very lightly, for two reasons. First, although it is certainly questionable whether any emergentism can count as a form of physicalism (see Kim, 1999, 2006), some proponents hold that emergence is compatible with physicalism. Of course, everything hinges on the details of any given view, rather than the label a view claims for itself. So, some “emergentisms” may be forms of antireductive physicalism. Secondly, and perhaps more importantly, one question I am pressing here is how well antireductive physicalism itself fares as a form of physicalism. By my own lights, then, the distinction between emergentism and antireductive physicalism is not to be taken as crucial here. But suppose that emergentism and antireductive physicalism are at least distinguishable in a sense suggested by Jessica Wilson (2015), albeit in different terms³: the former denies the causal completeness of the world, and the latter denies the requirement that no effects have more than one sufficient cause (or set of sufficient causes). It is the latter view, the one that attempts to be a form of physicalism, with which we are concerned

³ Wilson calls what I’m calling emergentism and antireductive physicalism, “strong” and “weak” emergence, respectively. It is not entirely clear whether Wilson takes her taxonomy to exhaust the antireductive strategies, although its exhaustiveness is implied. If I’m right to understand her taxonomy that way, then I’m right to make the rough correspondence between her terms and mine.

here, and in more detail in Chapter 7.

Various positions of the latter stripe embrace the somewhat unintuitive ontological plenitude⁴ that results from coincidence and distinctness, and yet attempt to be compatibilists with respect to the coincident entities' causal powers. Compatibilists take the further ontological commitment entailed by the distinctness claim not be a matter of *over*-population, as I've called it, but more like an enrichment.⁵ The strategy is to say how the spatiotemporally-coincident yet distinct entities are not in fact causal competitors, by claiming that they are somehow closely related, despite being non-identical. I discuss this strategy in detail in Chapter 7, where I argue that it fails.

What about an even closer relation between, for example, mental and physical entities? Consider "token identity" in the philosophy of mind. Token physicalism is the view that mental kinds are not identical with physical kinds, but that every particular mental entity (event, property, or state) is identical with a particular physical entity (event, property, or state). Token identity, like necessitation, thus *appears* to avoid exclusion worries. The token physicalist can remain antireductive by denying the identity of mental kinds with physical kinds. But token physicalism is more plausibly physicalist by accepting that a given mental state *is* a physical state. Now, I don't know of any very close analog of token identity as a response to related problems formulated in terms of constitution or composition. But if there is a defensible token physicalist thesis with respect to the realization of minds by brains (or otherwise), it might shed light on puzzles of material constitution and grounding problems. If it did, that would suit my point here. On the other hand, the reason for the gap in the literature corresponding to this part of the logical space might be the incoherence of token identity in the case of more ordinary relata. And there

⁴ I emphasize the "ontological" aspect here because views I've characterized as deflationary, like Lewis's (1991) argument that mereology is ontologically innocent, can be "permissivists" similarly accepting of "universalism," the doctrine that any distinct fusion composes an entity; it's just that that entity is not a new one in the purported ontology. Whether anyone can get away with this uncommitted permissivism is controversial, of course.

⁵ This, I confess, is just my own characterization. I leave it intentionally vague, except to say that I mean to signal the oft-defended intuition that something is lost in the reductionist's picture of the world (especially the eliminativists').

is good reason to question the coherence of the view in the mind–body case. Kim (2012) thinks no intelligible thesis of token physicalism has been offered. That would be embarrassing. How could such a popular view be indefensible? If Kim is right, it could be that token physicalism has stood on the apparent merit of multiple realizability-based arguments against reduction, more than on positive construals of the thesis. I will not rehearse Kim’s arguments against his best reconstructions of token theses, however, because there is a very direct way to see the threatened incoherence. The token physicalist claims that mental entity types are distinct from physical entity types. But this means that mental tokens, supposedly identical with physical tokens, have differing sortalish properties (they fall under different kinds and the mental has the modal property of being possibly otherwise realized). But if mental tokens differ in their properties from physical tokens, they are distinct, offering up a grounding problem. In short, if one hangs onto the difference in types, one gets a difference in tokens as well.

I suggest that the grounding-like issues are similarly solved by antireductionism at the expense of distinctness. Let me illustrate this with respect to Baker’s (1999) antireductive physicalism formulated in terms of constitution. I understand Baker as antireductive because she explicitly takes her position to have an “antireductive thrust,” (p. 152), and thinks constituted entities have properties that their constituents do not. Baker thinks constitution, as she formulates it, holds between persons and their bodies, art objects and their media, monetary value and monetary instruments, national flags and bits of cloth, to name a few. If we say that a paper note has some value, that is so because the paper bill constitutes some thing with that value—a dollar—and not the other way around. The constituent can “borrow” properties that really belong to the constituted entity. Secondly, she thinks these properties aren’t spooky. So she dissents “from Allan Gibbard when he says, ‘If the statue is an entity over and above the piece of clay in that shape, then statues seem to take on a ghostly air’” (p. 151). Her “antireductive” view aims to be a form of physicalism because of her non-ghostly answer to the grounding problem: “it is in virtue of its relational properties that David exists. . . . Piece could exist in the absence of an artworld; David could not” (p. 151). The differing modal

properties of Piece and David depend on, or are somehow “grounded in” circumstances, presumably physical ones. Plenitude-embracing compatibilists of both constitution and realization stripes can (and do) appeal to circumstances to get the grounding done, and they should. But once corrected in this way, the resulting view fails to be antireductive in the way Baker takes her view to be, as we’ll see in more detail in Chapter 7. Notice, for now, that this appeal to sufficiency-in-the-circumstances is no self-sufficiency at all on the part of pieces of marble or neural processes. At best, it shows that things like statues and mental entities are not identical with the marble and the neural processes on which they *partly* depend. But we already knew that wholes aren’t identical with subsets of their parts. If the realization theorist or constitution theorist retreats to sufficiency-in-the-circumstances as grounding sortalish difference or sameness, and the necessitation of a realized or constituted entity by realizing or constituting entities-in-the-circumstances, they solve grounding and overdetermination problems by giving up distinctness—and thus antireductionism. That is, it is no longer clear how David is distinct from all of Piece-in-the-circumstances, or how a functional property is distinct from all of some physical property-in-the-circumstances. In light of the grounding problem, antireductive physicalism looks either less antireductive (on Baker’s account), or less like physicalism. For an example of antireductionism as dubious physicalism, consider Fodor’s (1997) conclusion that it is “entirely mysterious” why there should be “macrolevel regularities at all in a world where, by common consent, macrolevel stabilities have to supervene on a buzzing, blooming confusion of microlevel interactions” (p. 161). Hence the persisting worry that truly antireductive views have no answer to the question of what grounds the supposed difference between spatiotemporally coincident but non-identical entities.

Moreover, because grounding and overdetermination problems, and their possible solutions, cross-cut concepts of dependence, appealing to the relations themselves won’t provide the required remedy. Stipulating that the dependence between mind and body or between art objects and their media is a matter of realization or constitution in the circumstances doesn’t on its own go any distance to alleviating these problems while maintaining distinctness. It should be especially obvious that it is no answer to grounding or overdetermination

problems to insist on distinctness by saying that the relation which holds between, for example, body and mind, is one of realization or constitution. Yet this is arguably what at least early theorists about those relations have done. Concepts of dependence can directly beg an outstanding question by taking certain otherwise unaccounted-for intuitions about sortalish differences and simply baking these into the concept that was supposed to do the work of explaining, for example, mind–physical dependence. In the case of realization, the antireductive lesson that was supposed to be *drawn* from the possibility of multiple realization, but became baked into the very concept of realization, remained unquestioned for decades (see Kim, 1992b; Shapiro, 2000), and might still be the dominant view in the philosophy of mind and cognitive sciences. Even more recent theories of realization (see Gillett, 2002b, 2002a, 2013; Aizawa and Gillett, 2009) have been accused of relying on an unspecified “in virtue of” or “grounding” relation (see Endicott, 2016a). Similarly, Chisholm (1973) appealed to an unexplained notion of constitution to account for the persistence of objects through change of parts. Baker’s (1997, 1999) view is largely stipulative in a similar way: begin with the thought that a higher-level thing and its lower-level constituents have differing sortalish properties. Use these properties to deny their identity and to insist that the relation between them must be some other relation: one of constitution. Finally, invoke constitution (in the circumstances) to explain the sortalish differences.

Wrapping Up

In this chapter, I’ve argued that there’s a grounding-like problem for each of the above three ways of denying identity while affirming sufficiency, and, although the problem of overdetermination arises most often in the context of mind–body antireductionism, it is actually common to every instance of putative sufficiency paired with distinctness owing to differing sortalish properties. Following Bennett (2004), but extending the analysis to realization, and calling the latter the “many problems of the many,” highlights the problematic overdetermination or exclusion problem resulting from overpopulation. Similarly, calling all of the former “grounding-like” problems highlights the resulting “in virtue of what”

questions: in virtue of what does just one of many possible fusions compose an entity; in virtue of what are distinct kinds of bases sufficient for some unified higher-level kind; in virtue of what are coincident entities distinct? Finally, no putative synchronic dependence relation is immune to these problems. They arise from antireductive approaches to the synchronic dependence of higher-level entities on lower-level ones, whether formulated in terms of composition, constitution or realization, which I suggest is a reason to suspect that these may not be such different forms of dependence after all. At least, no mere appeal to these relations secures antireductionism. Moreover, views appealing to circumstances to get the necessitation and grounding needed to solve the overdetermination and grounding-like problems simply doesn't rule out identity. If anything, the best efforts to deal with overdetermination and grounding-like problems are appeals to circumstances that may yield a form of identity, if a widened one, after all. But I am only just beginning to defend this view. It remains to be seen, in the following chapters, why intuitions in the vicinity of multiple realizability, as the initial motivations for antireductionism, fail to yield the distinctness that is supposed to block reduction.

Chapter 5

Antireductionism and Multiple Realizability

We've seen that problems persist for antireductionism in a rather patterned way, owing to commitments to both physicalism and to non-categorical properties, when the entity bearing those properties is taken to be coincident with some physical conditions on which it depends and yet which are insufficient for those properties. While I take the puzzles as reasons to doubt antireductive physicalism can be coherent, I don't take them to be a reason to give up on physicalism generally. Two straightforward reasons to pursue physicalism are the lack of viable alternatives, and the historical tendency toward satisfactory explanations of higher-level entities in terms of lower-level ones. Of course, the idea that we are partly what we make, is an extension of physicalism—the idea that we are physical things, and that physical things like technologies can partly explain physical things like us. Part of my purpose in this chapter, then, is to reject one of the arguments for antireductionism: the argument from multiple realizability. But one central way of countering antireductionism does so by rejecting the functional kinds to which the multiple realizability argument appeals as being unscientific, or “unprojectible.” The idea is that since the members of functional kinds inherit their causal powers from their realizers, and scientific kinds are causally individuated, putatively reduction-blocking differences between realizers will mean that they fall under different scientific kinds. Correspondingly, those that do fall under a scientific kind aren't “multiply” realized; differences among members of such kinds are said

to be “mere variations.” I argue in this chapter that these arguments against antireductionism fail. I argue in the next chapter that realizers need not share all their causal powers to fall under a kind, and that kinds come in degrees of similarity between their members, and hence degrees of projectibility. As I argued in Section 2.1, it matters to the seriousness of externalism that functional kinds are not unscientific. That would threaten with elimination those broadly functionalist models of mind current in the cognitive sciences on which the appeals of EM—to parity or similarity with extended states or processes—are based. It would in fact threaten many scientific kinds from disciplines other than physics. My aim here, then, is to navigate between the Scylla of antireductionism and the Charybdis¹ of eliminativistic reductionism. Externalists must be able to account for the projectibility of functional kinds, and I must do so in a way that is amenable to reduction. As I said in Section 1.2, a broadly reductive functionalism is required for the inference from our partial constitution by technologies to our partial identity with technologies. The question of EM hinges on nothing less than the question of mind–world relations, to which issues about multiple realizability are central. Again, as Bickle (2016, sec. 3) says, “what is at stake here should not be underemphasized. . . .” The issue of multiple realizability remains central to the direction of fruitful sciences of minds and human persons.

The issue will occupy us for this chapter and the next. In Section 5.1, I introduce the antireductive argument in more detail and summarize some ways it might be attacked. In Section 5.2 and Section 5.3, I explain that the best existing line of attack fails. Then, in Section 6.1, I elaborate a notion of similarity that puts kinds on a spectrum of similarity. This view of kinds will eventually allow me in Section 6.3 to deny a different premise of the antireductive argument.

¹ In Homer’s *Odyssey*, Scylla and Charybdis were legendary monsters inhabiting two sides of a strait so narrow that it was difficult to sail around Scylla without passing close to Charybdis, and vice versa.

5.1 The Argument for Antireductionism from Multiple Realizability

A widely held view in philosophy, largely attributable to Putnam (1967), remains that multiple realizability repudiates mind-brain identity, precludes the reduction of mental states to brain or other physical states, and secures the autonomy of psychology. Fodor's (1974) "Special Sciences" then generalized the argument from multiple realizability to all "special sciences"—those sciences other than physics. The idea, in short, is that as much as the world is populated by "lower-level" stuff like particles, molecules, and gravitational fields, it is also populated by "higher-level" things like organisms, institutions, economies, norms, values, art, people, goals, feelings, and minds. These are all quite real and scientifically interesting things that are not just difficult to identify with the lower-level kinds, but, the idea goes, *impossible* to identify with such kinds because the higher-level things are just not the right kinds of things to be so identified. They're not the right kinds of things to be identified that way because what it is to be an institution, value, norm, person, or mind is to *function* in a certain way or set of ways. And functions are *multiply realizable*: they can be performed by a variety of mechanisms, each supposedly sufficient and yet none specifically necessary for the higher-level thing identified by that function. A functional kind, then, is not identical with *any* physical kind. But this gets us into the problems we saw in Chapter 4. And it would preclude making claims about our *identity* on the basis of any mental properties we might have that are realizable by partly technological physical properties. Finally, if constitution and realization are as closely related as I suggest, the same sort of considerations might apply to, for example, multiply constitutable cognitive systems. Being partly constituted by technologies wouldn't, on the antireductive view, mean that we are partly what we make. The argument for antireductionism, in a simple form which I adapt from Bickle (2016), may be characterized as follows:

[**Multiple Realizability**] Many special science kinds are multiply realizable by distinct physical kinds.

[**Unity Condition on Kind Identity**] If a kind is multiply realizable by distinct

physical kinds, then it cannot be identical to any specific physical kind.

Multiple Realizability and the *Unity Condition on Kind Identity* straightforwardly entail,

[**Non-Identity Thesis**] Many special science kinds are not identical to any specific physical kind.

This conclusion gets combined with a condition on reduction, such as:

[**Identity Condition on Reduction**] If a kind is not identical to a physical kind, then it is not reducible to physics.

From the *Non-identity Thesis* and the *Identity Condition on Reduction*, we can conclude,

[**Antireduction Thesis**] Many special science kinds are not reducible to physics.

I will endorse *Multiple Realizability*, but I'm eventually going to reject the *Non-identity Thesis*, and the *Antireduction Thesis* with it.

Of course, the *Antireduction Thesis* has not gone unchallenged, and the issues remain outstanding. In his review of the literature, Bickle (2016) says, “the renewed critical interest in multiple realizability, begun more than a decade ago, continues to the present day” (sec. 3). First, the plausibility of the *Identity Condition on Reduction* has been questioned; the *Identity Condition on Reduction* rests on Nagel’s (1961) account of intertheoretic reduction, or something like it, and reductionists can deny that the *Identity Condition on Reduction* captures the right condition on reduction, and can subscribe to an account of reduction revised in light of multiple realizability (see Sober, 1999). But I will not dispute the *Identity Condition on Reduction*. To see why, recall that physicalist commitments to non-overdetermination and the causal closure of the physical, together with an entity’s dependence on lower-level conditions sufficient for it, become problematically inconsistent when combined with the thought that an entity has properties for which those conditions are *not* sufficient. If one accepts the *Non-identity Thesis*, that many special science kinds are not identical with any particular physical kind, one already has the grounding and overdetermination problems for special science properties. So

these problems arise independently of further inferences from distinctness to irreducibility. So although I have been speaking of “antireductionism,” my issue is really with “anti-identificationism,” a precursor to antireductionism, though one can usually put these two positions together, as I did throughout Chapter 4. From my point of view, what matters most is simply that we not posit coincident but non-identical entities. That’s because a plausible physicalism, whether properly called “reductionist” or not, must identify everything in the world with physical conditions. If these physical conditions are expansive, “reduction” seems to be a less apt label for the identification. But by denying the *Non-identity Thesis*, I needn’t worry about the rest. Having said that, I suspect antireductionism goes, too.²

The *Multiple Realizability* premise has also come under attack. There are two ways of denying that *Multiple Realizability* is true of a particular special science kind. The first is to deny that *the same* kind is realized by *truly* multiple physical kinds (see Shapiro, 2000, 2008; Bechtel, 2001; Polger, 2007, 2009; Polger and Shapiro, 2008, 2016; Shapiro and Polger, 2012). The second, related strategy is to deny that multiply realizable special science kinds can be properly scientific (see Kim, 1992b; Shapiro, 2000). These two approaches are closely related because a condition of the second strategy is that some putative kind *is* realizable by quite distinct physical kinds. Together, these strategies might be turned into a general counterargument to the *Antireduction Thesis* by running an apparent dilemma in which the *Multiple Realizability* premise turns out false in every particular case, like the dilemma posed by Lawrence Shapiro (2000):

Either the realizing kinds truly differ in their causally relevant properties, or they do not. If they do not, then we do not have

² If the *Non-identity Thesis* is rejected, then the argument for the *Antireduction Thesis* via the *Identity Condition on Reduction* is unsound. And it is not at all obvious what plausible condition on reduction replacing the *Identity Condition on Reduction* could be strong enough to block reduction without trivializing irreducibility (and reducibility). But, again, I needn’t settle that. I take it that the best general picture of intertheoretic reduction will be a consequence of an account of kinds and regularities in science and that reductionists ought not to beg the question against antireductionism by redefining reduction or relying on an account of reduction whose primary motivation is to deny the *Antireduction Thesis*.

a legitimate case of multiple realization. If the realizing kinds do genuinely differ in their causally relevant properties, then, it seems, they are different kinds. But if they are different kinds, then they are not the same kind, and so we do not have a case in which a single kind has multiple realizations. (p. 647)

I think cases that slip through both horns are the norm, and that the attempt to deny *Multiple Realizability* fails. Interestingly, I think the reasons to believe that multiple realizability is common are reasons to believe the *Unity Condition on Kind Identity* is false. Thus, to see why we should deny the *Unity Condition on Kind Identity*, it's important to see why no way of denying *Multiple Realizability* is plausible.

First, in contrast to the claim that multiple realizability is rare, I argue in Section 5.2 that it is ubiquitous. Second, in contrast to the claim that multiply realizable kinds are unscientific, I argue in Section 5.3 that they can be projectible. I will then argue in Chapter 6 that all kinds whose members are physical individuals, and kinds of such kinds, are physical kinds, and it is false that if a kind is multiply realizable by distinct physical kinds, then it cannot be identical to any specific physical kind.

5.2 A Dilemma for *Multiple Realizability*, and its First Horn

With a view to clarifying the notion of multiplicity involved in *Multiple Realizability*, Shapiro (2000) introduced a “causal” relevance criterion for the sorts of differences that count as differences among realizers. Not just any difference between mousetraps, for example, is relevant to the function of trapping mice, so not just any difference makes them different as mousetraps. For example, we shouldn't think that if you take a regular old snap-trap and paint it aquamarine on the underside that you have come up with another realization of a mousetrap. Shapiro makes the same point with respect to differently-coloured corkscrews. Shapiro believes that, “once various conceptual issues are clarified,” by which he means testing claims of *Multiple Realizability* against a relevance criterion, “the task of demonstrating multiple realizability smacks the hard surface of

empirical fact, and, I shall argue, leaves MRT far more difficult to establish than philosophers currently acknowledge” (pp. 636-7). Shapiro argues that “realizations ought to be distinguished in kind only if they differ according to how they achieve the capacity that serves to individuate the kind they realize” (p. 650). Failure to satisfy this condition is the first horn of the dilemma for putative cases of multiple realizability. And Polger and Shapiro (2016) further challenge *Multiple Realizability* using a relevance criterion, arguing that multiple realizability is special and not nearly so ubiquitous as some might think: “Variation is everywhere in nature, but multiple realization is not” (p. 70). On Shapiro (2000) and Polger and Shapiro’s (2016) view, multiple realizability occurs, but is not ubiquitous, in contrast to especially Carl Gillett’s (2002a) account of the realization relation.³ And it is on the issue of “relevant” differences that Shapiro (2000)⁴ and Polger and Shapiro (2016)⁵ stake this claim. The criterion is supposed to vitiate apparent cases of multiple realizability. In Shapiro, there are a few statements of the relevance criterion. The clearest statements are, “to say that a kind is multiply realizable is to say that there are different ways to bring about the function that defines the kind” (p. 644), and, “realizations ought to be distinguished in kind only if they differ according to how they achieve the capacity that serves to individuate the kind they realize” (p. 650). Let’s refer to it this way:

[Causal Relevance Condition] A kind is multiply realizable only if realizers differ in how they bring about the function that individuates the kind.

³ Gillett (2002b, 2002a, 2007), Aizawa (2009), and Aizawa and Gillett (2009), have sought to characterize realization as a sort of non-causal and synchronic dependence that instances of a single property or function have on physical conditions that are not qualitatively identical to one another, no matter where they fall on a spectrum of similarity. Because so many differences, including small differences, among the implementations of a property or function amount to multiple realizability on that account, Polger and Shapiro accuse Gillett, and Aizawa and Gillett, of promoting a notion of realization that entails an “undesirable profligacy” (p. 45) of instances of the realization relation.

⁴ Every further reference to Shapiro in this chapter and the next is to Lawrence Shapiro (2000), unless otherwise specified.

⁵ Every reference to Polger and Shapiro in this chapter and the next is to Polger and Shapiro (2016), unless otherwise specified.

At a glance, there are at least two ways to understand different “ways” and differing in “how.” One would emphasize parts and their arrangements. Another would emphasize functionally-defined properties “belonging” to an individual realizer. But no one should reply to the question of how a mouse trap traps mice with, “by having the disposition to trap mice.” The how-question is only sensibly answered in terms of parts and their arrangements. Shapiro says,

I do not deny Putnam’s [1979] claim that “two systems can have quite different constitutions and be functionally isomorphic.” I do, however, deny that two systems that are functionally isomorphic are, in virtue of this fact, realizations of the same kind. (p. 638)

The denial here can be clarified in terms of scientific, natural, or projectible kinds. Since the differences Putnam cites are differences in constitution—most plausibly interpreted here in terms of parts and arrangements—and the sameness he cites is functionally characterized, it would be consistent with traditional appeals to *Multiple Realizability* to accept a causal relevance criterion for differences when answering the “how” question in terms of parts and arrangements. For instance, mouse traps differ in how they catch mice by differing in the parts or arrangements without which they could not catch mice. But this must not be what Shapiro means by “how,” because he uses the criterion to “screen off” some differences of the above sort as not amounting to relevant differences in realizers. On Shapiro’s view, a steel corkscrew and an otherwise very similar aluminum corkscrew don’t count as different realizers of corkscrews. They do not relevantly differ, apparently, since both are rigid. “The means by which steel and aluminum achieve rigidity are, in this context [the context of corkscrew realization], beside the point” (p. 645). Why? “Relative to the properties that make them suitable for removing corks, they are identical” (p. 644). By apparent contrast, Shapiro suggests that a lever-style corkscrew, as compared to a winged corkscrew (which adds a rack-and-pinion), count as different realizers of the corkscrew function.

One might wonder whether levers and rack-and-pinions could all be functionally described, lumped together in much the way steel and aluminum are lumped as having “rigidity.” That is, levers and racks-and-pinions could be described as *devices yielding mechanical advantage in applications of force*.

Wouldn't that screen those differences off, too? In that case, as much as Shapiro can say, "relative to the properties that make [aluminum and steel] suitable for removing corks, they are identical" (p. 644), he should say that, *relative to the properties that make the two mechanisms suitable for removing corks, they are identical*. But this is like answering the how question by reference to a disposition. How do corkscrews remove corks? By being able to remove corks. On the other hand, admitting differences between aluminum and steel as differences between causally relevant parts, the appeal to their being "relatively identical" is subsequently just the *Multiple Realizability* claim, that things not identical in an absolute sense might perform the same function. Aluminum is not rigid simpliciter; it is rigid enough *as* a lever of a certain gauge in a certain application at a certain temperature and pressure. Steel and aluminum don't have identical rigidities, and their differences track all the way in to their differing atomic structures. As causally individuated, then, aluminum and steel, far from being identical, are only similar insofar as they have similar effects or share particular effects as they do in corkscrew function. Moreover, no physicalist would deny that *all* differences in corkscrews coincide with microphysical differences. Winged and lever corkscrews also differ in their parts and arrangements, and the other functional differences between the steel and aluminum cannot hide behind the property of being rigid or a specific functional description any more than the differences between levers and other force-focusing mechanisms can hide behind the functional description of providing mechanical advantage. No differences in parts of the corkscrew get screened off simply by being differences among smaller parts; there is no other source of difference.

There is a version of the difference criterion that I do not deny. I agree with Shapiro that not all of what we might call a corkscrew does any corkscrewing work. Lazy tag-along parts can be omitted. Shapiro notes that some properties, like colour owing to paint, are causally irrelevant. Now, aluminum and steel differ in colour. Does that make their colour properties relevant, since aluminum or steel or some material must be involved in the work, and all have colours? But this asks the relevance question of the wrong candidate. We will want to deny that colour is a difference-maker to the function in question. Of course,

for more specific dual-functions, for example, flaunting status while opening a wine bottle, a certain kind of corkscrew, for example, a golden one, might be required. But any downstream effects which are extraneous, *ex hypothesi*, to the function or functions in question, while nonetheless being effects of the working parts and their arrangement achieving the function or functions in question, are pertinent to the question of multiplicity, because it is those extra-functional effects that individuate the various realizers of a function. It is differing parts, not functional properties, that can be tested as to their relevance to the function in question. If the putatively property-bearing parts, like paint, of the irrelevant property, colour, could be omitted altogether and still get the job in question done, the *paint* is irrelevant, and the paint colour goes with it. One cannot altogether omit certain materials, of one sort or another (but not any sort), any more than one can omit a mechanism of one sort or another (but not any sort) and still realize a corkscrew. What suggests aluminum or steel and winged or lever-style corkscrews as potential cases of *Multiple Realizability*, having such multiplicity in the realizers, are differences between parts which are not just along for the ride, but doing work within the sufficient base of the realized function—parts which, though they can be traded, cannot be eliminated and still yield the realization. The appropriate differences to be tested for relevance are differences in parts or arrangements, and causal relevance should be decided counterfactually. I suggest that,

[**Causal Relevance**] An entity *a* is causally relevant to a function *e*, if and only if without *a*, and without compensatory work performed by other parts, and without replacement of *a*, that instance of *e* would not occur.

Two notes are in order. First, the “no-compensation” clause in *Causal Relevance* above must be in place in order to test causal relevance, or else the counterfactual will come out false *because* of multiple realizability. It must be in place in order not to exclude working parts—for example, the very tip of a very sharp corkscrew—as causally irrelevant just because a loss of some degree of sharpness might not preclude general function. In contrast, Polger and Shapiro claim that,

it is plain that isotopes of gold are not different ways of having the atomic number 79 (i.e., being gold)—they are not multiple

realizations of gold—because they all have their atomic number in the same way: Namely, by having 79 protons in their nuclei.

Indeed, elements are grouped by the number of protons in their nuclei. That number corresponds to the number of electrons the nucleus tends to attract, and that relationship is somewhat more predictive or definitive of the reactivity or causal profile of any one of an element’s observationally stable isotopes than the number of neutrons in the nucleus. Thus, the property of being gold and the property of being an element with 79 protons in its nucleus are identical. Now, is the latter multiply realizable? Polger and Shapiro want to say, no: isotopes have their atomic number “in the same way.” But that is like answering the question of the way or the how of mouse-trapping by reference to a thing’s being a mousetrap. Rather, according to the clarified causal relevance criterion I have recommended, protons are causally relevant because, being gold is *not* just being 79 protons; one from my left ear, another from your car steering wheel, another from Shakespear’s corkscrew, another from Saturn’s ring, and so on. Being gold is being any *element-with-79-protons-in-its-nucleus*, and being an *element* with that many protons depends on having between 90 and 126 neutrons. Neutrons are essential to the stability of the nucleus, hence the range of stabilities of isotopes of a given element. Without the neutrons, the element would not be gold. Now, it happens that there is only one observationally stable isotope of gold. But it would be false to think that a gold-195 (79 protons and 116 neutrons), would not be gold, or that a block of it—if a block could be synthesized—with a half-life of 186 days, would not be a gold–platinum alloy over the course of its relatively leisurely change into increasingly pure platinum (98% pure after a few years, by my calculations). Many isotopes of a single kind of element are observationally stable. Hydrogen, for example, is mostly two observationally stable isotopes—protium and deuterium. These isotopes differ somewhat in their properties and reactivity—they can be causally individuated—and can be sub-classified (see Bursten, 2014).⁶ Water with an

⁶ For some of the same reasons, Julia Bursten (2014) follows a number of philosophers of chemistry in rejecting the microstructural essentialism—inherited from the Kripke-Putnam view of the extension of our natural kind terms—that might underlie Polger and Shapiro’s view.

unusually high proportion of deuterium is known as heavy water, and differs somewhat in its effects (e.g., on a weigh scale), but perhaps not in some of its other effects in particular branches of a causal network, eg, slaking thirst. Protium and deuterium differ in causally relevant ways. By way of analogy, think of standard stud wall construction. A wall without enough studs is not stable. The design of a wall with too many studs is not economical. Still, in any case, the work of holding up a roof above is spread among the studs. Likewise with corkscrews composed of levers differing in length, shape, or thickness. It would be mistaken to think that a bit of aluminum was causally irrelevant—not participating in force transmission—just because it could be drilled out, or cut away. Lots of things—nuclei, houses, corkscrews, bureaucracies—are overbuilt relative to their targeted or “safe” working loads. If differences in underworked—but not *unworking*—parts or arrangements are screened off as irrelevant because of functional similarity realized by a narrower set of parts, *Multiple Realizability* is denied at one level of description or explanation only by affirming it at another.

Second, the “no-replacements” clause in *Causal Relevance* above must be in place in order to test causal relevance, or else the counterfactual will come out false *because* of multiple realizability, as we saw in Section 3.3.3. It would be very strange if there was no multiple realizability because of multiple realizability. If differences in parts or arrangements are screened off as irrelevant because of functional similarity realized by micro parts, *Multiple Realizability* is once again denied at one level of description or explanation only by affirming it at another. So it is quite puzzling that Shapiro says,

For all I know, rigidity is a disposition that is multiply realizable. If rigidity is a disposition that can be brought about in various ways, and if steel and aluminum differ in respect to how they produce rigidity, then steel and aluminum are alternative realizations of rigidity. But, relative to their function in a corkscrew, rigidity screens off the differences between steel and aluminum (pp. 644-645).

We’ll return to this presupposition of *Multiple Realizability* at one level of explanation in order to deny it at another. And we’ll see that such a boundary may play an explanatory role. But, there is no “rigidity” as such

that could screen off these differences. If we sometimes speak this way, doing so posits an intervening boundary that is not a metaphysical boundary. Imagine reconstituting some aluminum oxide via electrolysis to produce some pure aluminum. Now, must one do any further building, in any metaphysical sense, to produce a lever or even a whole corkscrew? No; a specialized implementation of the Hall–Héroult process might have built the lever or corkscrew directly from aluminum oxide, and Hephaestus might have done so from aluminum atoms. It is pragmatic divisions of labour, rather than issues of metaphysical transitivity that restrict aluminum refineries to outputting stock material rather than corkscrews. Ontologically, the corkscrew and its rigid parts are on par. Shapiro provides no reason why realization isn’t transitive all the way in to the finest available grain-level of description.

While Polger and Shapiro take winged and waiter’s corkscrews to be multiple realizations of corkscrews, Polger and Shapiro follow Shapiro in viewing aluminum and steel differences as insufficient for multiple realizability. Likewise, while they think that camera eyes (like mammalian and octopus eyes) on the one hand, and compound eyes (like those of insects) on the other are multiple realizations of eyes, Polger and Shapiro think that mammalian and octopus eyes, despite their differing evolutionary histories and the resulting differences in the kinds of tissues that compose the parts of their eyes, are not relevantly different as eyes.⁷ Further, it seems plausible that Polger and Shapiro have causal relevance in mind insofar as they follow Shapiro in appealing to examples of irrelevant properties like colour:

Because [waiter’s corkscrews and winged corkscrews] make use of different mechanical principles—one levers, one rack and pinions, and so forth—we conclude that they do the corkscrew job in different ways, as considered from the point of view of our imaginary science of mechanical artifacts. In contrast, if we consider two waiter’s corkscrews that differ in the material of which they are made (e.g.,

⁷ Shapiro, and Polger and Shapiro, make a point of pitting Putnam’s analysis of the camera eye example against Block and Fodor’s (1972) analysis. I intend my analysis of explanatory versus causal relevance to help illuminate this interesting historical difference among functionalists. Putnam must have something like Polger and Shapiro’s relevance criterion in mind, and if so, will be subject to the same critique.

aluminum, stainless steel, or wood) or in their color (e.g., silver, black, or pink), the imagined science of mechanical artifacts tells us that these differences are not relevant differences (Shapiro 2000). The reason is that the same explanation of how the device does its corkscrew job applies to all of them, regardless of material composition and color. (Polger and Shapiro, 2016, p. 68. See also p. 69, p. 99)

Moreover, Polger and Shapiro take it that “realization is a synchronic ontological dependence relation, distinct from identity, and that transmits physical legitimacy from physical realizers to what is realized” (p. 28). We saw, however, that causal relevance doesn’t do the work Shapiro intended for it, and that as a species of dependence, it is not obvious how realization could fail to be transitive. Transitivity is arguably partly definitive of dependence relations. Either realization is transitive, or realization is not an ontological dependence relation. But Polger and Shapiro deny that realization is transitive.

On our view, eye is a multiply realized kind, but not because of variation in cellular composition. That is not the right kind of variation. Human eyes and octopus eyes are the same kind of realizer of eye, namely, camera eye. But camera eye might be multiply realized. . . . The differences between the cells in human and octopus eyes may well constitute a difference in the realization of part of an eye, or part of a part, etc. of an eye. But as it happens, it does not suffice for multiple realization of eyes themselves. And eyes do not inherit the multiple realization of their parts or realizers. Multiple realization doesn’t percolate up in this way. (p. 64)

They then attempt to explain the intransitivity in a footnote to the latter claim about percolation:

The reason is simple: Multiple realization is a claim about the dissimilarity of two taxonomies, and dissimilarity is not transitive. P might be realized by Q and Q might be multiply realized by R and S; but that does not imply that R and S are multiple realizers of P. One way this could happen is that the difference between R and S is relevant to kind Q and irrelevant to kind P. (note 1 on p. 81)

I grant that dissimilarity is not transitive.⁸ If indeed “multiple realization is a claim about the dissimilarity of two taxonomies,” I take the idea here to be that the taxonomy in which eyes are a kind is dissimilar to the taxonomy in which camera eyes are a kind, which in turn is dissimilar to the taxonomy in which cellular components like proteins are kinds. From these dissimilarities we cannot infer that the taxonomy in which eyes are a kind is dissimilar to the taxonomy in which cellular components are kinds. Suppose these taxonomies are somehow “similar.” I’m not sure what this is supposed to show. Consider an example of such a possibility. Suppose cars are multiply realizable by those with diesel and those with gasoline engines. Suppose that diesel vehicles are multiply realizable by trucks and cars. In other words, one way of classifying vehicles aligns with particular lower-level differences while ignoring others. Other ways of classifying vehicles align with other lower-level differences, while ignoring still other differences that are nonetheless causally relevant as counterfactually judged. Diesel cars and diesel trucks—both diesel vehicles that can have the same effect, say, getting to the office using vegetable oil—have differing frames and engine sizes; these differing components cannot be omitted, and they are part of how the individual does the diesel vehicle job. But this just makes it plain that no way of classifying vehicles unless by individual serial number can pay attention to every difference among ways of being a vehicle, and *that* is what *Multiple Realizability* claims amount to. Notice, too, that if kinds in two taxonomies are identical, and one is multiply realized by kinds in another taxonomy, both are multiply realized in virtue of being identical. Likewise, an identity in a synchronic ontological dependence chain that includes realization doesn’t break the transitivity of the dependence. We saw this above with gold. Atomic gold is any element-with-79-protons-in-its-nucleus. There is more than one “way”—not in the question-begging functional sense, but in the non-question-begging sense of differing compositional ways—to be an

⁸ For example, if A is dissimilar to Z and Z is dissimilar to B, one can no more conclude that A is dissimilar to B than that they are identical. Apples are dissimilar to zebras, zebras are dissimilar to bananas, but those dissimilarities are not a measure of the dissimilarity between apples and bananas. Unlike relations such as “bigger than,” and “depends on,” nothing about the relation between A and B can be known by their both being dissimilar to Z.

element-with-79-protons-in-its-nucleus. So, there is more than one way to be atomic gold. Likewise, if A (a camera eye) is B (a lens focusing light on a photosensitive surface), and B depends on C (cellular composition of the lens), then being identical, A and B must stand in the same relation to C , and A also depends on C . And so with eyes generally. If Y (an eye) depends on A , and A is B , and B depends on C , then Y depends on C . Nothing about the intransitivity of dissimilarity bears on the transitivity of realization as a dependence relation. When Polger and Shapiro claim that *Multiple Realizability* “is a claim about the dissimilarity of two taxonomies” (note 1 on p. 81), they might mean that realization, as they define it, is a special form of dependence an individual of one kind has on one of a number of individuals of another kind, such that the former kind is supposedly not reducible to the latter. But the putative irreducibility of one taxonomy to another is a supposed *further consequence* of multiple realizability; it is not, in the first place, what *Multiple Realizability* is “about.” The invocation of *Multiple Realizability*, if realization is a species of dependence, first of all involves dissimilarities other than taxonomic ones, namely, causally relevant physical differences between realizers. It also involves sameness, that is, sameness of function. It is simply off track to cite *putative* intertaxonomic irreducibility and the intransitivity of dissimilarity as bearing on the question of the transitivity of any proposed ontological dependence relation.

Of course, it might be that Polger and Shapiro’s view is unclear as to the notion or notions of relevance at work. Besides following and citing Shapiro’s example of colour and the claim that realization is a synchronic ontological dependence relation, it is not clear that the relevance criterion Polger and Shapiro have in mind is a causal one. For one thing, whereas Shapiro named his relevance criterion “causal” relevance, Polger and Shapiro do not name their relevance criterion or criteria.

What might go some way to explaining the thought that there is an interesting difference between winged and waiter’s corkscrews, but not between aluminum and steel corkscrews, is the fact that a lever, for example, performs a sub-function which can be taken as a sub-explanandum, and has as its dependence base a subset of the whole corkscrew. Suppose there can be a more

narrow range of mechanisms, like winged corkscrews, that can perform a more specific function or functions. If the role of a lever as a part of this system can be occupied by levers differing in causally relevant ways, then the idea might be that the *lever*, not the corkscrew, is multiply realizable. This seems to be Polger and Shapiro's view in the case of camera eyes as realizations of eyes. Camera eyes are realized by "lenses that focus light on a concave retina that is located at the back of the eye" (p. 51). They say,

Vision scientists regard human and octopus eyes as the same in type despite differences in their cellular composition. Cellular differences do not make a difference to how the eyes perform their function: They both focus light on a photosensitive surface in the same way. They rely on the same kind of mechanism. (p. 64)

Again, while camera eyes may be the same kind of mechanism, it is not true that camera eyes differing in cellular compositions focus light on a photosensitive surface in a way that is in every sense the same. But what I want to highlight now is Polger and Shapiro's focus on what actual sciences say:

the sciences that deal with the putative realizers tell us which differences in them amount to different ways of producing their effects and which do not. . . . As before, the relevant differences are those that make for differences in kind according to that science. (p. 66)

This suggests that the notion of relevance that Polger and Shapiro are applying to screen off compositional differences as "mere variation" (p. 209) is what we might call "explanatory" relevance. For example, we might think that we were doing a better job of explaining the situation with eyes by refraining from saying that differing cellular compositions amount to different ways of performing the eye function. Or, at least, we might think we'd do a better job by being more specific. As an explanation, the fact that a corkscrew depends on some aluminum sheds no light on, perhaps even obscures, the arguably more interesting variety of ways that the very same material might be arranged in various mechanisms able to perform the same function. On the other hand, if aluminum and steel were more rarely found doing the same job except in corkscrews, and if the range of mechanisms that could extract corks

was narrower, it is conceivable that reference to materials at a finer grain-level would be doing a greater share of the explanatory work, epistemically speaking. Imagine a complex lock, for example, employing a retinal scanner, as a case of a functional role that is more discriminating of occupants of the key role. The explanation of how the retinal key works might well involve quite fine-grained descriptions of blood vessels, particularly in the case that a spy is infiltrating a high-security facility by impersonating a chief of defense. Recall, Polger and Shapiro say that the differences between aluminum and steel are irrelevant because “the same explanation of how the device does its corkscrew job applies” (p. 68) to all corkscrews of the same mechanical type with differing material types. But there is simply more than one “explanation,” available in the sense at issue. If one asks how a corkscrew works, or how a bottle was opened, the best answer depends on who is asking, why, when, where, and how. What’s difficult to reconcile is that Polger and Shapiro acknowledge this very point. Following Sober (1999), they reject, as I do, “Putnam’s claim that the microphysical explanation of why the square peg will not go through the round hole is not an explanation at all, or is at least a terrible explanation” (p. 200). Rather, they think, “Both microphysical and macrophysical explanations can be good ones,” and “this is what should be said about cognitive and neuroscientific explanations, as well.”

There may of course be various reasons to prefer one over the other:
but the reasons will not be that one “explanation” fails to be an
explanation at all. (p. 200)

As Bennett puts it, the sort of explanation that “renders [some explanandum] intelligible, sheds some light on how or why it happened, or perhaps puts an end to a line of questioning” (p. 61), “is surely context sensitive in that it matters what exactly we are asking, and what we already know; it is also subject to Gricean considerations about cooperativeness and so forth” (p. 61). The appearance of a sort of variability that is *prima facie* more interesting than very small differences (“mere variation”), is a feature of cases like that of corkscrews, in which the straight-forward constraints imposed by corked bottles make for somewhat more (but not completely) open-ended engineering possibilities. Of course, Polger and Shapiro are right that the job of the camera

eye imposes constraints that make for less variation—less multiplicity—in camera eyes than in eyes generally. But again, this is a matter of degree, and less variation needn't be less interesting; for example, the apparent convergent evolution in the case of human and octopus eyes may be interesting in its own right. So we ought not to decide what is “interesting” in advance. As it happens, it is fairly common knowledge that aluminum and steel can perform somewhat similarly in spoons and in bicycle frames. That, I take it, is old news to some, and exciting to others. Suppose a culture of woodcarvers has managed to make a great variety of things from wood. They are so adept at carving and so mechanically ingenious that they do not find the differences between wooden dovetail joinery and wooden gears interesting. But they are novice genetic engineers interested in improving wood quality. Indeed, they are even more intrigued when an alien arrives with corkscrews made of carbon fiber. Closer to home, the case of the corkscrew might furnish an opportunity to explain to an eight-year-old that the two levers and rack-and-pinions might have been exchanged for a single lever, and it also might furnish an opportunity to explain that the aluminum might have been exchanged for steel, this being a case where those differences make no difference (by replacement) to the general function, and yet, that such differences are indeed difference-makers elsewhere and in other circumstances. Aluminum-single-lever corkscrews are preferable when portability is at issue; double-lever-corkscrews are preferable (or not) for their ease-of-use by eight-year-olds. By contrast, a “complete” explanation, or “metaphysical” explanation, surely ought to mention the materials, since these are what make up the items and ultimately make them able to do their jobs. One couldn't simply go without those materials. In the case of eyes again, it may be that the fact that certain proteins compose a lens does not “explain” why the eye of which the lens is a part counts as an eye. And “explains”—understood epistemically—may not be transitive. However, I need not settle that. None of this shows that realization is not transitive. If realization is a synchronic ontological dependence relation, and if A realizes B , and B is or realizes C , then A realizes C . And since many differences satisfy the plausible *Causal Relevance*, multiple realizability is as common as “mere” variation. In fact, we have no reason, so far, to see these as anything other than rough

differences in degree of similarity.

5.3 The Second Horn and Similarity

That brings us to the second horn of the dilemma, which was that, “If the realizing kinds do genuinely differ in their causally relevant properties, then, it seems, they are different kinds” (p. 647). To clarify, I take it that Shapiro means that if *realizers* differ in their causally relevant properties, then they are different scientific kinds. Shapiro sketches an argument for that premise. First, he advocates an inheritance principle for laws:

Microlevel laws determine the macrolevel laws, in the sense that their truth physically guarantees the truth of the macrolevel laws. Macrolevel laws about water are true because certain microlevel laws about water are true. But, more significantly, macrolevel laws about all samples of water are possible because the same microlevel laws are true of all samples of water. (p. 649)

In order to generate an inference to Shapiro’s contention that, “if two realizations contribute to a capacity in causally distinct ways, then this must mean that there are no laws common to both of them” (p. 648), the inheritance must be taken together with another plausible, albeit unstated, principle of inheritance of differences. He must think that realizers that differ in their causally relevant properties do so in virtue of differences in their lower-level compositions or constitutions. Next, quoting and agreeing with Fodor in one respect, Shapiro (p. 647) endorses a projectibility criterion for natural kinds:

What justifies a taxonomy, what makes a kind “natural,” is the power and generality of the theories that we are enabled to formulate when we taxonomize in that way. (Fodor, 1968, p. 119)

The reconstruction can now be fleshed out as follows:

- (1) Microlevel laws determine the macrolevel laws.
- (2) Realizers that differ in their causally relevant properties do so in virtue of differences in their lower-level compositions or constitutions.
- (3) Realizers that differ completely in their causally relevant properties have no laws common to both of them.

- (4) What makes a kind natural is the explanatory power of generalizations in which the kind figures.

So:

- (5) Realizers that differ *completely* in their causally relevant properties are different natural kinds.

But as we're about to see with Kim's (1992b)⁹ earlier and more thorough argument to the same effect, the completeness of difference required for the inference to be valid is a problem. Whereas Shapiro focuses on the first horn of the dilemma he presents, which we saw above, Kim is aware that the second horn is crucial for a general refutation of the *Antireduction Thesis*, and he thus focuses almost entirely on offering an argument for the disastrous effect of the second horn to the *Multiple Realizability* claim, and the *Antireduction Thesis* with it. The second part of the dilemma is crucial, because if any of the objects of a special science *are* multiply realizable, which no parties to the debate deny,¹⁰ the promise of denying *Multiple Realizability* in *general* rests with those "kinds" falling on the second horn of the dilemma, and not being genuine scientific kinds. Notice, too, that the essential "condition on the nature of [cognitive] processing" (Adams and Aizawa, 2001, p. 51) we saw in Section 2.1 is reminiscent of this latter strategy. Recall that a corollary of a notion of kinds according to which heterogeneous collections are unscientific is that EM is trivialized to the extent that it relies on the similarity of classically cognitive processes and mental states with extended ones. The objection is that the similarity is too weak to support inductive reasoning or prediction.

⁹ Every reference to Kim in this chapter and the next is to Jaegwon Kim (1992b), unless otherwise specified.

¹⁰ No one, to my knowledge, wants to go so far as to claim that no objects of any special science are multiply realizable, even if they argue that multiple realizability is not nearly so widespread as antireductionists have assumed. Shapiro (2000) and Polger and Shapiro's (2016) favourite illustration is that of corkscrews, and they think simple waiters' corkscrews employing a single lever, and corkscrews employing a rack-and-pinion actuated by double-levers, are legitimately multiple realizations of the corkscrew kind. So, too, with the compound eyes of insects as compared to the camera eyes of mammals—these are truly multiple ways of realizing the eye kind.

Kim opens by saying,

I will not here dispute. . . the possibility that extraterrestrial creatures with a biochemistry radically different from the earthlings', or even electromechanical devices, can "realize the same psychology" that characterizes humans. (p. 1)

Kim thinks this possibility entails the standard conclusion that multiply realizable special science kinds or properties are irreducible and that special sciences are thus autonomous. He then proceeds to argue that any such multiply realizable and irreducible "kinds" can only be disjunctive sets whose heterogeneity precludes their ability to figure in empirical laws. Witmer (2003) summarizes Kim's argument nicely:

Jaegwon Kim claims that [the argument for the *Antireduction Thesis*] proves too much: if we accept its premises, we will block psychophysical reduction only at the cost of being forced to accept the further, disturbing conclusion that mental properties themselves cannot appear in laws and are not genuine "kinds" in the first place. Whatever is wrong with the property expressed by the disjunctive predicate *infects*, so to speak, the mental property itself. This is Kim's challenge: if there is something about the "disjunctive" property that prevents it from appearing in laws of nature, how can mental properties themselves appear in laws of nature? (p. 60)

As Witmer (2003) also points out,

Kim's [1992b] challenge. . . arises for anyone who is inclined to accept both that there are psychological laws and that mental properties are multiply realized. If (like myself) you think the original antireductionist argument is fundamentally misguided, you may still need to find an answer to Kim's challenge. (p. 60)

In other words, the reductive functionalism I advocate requires me to answer Kim's challenge. Kim's argument proceeds from what Kim calls the "Physical Realization Thesis," which Kim thinks physicalist proponents of realization already accept, and which he describes as the conjunction of the "Structure-Restricted Correlation Thesis"—

If anything has mental property M at time t , there is some physical structure type T and physical property P such that it is a system of type T at t and has P at t , and it holds as a matter of law that all systems of type T have M at a time just in case they have P at the time (p. 5)

—with the belief that,

significant properties of mental states, in particular nomic relationships amongst them, are due to, and explainable in terms of, the properties and causal-nomic connections among their physical “substrates.” (p. 14)

To the physical realization thesis, Kim adds two more plausible principles his interlocutors do not typically deny:

[**Causal Inheritance**] The causal powers of an instance of a property realized by a physical implementation are identical with the causal powers of that implementation. (p. 18)

[**Causal Individuation of Kinds**] Kinds in science are individuated on the basis having similar causal powers. (p. 17)

From these principles we can conclude:

[**Relative Homogeneity of Kinds**] Properties realized by completely dissimilar physical implementations cannot be constitutive of scientific kinds.

This conclusion, together with a Physical Realization Thesis about mental kinds, or any functional kind, “effectively rules out mental kinds as scientific kinds” (p. 18). I call this a form of denying *Multiple Realizability*—the claim that many special science kinds are multiply realizable—because Kim’s argument denies that what is multiply realized is a kind at all. Then, insofar as the autonomy of a special science is secured by the irreducibility of its objects of study, such autonomy is bought at the cost of the projectibility and explanatory power of the generalizations made by that science, that is, at the cost of being a science at all.

But Kim’s argument appears to be susceptible to obvious and widespread counterexamples. Thus, for example, Fodor (1997) replies,

The very existence of the special sciences testifies to reliable macrolevel regularities that are realized by [heterogeneous] mechanisms. . . . Unimaginably complicated to-ings and fro-ings of bits and pieces at the extreme microlevel manage somehow to converge on stable macrolevel properties. (p. 160)

Fodor continues,

On the other hand, the “somehow” really is entirely mysterious, and my guess is that that is what is bugging Kim. He just doesn’t see why there should be (how there could be) macrolevel regularities at all in a world where, by common consent, macrolevel stabilities have to supervene on a buzzing, blooming confusion of microlevel interactions. (p. 161)

Note that Shapiro, Kim and Fodor, among others, seem to agree that multiplicity, regularity, and reduction don’t readily mix (although this is exactly what I’ll go on to deny in Section 6.3). Shapiro and Kim want to downplay multiplicity, and keep the other two. Fodor wants to keep multiplicity and regularity and dispense with reduction. So Fodor concludes, “Kim wants to make it true by fiat that the only projectible kinds are physically homogeneous ones” (p. 161).

But Fodor’s appeal to mystery is hardly preferable to Kim’s supposed untenable fiat. Furthermore, there is no such fiat. Kim’s argument only requires “similarity” for physical implementations of kinds, not “homogeneity.” However, *similarity* is the real source of the problem with Kim’s argument. Notice that I reconstructed Kim’s argument for the *Relative Homogeneity of Kinds* with reference to “complete dissimilarity,” whatever that might be. Since Kim says no such thing, it is perhaps more true to his view to omit the completeness. But without completeness of dissimilarity, the argument simply isn’t valid. In a crucial passage, Kim offers invalid reasoning as to why the two principles rule out mental kinds as scientific kinds:

The reasoning is simple: instances of *M* that are realized by the same physical base must be grouped under one kind, since *ex hypothesi* the physical base is a causal kind; and instances of *M* with different realization bases must be grouped under distinct kinds, since, again *ex hypothesi*, these realization bases are distinct

as causal kinds. Given that mental kinds are realized by diverse physical causal kinds, therefore, it follows that mental kinds are not causal kinds, and hence are disqualified as proper scientific kinds. Each mental kind is sundered into as many kinds as there are physical realization bases for it, and [psychology] as a science with disciplinary unity turns out to be an impossible project. (p 18)

It does not follow that mental kinds are not causal kinds unless “diverse” is understood as involving no similarity whatsoever. The argument involving complete dissimilarity and the valid reconstruction above is thus not entirely unwarranted. That’s because the notion that multiple realizability involves complete dissimilarity is invited by the strong language antireductionists have historically used in their appeals. Indication that the target antireductionist thesis or principle involves complete dissimilarity is given in the quotations Shapiro elects to cite from his primary disputants, especially Fodor (1968). Fodor says that, “because it is (roughly) a sufficient condition for being a mousetrap that a mechanism be customarily used in a certain way, there is nothing in principle that requires that a pair of mousetraps have *any* shared mechanical properties” (pp. 115-116, my emphasis). And Kim says,

the functionalist, especially one who believes in MR, would not, and should not, look for something common to all pains over and above *H* (the heart of functionalism, one might say, is the belief that mental states have no “intrinsic essence”). (p. 23)

It might look as though the more dissimilar the implementations, the better the antireductionist’s case. Putnam (1979) is famous for claiming that “whatever our substance may be, soul-stuff, or matter or Swiss cheese, it is not going to place any interesting first order restrictions on the answer” to the question: “*What is our intellectual form?*” (p. 302, emphasis in original). But Kim’s good argument against that sort of multiple realizability of scientific kinds stands: complete difference among functional kinds would indeed preclude their figuring as scientific kinds. So, despite rhetorical appeals to Swiss cheese and a lack of interesting physical similarities, the *complete* dissimilarity or difference that Kim’s and part of Shapiro’s argument target is not really a live option—even if only in light of those arguments. Recognizing this, Polger and

Shapiro say, “We know that realization theorists and functionalists made such claims, but we think that they should not have done so” (p. 38). They cite Block (1997), an antireductionist, who distances himself from this extreme by appealing to the “Disney” principle—that teacups can’t really walk and talk because, “constraints impose similarities” (p. 121). Multi-functional specificity imposes further constraints, for example, among portable corkscrews, which have to be relatively light, or among orally-administered antibiotics, which have to be absorbable by the digestive tract. In the end, Kim’s argument from causal inheritance and causal individuation of kinds is effective only against an extreme position, even if that straw man is one of their opponents’ making. Kim’s arguments can’t guarantee that there are no regularities among functional kinds because anyone granting that kinds are individuated on the basis of their causal powers must allow that even very diverse mechanisms underlying a function *are just that similar*.

Now recall the dilemma Shapiro offers against *Multiple Realizability*. The general thought is that by the time realizers differ enough to count as multiple realizations, they differ too much to be projectible and to count as members of a scientific kind. From this Shapiro attempts to conclude that functional kinds “never enter into (nonanalytic) laws of the form ‘All *Ps* are also _____’ ” (p. 654). But this conclusion would also require that the claim, “if two realizations contribute to a capacity in causally distinct ways, then this must mean that there are no laws common to both of them” (p. 648), be clarified in terms of complete distinctness. And the claim that there are no laws about functional kinds is notably absent from Polger and Shapiro, anyway.¹¹ Despite the multiple realizability of corkscrews, we can indeed specify regularities like, *cork extractors apply upward force* $> x$ *to a cork*,¹² where x is the higher

¹¹ Actually, this isn’t totally clear. Polger and Shapiro say, “the reality of corkscrews is not redeemed by the regularity that corkscrews tend to have the capacity to remove corks from bottles, for this regularity just repeats the functional specification of corkscrews” (p. 196), and this is said in the context of the requirement that real properties be invoked in regularities that are “empirical rather than analytic,” implying that corkscrew regularities are analytic. By contrast, the regularity I gave is clearly empirical. And corkscrews are indeed multiply realizable on their view.

¹² I choose the more general functional kind “cork extractors,” to show that quite a broad

limit of the range of frictional forces exerted by typical corks compressed into typical bottlenecks. The range of rigidities and tensile/flexural, shear, and compressive strengths among corkscrew parts can thus be predicted by those of the other parts given x , as well as the properties of the compressed corks to be penetrated, and the typical force-generating capacities of the actuating organism. Now, if that's not a very clear example of a useful projection, think of something like an antibiotic as a member of a functional kind that we might make generalizations about. Perhaps, *antibiotics bind with catalytic enzymes essential to cell wall biosynthesis*. Notice, crucially, that these generalizations involve facts about bottles and parasitic bacteria respectively. For the time in the history of medicine that the generalization about antibiotics is or was a true generalization,¹³ it is or was true in virtue of the particular enzymes essential to bacteria. Indeed, Shapiro also seems to acknowledge that if realizers “have no or only few causally relevant properties in common, then there are no *or just a few* laws that are true of both of them” (p. 649, my emphasis).

So the counter dilemma for Shapiro's dilemma is this. On one hand, without completeness, the argument yields no strong conclusion about functional kinds' utter inability to enter into nonanalytic generalizations. Although the

generalization covers an even greater variety of mechanisms than “corkscrew.” Polger and Shapiro take “corkscrew” to be an “anchored kind” that makes use of an “identity” (p. 39), because it specifies the use of a screw, which they take to be a kind in the “realizing mechanical science.” “Identity,” in this sense, is supposed to hold between two classes which admit a degree of variation, or multiple realizability in my, and Gillett and Aizawa's sense, among their members. I'm unclear as to what the anchoredness is supposed to show, because we can always add more specificity to pick out a subclass of functional kinds. “Aluminum winged corkscrew” is one such subclass, and might be more suited to certain applications. The location of a spark plug to be loosened or tightened might specify not just a 21mm socket, but must be reached by a 21mm-socket-on-a-long-drive-with-a-u-joint-in-the-middle. If kinds and individuals are causally individuated, it will be possible to specify a set of effects more and more narrowly until just one individual satisfies the criteria, for example, when retinas or fingerprints are used as keys in high-security locks.

¹³ This is indeed the way β -lactam antibiotics, the largest family of antibiotics, including penicillin, works (Blair, Webber, Baylay, Ogbolu, and Piddock, 2014). Since penicillin was the first pharmaceutical antibiotic to be developed, there was a time in history when the above generalization was true.

prevalence of multiple realizability might be questioned, in that case, a causal relevance criterion for differences must do all the work. But according to a non-question-begging causal relevance criterion clarified in terms of counterfactual dependence, minor variation among realizers is not ruled out as causally relevant difference. And explanatory relevance understood epistemically is not relevant to issues of synchronic ontological dependence. On the other hand, with completeness, the argument does not target any live claim of *Multiple Realizability*. The valid conclusion involving complete dissimilarity doesn't cover functional kinds in its scope, and leaves a foothold for the antireductionist in the standoff we've seen.

Wrapping Up

To summarize, the second horn of the apparent dilemma for functional kinds works with completeness of difference among realizers, but doesn't target any live view. And since complete difference presupposes causally relevant differences among realizers, the first horn of the apparent dilemma is only pressed into service *without* imagining complete differences among realizers. What look like horns of a dilemma don't actually work together. Rather, we've seen two distinct arguments against *Multiple Realizability*: one is a failed argument against the prevalence of multiple realizability, and one is an effective argument against the scientific seriousness of merely imagined functional kinds like minds that could be made of Swiss cheese.

So far, antireductionism still has its *Multiple Realizability* to appeal to, and still carries the day. Even supposing multiple realizability is less prevalent than some of us think, it is not clear what such a conclusion would do to advance the larger debate. Antireductionism still goes through for those less prevalent functional kinds, and the puzzles to which antireductive physicalism leads are left doubly unilluminated. First, because the restricted notion of multiple realizability still grants the irreducibility of multiply realizable kinds such that the same puzzles will arise in those cases. Second, because the reducibility of kinds admitting of variation remains a starting assumption of the identity theorist, and in the absence of a justification of that assumption,

the antireductionist's argument can thus be run in terms of variation. The attempt to distinguish between variation and multiple realizability just focuses the debate in the wrong place. Instead, as we're about to see in the next chapter, by justifying the reducibility of kinds admitting of variation, and running together variation and multiple realizability on a spectrum of similarity, reductionists can deny a different premise of the antireductionist's argument. Hence, they can dispense with antireductionism while keeping functional kinds, which, as I said at the outset of the thesis and of this chapter, is what would be required for our identity with the things we are constituted by, for example, technologies.

Chapter 6

Similarity and Reduction

In the last chapter, I argued that it doesn't work to counter antireductionism by denying either the prevalence or the scientific legitimacy of functional kinds. First, I argued in Section 5.2 that causal relevance doesn't distinguish between "mere variation" and multiple realizability. Causal relevance is no reason to think that multiple realizability is not ubiquitous, and that way of countering antireductionism fails. I then considered Kim's (1992b) argument against antireductionism premised on the idea that, "Kinds in science are individuated on the basis of having similar causal powers" (p. 17). But unless the dissimilarity between members of a functional kind is somehow complete, the idea yields no strong conclusion about functional kinds' utter inability to be projectible. On the other hand, with completeness, a valid argument involving complete dissimilarity doesn't cover functional kinds in its scope, and leaves *Multiple Realizability* intact as a foothold for antireductionism.

In this chapter, I offer a positive argument for the idea that variation between members of a kind and multiple realizability of a kind involve the same measure of sameness and difference. In Section 6.1, I offer a causal account of similarity according to which kindhood and projectibility come in degrees. I concluded the last chapter by noting that the attempt to distinguish between variation and multiple realizability, even if it was successful, wouldn't advance the debate between antireductionists and reductionists. In Section 6.2, I explain why that is, and hence why it is important, as groundwork for reductive functionalism as a plausible form of physicalism, that variation and multiple realizability be understood as ends on a spectrum of similarity. In Section 6.3, I argue

that this view of kinds, even while granting the *Multiple Realizability* claim—that many special science kinds are multiply realizable by distinct physical kinds—allows us to deny the *Unity Condition on Kind Identity*. That was the further premise: if a kind is multiply realizable by distinct physical kinds, then it cannot be identical to any specific physical kind. We can hence reject the *Non-identity Thesis* that would follow—that many special science kinds are not identical to any specific physical kind. Having rejected the latter, even while accepting a plausible condition on reduction such as the *Identity Condition on Reduction*—that if a kind is not identical to a physical kind, then it is not reducible to physics—we can reject the *Antireduction Thesis*. That was the conclusion that many special science kinds are not reducible to physics.

6.1 Accounting for Similarity

A crucial claim in Kim’s argument was that “Kinds in science are individuated on the basis of having similar causal powers” (p. 17). While I agree, this says little about how to understand kinds and similarity. It is circular to define kinds in terms of similarity if similarity is defined in terms of kinds, which it is. At least since Leibniz, “the standard way of cashing out what it means for an object to be similar to another object is to require that they co-instantiate properties” (Frigg and Nguyen, 2016, sec. 3.2). “Famously, Leibniz construed identity as the sharing of all properties and, analogously, similarity as the sharing of at least one property” (Decock and Douven, 2011, p. 68). Goodman (1972) criticized the very notion of similarity as “slippery.” “As it occurs in philosophy, similarity tends under analysis either to vanish entirely or to require for its explanation just what it purports to explain” (p. 446). Contra Leibniz, this prevents us from meaningfully defining similarity in terms of shared properties, an issue now well-known (e.g. Lewis, 1983, p. 346; Armstrong, 1989, p. 40; Gärdenfors, 2004, p. 111; Maclaurin and Sterelny, 2008, p. 15), and which remains controversial (see Yi, 2018; Weisberg, 2012). Goodman considers the possibility that “*a* and *b* are more alike than *c* and *d* if the cumulative importance of the properties shared by *a* and *b* is greater than that of the properties shared by *c* and *d*” (p. 443). But Goodman leaves that direction

unexplored, rejecting importance as relative—“a highly volatile matter, varying with every shift of context and interest, and quite incapable of supporting the fixed distinctions that philosophers so often seek to rest upon it” (p. 444). For example, Goodman says,

suppose we have three glasses, the first two filled with colorless liquid, the third with a bright red liquid. I might be likely to say the first two are more like each other than either is like the third. But it happens that the first glass is filled with water and the third with water colored by a drop of vegetable dye, while the second is filled with hydrochloric acid—and I am thirsty. Circumstances alter similarities. (p. 445)

Goodman’s point here is that if similarity judgements are made on the basis of properties, the relative importance of properties to those judgements varies with context. Indeed, I think we must say something of the same form if we take similarity as primitive and try to explain property judgements on the basis of similarity.¹ Yet, the possibility of making judgements at all, and making more or less fruitful ones, must depend on the world being the way

¹ Similarity *judgements* have received significant attention in the cognitive psychological sciences and philosophy of psychology, following Carnap (1967), who took similarity as primitive, and attempted to explain property attribution in terms of similarity. Contra Carnap, Goodman (2012) introduced two problems. The problem of “imperfect community” is a problem for defining kinds or properties by saying that a kind is a set whose members “are more similar to one another than they all are to any one thing outside the set” (Quine, 1969, p. 9). This definition allows for strange disjunctions which are intuitively not kinds. The problem of “constant companionship” is the problem of how unfavourable circumstances might lead us to glean the wrong properties from the world, and to miss others altogether, in ways which we ostensibly do not. The details do not matter here, and I simply note that later elaborations seek to overcome difficulties with Carnap’s account (see Tversky, 1977; Gärdenfors, 2004, 2014). These may be good initial models of how we make property attributions, but since they take similarity as primitive, they are silent on the question—if it is a question—of what it is for entities to be more or less similar *independently* of our judgements such that judgements are possible and more or less fruitful. Alternatively, if similarity is defined, following Leibniz, in terms of the sharing of properties, we might construct analogous models of human or animal similarity attribution in terms of properties, and yet these models, taking properties as primitive, would be silent on the question of what it is for entities to share in a property independently of our judgements. We can expect the same issues to arise for kinds or natural resemblance classes. And this is the problem—in

that it is somewhat independently of our judgments. As Quine (1969) says, the “dubious scientific standing of a general notion of similarity, or of kind,” “is itself a remarkable fact. For surely there is nothing more basic to thought and language than our sense of similarity; our sorting of things into kinds” (p. 6). Not unlike Goodman’s “importance” criterion, Quine (1969) thinks we define similarity only by being more restrictive about the classes that constitute properties. He appeals to the “significance” of similarity, saying,

If properties are to support [Leibniz’s] line of definition [of similarity] where sets do not, it must be because properties do not, like sets, take things in every random combination. It must be that properties are shared only by things that are significantly similar. But properties in such a sense are no clearer than kinds. To start with such a notion of property, and define similarity on that basis, is no better than accepting similarity as undefined. (p. 7)

Quine (1969) ultimately takes similarity to be more easily defined in terms of kinds than vice versa. Quine gives “a somewhat limping definition” that “works only for finite systems of kinds”:

a is more similar to *b* than to *c* when *a* and *b* belong jointly to more kinds than *a* and *c* do. (p. 8)

For my part, I’m unsure whether the restriction to finite systems is a problem. In any case, even if “the relation between similarity and kind... is less clear and neat than could be wished,” Quine thinks the “notions are in an important sense correlative” (p. 10). So, too, with properties.

While Quine perhaps took kinds, similarity and properties to be “unreduced,”² the point is that no term we are seeking to define can figure in the definition. We cannot meaningfully define similarities in terms of “similar”

whichever guise—of concern in the main text. I simply note that most philosophers seem to have continued to follow Leibniz in thinking that similarity is more easily defined in terms of properties or kinds than vice versa. So Quine says, “Definition of similarity in terms of kind is halting, and definition of kind in terms of similarity is unknown. Still the two notions are in an important sense correlative” (1969, p. 10).

² If it’s right that Quine thought kinds are ultimately unreduced, I don’t think it is entirely clear why. In answer to the question, “Why should our subjective spacing of qualities have a special purchase on nature and a lien on the future?” (p. 13), Quine suggests,

effects. We cannot define kinds in terms of having the same *kinds* of effects. So, we cannot meaningfully—that is, non-trivially—say with Kim that “kinds in science are individuated on the basis having similar causal powers” (p. 17). That is true, but trivially so, given that similarity and kinds are correlative or interdefined. For example, if similarity is understood in the way Quine suggests, in terms of degree of joint kindhood, then Kim’s premise only says that kinds have the same kinds of effects.

Yet, I take it that causal individuation of entities and properties still shows promise. “Alexander’s Dictum,” named by Kim (1992a, p. 134) after a passage in Alexander (1927), is the idea that what exists is causally efficacious. Kim (1992a) argues that his functionalist interlocutors are committed to it as well as he. Although Kim’s name for the principle is now popular, the same idea is present in Plato’s *Sophist* (247d-247e). Elder (2003) says that the status of Alexander’s Dictum “as a regulative principle seems unassailable,” and argues that it is “as bland and as true as it appears to be...” (p. 22). Add to it the widely accepted notion of identity as the sharing of all properties noted in Section 3.1.5. Taken together, these entail the “indiscernibility of identicals,” and the discernibility of non-identicals. Properties are thus plausibly identified with causal powers, and particular entities can be identified by their whole particular sets of effects.

So we might start with the thought that to have a particular property is to have a particular effect. But suppose that infinite shades of red are infinitely many properties, since entities with those properties have correspondingly many effects on, for example, an analog spectrometer. If those effects individuate those properties, *ex hypothesi*, there is no single effect that can bring all these shades of red under a single property or kind. As Quine (1969) puts it, “Cosmically, colors would not qualify as kinds” (p. 14). Of course, it seems intuitive that these effects of shades of red would be similar, and the world

There is some encouragement in Darwin. If people’s innate spacing of qualities is a gene-linked trait, then the spacing that has made for the most successful inductions will have tended to predominate through natural selection.

This, and a comment that Quine makes about more methodologically sophisticated systems of kinds and similarity standards—which we’ll see shortly in the main text—suggest to me that kinds are further analyzable.

is full of examples bearing out this intuition. An elephant and a truck are similarly massive compared to a proton, the former two having “similar effects” on a weigh scale (on Earth). The more it rains, the more a mountain erodes. The greater the solar incidence and absorbance of some material, the greater the mean kinetic energy of that material when the sun shines. Again, however, it is no analysis of kinds that appeals to the similarity of their effects. We’ve already seen that similarity of effects just assumes what it is supposed to explain. Causal powers, varying as finely as they often do, will not yield kinds if we take there to be one effect per causal power or property, because such effects would individuate properties too finely.

However, this just means we should not think that there is one effect per property or causal power. After all, why should the causal power of reflecting or emitting a particular wavelength of light be individuated by a spectrometer and not something else—an eye, a piece of black-and-white-film, or a molecule of chlorophyll b? It would seem that there can be many effects of the same causal power of reflecting a particular wavelength of light: on eyes, on film, on chlorophyll b. Moreover, I suggest that the relation between causal powers and effects is many–many, and that if causal individuation is going to do the required work in individuating kinds, what we need is a notion of causal individuation in which the *same* effect or effects figure.

Indeed, the world is not only a collection of weigh scales and spectrometers. Nature is also full of what I’ll call “switches”. Let me explain. Not only can there be differing effects of the same causal power, differing causal powers can have some of the same effects. That there is a planet hospitable to life orbiting our sun, if the nebular hypothesis is right, is to be explained by the distribution of matter and the resulting balance of forces—stellar wind, gravitational accretion—operating in the protoplanetary disk surrounding our sun as a young star. Mass scales smoothly, but some *critical* mass makes the difference between a planet’s present being and non-being. Chlorophyll b is excited to precisely the same degree by up to four quite different wavelengths of light. Graphing absorbance on the *y*-axis and wavelength on the *x*-axis yields a curve for chlorophyll b that resembles a double-peaked mountain, albeit truncated at a threshold for *x*. There are thus four values for *x* that satisfy each

of many values for y . In other words, four distinct wavelengths of light—four otherwise distinguishable causal powers—can elicit a single effect as far as photosynthesis is concerned. I think a hint of this idea of switches can be seen in Quine (1969). In contrast to our “innate, scientifically unjustified similarity standards” (p. 14), Quine describes,

a development away from the immediate, subjective, animal sense of similarity to the remoter objectivity of a similarity determined by scientific hypotheses and posits and constructs. Things are similar in the later or theoretical sense to the degree that they are interchangeable parts of the cosmic machine revealed by science.
(p. 19)

I take the notion of interchangeability of parts to be very close to the idea of differences that make no difference to function, but are nonetheless discernible differences among things required for the function. Consider, again, the interchangeability of wavelengths for a photosynthetic effect. The light is required, and the wavelengths differ. Outside the photosynthetic effect, among other effects, the difference is detectable. For example, the same values are not interchangeable as far as human vision is concerned. It seems characteristic of organisms, in particular, to employ even apparently analog devices at sensory surfaces in further state-of-affairs switching, allowing similar entities to have, quite literally, the same effect. For example, a scale or balance can be used as a sensor to discriminate among inputs, and in the simplest case, it can be used as a switch that distinguishes between items in a binary way. It might seem as though the input from an analog device should be kept raw to best discriminate reality. But this, again, would have cognitive systems operating only with particulars and without kinds, precluding induction. If comic strip character Calvin uses a bathroom scale, recessed into the floorboards and hidden under a rug, as part of a booby trap in the doorway to his room, he can tune various functions on his (imaginary) “transmogrifying” gun to trigger for various weight ranges, and thus discriminate victims by their weights. By selecting trespassers under 20kg, he can transmogrify his neighbour Susie into a frog. By selecting trespassers between 20 and 60kg, he can transmogrify his Mom into a “scarecrow,” and so on for Dad and Mrs. Wormwood. Of course Hobbes (usually Calvin’s ally) might fall prey to the prank intended for Susie,

and Calvin will have to add another setting to avoid the technical difficulties associated with de-transmogrification and Hobbes' wrath. But if Hobbes happens to be carrying a bowling ball or the like such that he approximates Susie's weight, it becomes obvious that no amount of tuning can safeguard Hobbes with 100% accuracy. Calvin will have to dedicate other sensory modes and thus even more resources to make the discrimination.

The point is that, given limited time and material, cognitive systems can't afford to respond differentially to an indefinite number of *individuals* as stimuli—they must be treated in kinds. As Quine (1969) put it, "Color is helpful at the food-gathering level" (p. 14). Looming objects within a size range—say, bumblebee to bowling ball—might cause the same ducking response. Looming protons won't be detected, and larger looming objects should cause my *diving* out of the way instead. Failing to dive, my being crushed is multiply causable by trucks and elephants. Now, take the multiple causability of my being crushed, together with what I'll call the "branching" nature of causal networks. The branching nature of causal networks is evident from the fact that two simultaneous effects of a cause—or even nearly simultaneous effects of a cause if they are far enough apart in space—cannot be causes or effects of one another, because influence propagates spatially over time. The effect or effects in question are thus spatiotemporally contained. Together, these features mean that the sameness of at least one property of the event of my death in both cases—my being dead—is confirmed by the sameness of subsequent effects of that property or state of affairs, for example, my wife's having to dispossess our second car.

An objection here might be that it is only at some higher level of abstraction, perhaps a propositional one, that the states of affairs in question are the same. After all, my death by elephant might be slightly messier than my death by truck; it might take that much longer to identify my body, that much longer for my wife to arrive at the morgue, and that much longer to sell the car, and the car might get sold to an entirely different buyer. Indeed, this objection is like the response that we saw Lewis (2000) give to the worry that causal trumping or preemption—cases like the one in which two stones are thrown accurately at a bottle a split second apart—can render false the counterfactual

that had the first stone not hit the bottle, the bottle would not have broken. The latter is false because the second stone would have broken the bottle. In cases of trumping and preemption, the event would have occurred anyway. Recall Lewis's view that we might rescue our intuition that the first stone was the cause by saying that a *different* breaking occurred—taking events as fragile. On the other hand, we might rescue our intuition that the first stone was the cause with the counterfactual claim that the bottle would not have broken at just that time in just that way—making our counterfactual fragile. I defended the counterfactual view of dependence, and yet I want to permit saying that in those two possible worlds (the one in which the first stone strikes and the one in which it misses while the second one breaks the bottle), *something* could be the *same*. For example, a bottle collector is disappointed in the same way, and to the same degree. In the two possible worlds in which I am crushed by different heavy objects, certain mental states might come about that are identical either way, for example, my wife will be upset with the hassle of dispossessing the car. In case the bottle collector is differently disappointed in each world—perhaps he lingered longer on the possibility of fixing the bottle that Suzy's stone broke, but not the one Billy's stone broke—he might write the same letter of complaint to the city council citing broken glass and calling for better nighttime lighting in the local park to dissuade wayward youth. These examples invoking people shouldn't make us think that the only true switches in nature are the mechanisms from which representational systems are built,³ or artifacts of human design (literal switches the council might turn on to light the park). Written words and machine states are surely paradigmatic cases of effects where differences of input within a certain range really make no difference. But this is no reason to think that earlier in the causal chain

³ Although I think switches could go some way toward a realist account of certain intuitions that tend to motivate antirealism, I leave that aside. In any case, the claim is not that similarity, or what I've just called proximity of causal powers, is somehow mind-dependent. Entities have causal powers independently. I am simply teasing out how such causal powers amount to bringing their bearers under kinds. Recall that it says nothing substantive (though nothing false) about kinds or about similarities to claim that they have the same kinds of effects or similar effects. Instead, similarity and kinds can be understood in terms of possible partial sameness in their effects.

producing a representation, two otherwise strictly distinct entities might both apply a critical force to flip the *same* switch, and I see no reason not to think switches are much more widespread and simple, as examples of molecules and protoplanetary disks suggest. Switches certainly seem characteristic of complex systems, however. Indeed, this may be what makes them systems in the mechanical rather than merely thermodynamic sense, viz., the capacity to accommodate a range of inputs and produce the same effect.

Anyway, the fact that some effects of distinct entities might be identical should be no problem for the counterfactual rendering of dependence. The claim that distinct entities can have *some* identical effects is not incompatible with the requirement that distinct entities have distinct *whole* sets of effects. Indeed, the fact that some effects are multiply causable explains why, when investigating the causes of an event, we often have to look more widely at the causal network than the branch in the causal network springing from a single effect. We triangulate, so to speak, the actual cause by additional fragile counterfactuals.⁴

The upshot is that kinds and similarity can be further analyzed this way: for an instance of a causal power or property to be more or less proximate to another is for there to be more or fewer possible switches that a member of the kind can operate,⁵ and every switch makes its actuators of a kind, albeit

⁴ Suppose Suzy pleads not guilty to breaking the bottle, saying that Billy threw a stone that might equally have broken it. And suppose no ballistic evidence, no difference in the shatterings, can determine which stone broke the bottles. Still, we can question Billy, hook both witnesses up to lie-detectors, check the security cameras, etc. Unless the entire local causal network is made of switches multiply actuatable by the same two entities—the camera doesn’t have adequate frame rate or resolution, Billy and Suzy don’t have 20/20, and so on, for each way the two throws influenced the world, as though an entity could preempt all the effects of another—their twoness will be discernible. Even supposing causal influence is only ever had on switches “all the way down”—imagine that our apparently analog weigh scale, at the atomic grain-level or finer, is really pixelated, made of tiny steps each actuated at its own critical threshold—to further imagine that this might mean two entities could differ indistinguishably would be like imagining that protons could be dented or scratched in a way that could never be discovered. According to Alexander’s dictum, it just becomes difficult to maintain that such differences exist. Certainly, we needn’t worry about them.

⁵ At this point, I am inclined to think it makes no difference whether the claim is about

“wispy.” To imagine a wispy kind, consider a case where almost anything with mass and acceleration might flip some other switch in nature. Someone is in a bad mood, perhaps, and a fly or someone’s snoring could set that person off. This is not to suggest that such a fly or snore would have what is considered to be the *proper* function (see Millikan, 1984, 2017; Kingsbury, 2004, 2006, 2008) of setting someone off. The point here is to characterize the limiting case on a spectrum of similarity when similarity is characterized on the basis of sameness of effect. The psychological example above should not pose a problem, but we can imagine non-psychological examples of multiply causable effects. Of course, switch-actuators of this kind only support a single axis of induction (*all x-switch-actuators have mass and acceleration greater than y*)—not enough similarity to make them normally worth tracking for cognitive systems undergoing selection. Of course, kinds, in the more ordinary sense of the word, are often multifunctional; their members have many proximate causal powers. If the condition on kinds is that under a particular circumstance members have one of the same effects, and the condition on indiscernibles is that they have all of the same effects in every circumstance, then more “robust” kinds will be those that have more of the same effects in more circumstances. And, at last, we can now also sensibly say that members of a kind tend to have similar effects, because we can characterize *similarity of effects* in terms of sameness of *further* effects.

6.2 Variation and Reduction

I concluded the last chapter by noting that the attempt to distinguish between variation and multiple realizability, even if it was successful, wouldn’t advance the debate between antireductionists and reductionists as I understand it. Let

actual switches, merely nomologically possible ones, or metaphysically possible ones. I take it as plausible that the number of switches that could be built in any given possible world—even by a divine agent otherwise obeying the laws in that world—that could be actuated by both a fly and an elephant is less than the number of switches that could be built and actuated by the housefly and the horsefly. Metaphysical possibility is tied down by nomological possibility in this way, and nomological possibility is tied down to actuality. Or so I suspect, and beyond that, defer.

me say a bit more about what I mean, and hence, by contrast, why running together variation and multiple realizability makes a difference to that debate. Polger and Shapiro say,

The identity theory holds, according to us, that there are important and explanatory mind-brain identifications. This is a significantly more modest view than has been frequently attributed to identity theorists, but still one at odds with realizationist or functionalist theories according to which multiple realization is so ubiquitous that no interesting or useful mind-brain identifications could be discovered. We do not deny that multiple realization occurs; we deny only that it occurs to an extent that renders the identification of psychological kinds with brain kinds impossible or unimportant. (p. 144)

Of course, Polger and Shapiro are right to note the history of the *Multiple Realizability* claim in connection with antireductionism. But they think, “the claim that multiple realization is incompatible with identification and reduction is most plausible if the variation is both frequent and large. If the differences between realizers are few or small, this opens the door to useful psycho-neural identifications” (p. 56). Correspondingly, Polger and Shapiro take variation to be a “mere fact” (p. 80) that requires no further explanation. But note that if the status of a kind in relation to physics rested in some way on the degree to which its members varied, it would seem as if reductionism and antireductionism could *both* be true—one in some bits of the world, and one in another. This may be what Polger and Shapiro have in mind when they say that, “textbook ‘all-or-nothing’ identity and functionalist theories are themselves idealized—caricatures, even” (p. 39). Of course, there could well be a separate debate to be had about whether to be an eliminative reductionist about a certain thing or kind, say, minds. In any case, Polger and Shapiro accept that their modest identity theory means that kinds like corkscrews differ from kinds like gold with respect to reduction. Multiple realizability being rare wouldn’t preclude antireductionism. Thus carving out a modest identity theory in the logical space only does so within the antireductive camp. The division is a disagreement within antireductive physicalism as a broader metaphysical thesis that not every kind is a physical kind and that not everything is identical with

its realizers. The modest identity theory is that some higher-level kinds are identical with lower-level kinds. The contrasting thesis is simply that most or perhaps all higher-level kinds and things are distinct from physical kinds and realizers. Moreover, on Polger and Shapiro's view, the non-co-extensivity of scientific taxonomies—and attending *explanatory* relevance of differences—is the test as to whether realization obtains. But even if distinctions that appear in one science correspond to those being made in another, this is not the only tendency in science. Interestingly, the taxonomic test for realization implies that the *richer* our scientific taxonomies, the more such realization so defined there is, and thus the more putative irreducibility there is. It is reasonably clear that our systems of kinds have changed and will likely change; they are, perhaps irrevocably, relative to our form of life at a time.⁶ “Mere variations”—differences that are not categorized at any finer grain-level—*could* be so categorized, and “multiple realizations”—differences that are categorized at a finer grain-level—might not have been. Sub- and cross-classification may be possible that nonetheless leave a superclass intact, which Polger and Shapiro recognize.

By contrast, reductionism as I understand it—or, at any rate, the reductionism I am advocating—takes the claim that “everything is physical” to be totalizing, and denies that there could exist *any* things whose beings somehow went beyond their physical bases. It would be strange indeed if the amount of that sort of reduction changed with the progress of science. Rather, on that way of drawing a line between two views, it would be impossible for the truth of the physicalism in dispute to hold in different bits of the world. Since the reductionist believes the antireductionist is mistaken in their *general* theory of how “levels” of reality are related, the reductionist's claim ought not to be that some—even many—cases of putative multiple realizability turn out not to involve reduction-blocking patterns of variation. That modest claim would be too modest. The reductionist should claim that no patterns of variation among kind members block their reduction to physics. The disagreement between

⁶ After all, what makes distinct entities similar to a given degree is the extent of overlap between the sets of switches each can actuate, including our instruments and cognitive systems, which are changing.

Polger and Shapiro and me is thus foremost⁷ about whether there are *ever* any patterns of variation that block reduction, and patterns of variation which do not.

To shoulder that burden, reductive physicalists still have to be able to explain how multiply realizable kinds reduce to physics. Furthermore, members of kinds that “merely vary”, like isotopes of gold, can be further sub- and

⁷ There is one other less serious element to the disagreement. As I showed in Chapter 4, antireductive moves can be made in terms of other synchronic dependence relations. The reductionist about composition presumably thinks wholes are identical with all their parts, and likewise for constituted things and their complete constituents, and so should think realized things are identical with their total realizers. There are reductive realization theorists (myself), and there are antireductive constitution theorists (see Baker, 1999). Of course, having arisen, the dispute over the license to use a term needs resolving if our work as philosophers is to be directly helpful outside philosophy where we should not expect those from other disciplines to specify which of two or many philosophical terms of art they are invoking. The legacy of a term’s use might be one reason to prefer a continuous use over a less continuous use. But neither use on offer *is* entirely consistent with a single legacy. That’s not surprising, since the need to clarify notions of realization has only come to light in recent decades, and, having come to light, it appears that previous philosophers have diverged in much the same way that contemporary philosophers have—viz. differing in their analyses of the relevance of the sort of “multiplicity” one finds, for example, among camera eyes. As Shapiro (2000) and Polger and Shapiro (2016) show, Putnam (1979) calls the differences between octopus and human eyes irrelevant, and they agree with him, while Block and Fodor (1972), Gillett (2002b, 2002a, 2007), Aizawa (2009), and Aizawa and Gillett (2009) consider these to amount to multiple realizations of camera eyes. Notice, too, the presence of antireductionists and non-antireductionists in both usage camps. Polger and Shapiro (non-antireductionists) follow Putnam (an antireductionist) in their analysis of variation between human and octopus eyes. Gillett and Aizawa (non-antireductionists), and I (a reductionist), follow Block and Fodor (1972) (an antireductionist) in our more liberal use of “realization” in the case of variation between human and octopus eyes. It would be premature to take “realization” as a signal of antireductive commitments, because distinctions between kinds of synchronic dependence relations—to the extent they are plausible at all—are orthogonal to issues of reducibility. So we ought not to make “realization” proprietary in a way that it is not, and for those reasons, philosophers should not attempt to make it so. Talk of realization is simply too widespread, among scientists and the rest of the humanities, as well as philosophy, to expect that we might fruitfully stipulate this meaning. Most importantly, there are plenty of those who require fit-for-purpose dependence terminology without interest or perhaps even knowledge of what are largely separable issues as far as conceptualizing

cross-classified, according to the particles they involve, thereby breaking inter-taxonomic alignment. The question then becomes, why shouldn't *variation* block reduction? And that question is especially pressing if, as I have argued, a causal relevance criterion doesn't secure the distinction between mere variability within a kind and multiple realizability of a kind as involving different sorts of synchronic ontological dependence in the first place. Indeed, reductive physicalists should be able to justify the assumption that the "variation" among members of, for example, a neurological process kind is compatible with their reduction to physics.

Ultimately, then, it is the assumption that the physical credentials of kinds and their status in relation to physics rests in some way on the degree to which their members vary, that I am challenging. As I understand it and introduced it, the argument for the *Antireduction Thesis* involves a further inference from *Multiple Realizability* in addition to other premises. So the fact that "many special science kinds are multiply realizable by distinct physical kinds" should not be conflated with the thesis that, "many special science kinds are not reducible to physics." An inference to the latter involves claiming something like, "if a kind is multiply realizable by distinct physical kinds, then it is not identical with any physical kind"—which I deny.

6.3 Variation and Kind Identity

A kind's multiple realizability, or great variation among its members, does not block reduction for the same reason that minute variation does not block reduction. To see this, consider an argument parallel to the one from *Multiple Realizability* against reduction:

- [**Variable Implementation**] Many robust scientific kinds are implemented by varying and thus distinct physical bases which also fall under sub-kinds.
- [**Unity Condition for Kind Identity**] If a kind is implemented by varying and thus distinct physical bases which also fall under sub-kinds, then it cannot be identical to any specific physical kind.

physicalism and intertheoretic relations.

Given *Variable Implementation* and *Unity Condition for Kind Identity*, we would conclude:

[**Non-identity Thesis**] Many robust scientific kinds are not identical to any specific physical kind.

This *Non-identity Thesis*, together with an *Identity Condition on Reduction* would lead to the conclusion that many robust scientific kinds are not reducible to physics. But, of course, few would accept such a *Unity Condition for Kind Identity*. We can say that a variably implementable kind, for example, gold, is not identical with any of its individual implementations, for example the isotope ^{197}Au , which everyone agrees is a kind of gold, if only a “mere variation” of gold. But we won’t claim that gold is not a physical kind. And we should reject the original *Unity Condition on Kind Identity* for the same reason. It stated that if a kind is multiply realizable by distinct physical kinds, then it cannot be identical to any specific physical kind. Except for being clouded by the historically murky notion of multiple realizability, the *Unity Condition on Kind Identity* is no better than the *Unity Condition for Kind Identity*. Of course we can say that a multiply realizable kind is not identical with any *one* of its members. But nor is gold identical with a particular isotope of gold. That should not lead us to think that gold is not a physical kind. Likewise, corkscrews’ not being identical with winged corkscrews should not lead us to think corkscrews are not a physical kind. After all, winged and waiters corkscrews are physical, as are their realizers in turn. Combined with nominalism about classes, the idea is that kinds are groups of things—whether switch-actuators or switches, I need not presently decide—which are entirely made up of those things. If something is entirely made of physical things, it is physical. So, multiply realizable kinds are physical kinds. I’m going to argue in Chapter 7 and Chapter 10 that the total realizing base of a functional property actually includes the switches individuating the property. So even apparently “wildly” heterogeneous realizers share parts of their bases. Note, too, that my liberal notion of kindhood is somewhat separate from questions of how kinds are reproduced and persist, and somewhat separate from the question of why we care about some kinds—like reproducible ones—more than

others, except insofar as our own lives and cognitive apparatus are among the realizers of certain kinds by being among the effects individuating them.

For now, however, the point above amounts to the simple claim that all physically realized kinds are themselves physical. The question of reduction is also somewhat separate, although, as I said at the outset to Section 5.1, once the *Non-identity Thesis* is denied, it is not clear what condition on reduction would block reduction. Kim gestures at the same thing when he asks, “what would be lacking in a reductive derivation if bridge laws were replaced by ‘bridge principles’ which do not necessarily connect kinds to kinds”? (p. 10) The bridge principles that reduce generalizations containing kind predicates that pick out robust kinds like gold, will be most simply written in terms of those predicates and the (very short) disjunction of physical realizers. On the other hand, the bridge principles that reduce generalizations containing kind predicates that pick out less robust kinds like corkscrews, may be most simply written in terms of those predicates and the (relatively shorter) disjunction of the switches (bottles) in virtue of which there can be an effect characteristic of the common function. The switches or circumstances that partly constitute the functional properties of entities—and, *mutatis mutandis*, their dispositional, modal, kind, or sortal properties—may themselves fall under a kind. This idea, of simplifying a generalization by employing what I’ll call a “circumstantial kind,” may be what Witmer (2003) has in mind when he denies that psychophysical reductionism requires bridge principles that link mental properties to disjunctions of physical realizers, saying that “functional definitions can *themselves* serve as appropriate bridge principles” (n. 1 on p. 60).

In asking what would be wrong with the above approach, Kim hints at something like my argument against the *Antireduction Thesis*, which he calls the “disjunctive strategy.” This strategy denies a claim like the one made by Fodor (1974), that, “if psychology is reducible to neurology, then for every psychological natural kind predicate there is a co-extensive neurological natural kind predicate, and the generalization which states this co-extension is a law” (p. 104). Now, Kim does dedicate a few central pages to arguing that Fodor’s (1974) view, in particular, provides no justification of the assumption that, in Kim’s words, “disjunctions of heterogeneous kinds are unfit for laws” (p. 10).

But the disjunctive strategy is undermined by Kim's second strategy.⁸ Kim can't, and doesn't, simultaneously pursue the disjunction strategy and deny the projectibility of multiply realizable kinds because his other argument against *Multiple Realizability* on the basis of the *Causal Individuation of Kinds*, if it was successful, would undercut the disjunctive strategy in the sense that the identity of multiply realizable "kinds" is not in question if they are not kinds at all. We simply needn't care about the identity of disjunctive sets as non-kinds. And Kim explicitly burns the bridges of the disjunctive strategy in that way, saying,

to dismiss such disjunctions as "wildly disjunctive" or "heterogeneous and unsystematic" is to label a problem, not to offer a diagnosis of it. In the sections to follow, I hope to take some steps toward such a diagnosis and draw some implications which I believe are significant for the status of mentality. (p. 10)

Kim then takes his argument from *Causal Inheritance* and the *Causal Individuation of Kinds* to be such a rationale. For Kim, the "local reductions"—in my view, subclassifications—which figure in the disjunctive strategy simply motivate splitting the putative functional kind into subkinds that may be kinds, and eliminating the superkind which is an upprojectible non-kind from the overall scientific taxonomy. Of course, what is at issue is whether disjunctions can figure in generalizations and thus constitute projectible kinds, and Kim thinks not. But as we've seen, the failure of the argument from the *Causal Individuation of Kinds* to target claims of *Multiple Realizability* in its valid scope leaves it susceptible to counterexamples of projectibility of kinds up and down the scale of "robustness." Switches—multiply causable effects—explain such projectibility. So there is no intelligible rationale for banning disjunctions. Moreover, the overall lesson is that the disjunctive strategy is not only an option, but that many kinds are basically disjunctive, and this does not rule out their projectibility. All but the *most* robust of kinds—qualitatively identical

⁸ Note that these strategies could be run as a dilemma only by admitting that one of the arguments fails. What makes it a dilemma, however, is that the reason one of the arguments fails (that similarity and kinds come by degree), is a reason to think many kinds are disjunctive. But Kim says none of this.

things—allow for variation. Stable isotopes of hydrogen can have one neutron or none in their nucleus, so stable atomic “hydrogen” is *deuterium-or-protium*. Hydrogen is nonetheless a physical, projectible, and perfectly scientific kind. These more robust kinds are just multifunctional kinds, that is, conjunctions of disjunctions. For example, to be a mammal is to be warm-blooded *and* nursing, while, on my view, there are as many ways (in the non-circular sense) to be warm-blooded as there are distinct warm-blooded things, and as many ways to be nursing as there are distinct nursing things. The kind is the intersection of these disjunctive classes. The more they can share effects, the more shared properties, the greater the number of intersecting classes each member of the kind is also a member of, and thus the more there are axes of projectibility of that kind, that is, the more are the generalizations that can be made about each member. These are what I’ve been calling “robust” kinds. I take it that most of our “natural” kinds are robust. Still, going in the other direction, we have the possibility that rather different things can fall under a minimally-projectible kind. Thus Witmer (2003) says, “While Kim worries that the predicate ‘pain’ hides the disjunctive character of the property, one might just as well worry that the disjunction predicate used to express the same property *imposes* a false impression of unprojectibility” (p. 65).

Not only would the familiar argument from *Multiple Realizability* to the *Antireduction Thesis* “prove too much” (Witmer, 2003, p. 60), it would prove *much* too much. Since there is no deeper distinction to be made between functional kinds and other physical kinds, and since there’s no denying *Multiple Realizability*, if we accept *Unity Condition on Kind Identity*, we would be forced to accept the doubly disturbing conclusion that chemical properties and the like cannot appear in laws and are not genuine kinds in the first place. Either what is wrong with lower-level disjunctions infects higher-level kinds, or whatever is right with so many higher-level kinds absolves disjunctions. I take the latter to be the right thing to say.

Let me defend this against a possible objection. Fodor (1997), for example, thinks there is a crucial difference between open and closed disjunctions, and the realizations of a psychological kind is an open disjunction due to the nature of the psychological property, since it is supposedly always possible that it

might be realized by something else. Even more interestingly, Fodor thinks functional kinds whose realizers are an *open disjunction* can be projectible, whereas it is *closed* disjunctions that are unfit for laws.

The functionalist assumes that there are laws about pains “as such”; so [the functionalist] infers that, though pain is multiply based, it is not (merely) disjunctive. So [the functionalist] infers that pain is unlike jade in the respects that are relevant to the question of projectibility. Kim, going the other way around, assumes that pain is (merely) disjunctive, hence that it is relevantly similar to jade, and hence that there aren’t any laws about pain. (p. 153)

To Kim’s view that, “What all samples of jade have in common is just these observable macrophysical properties that define the applicability of the predicate ‘is jade’ ” (Kim, 1992, p. 24), Fodor replies, “This, I say, is just wrong; and resolving the metaphysical issues about projectibility that Kim has raised turns on seeing that it is” (p. 154).

By his own lights, given the success of various special sciences, I think the burden is on Fodor to say why closed disjunctions are unfit for laws. Fodor has to demonstrate two things, first, that some kinds (e.g. jade) are closed disjunctions, and second, that closed disjunctions are unfit for laws. Here is Fodor’s defense of the first point:

Suppose that, pottering around in the cellar one day, you succeed in cooking up a substance—out of melted bottle glass, let’s say—that is, for all macroscopic purposes, indistinguishable from jade.... Have you, then, *created jade*? Not on your Nelly. (p. 154)

If you want to make some jade, you have to make either some jadite [sic] or some nephrite; just as, if you want to make some water, you have to make some H₂O. As with most of the metaphysical claims one comes across these days, the one that I just made relies on a blatant appeal to modal intuitions. But I think the modal intuitions that I’m mongering are pretty clearly the right ones to have. If you don’t share mine, perhaps you need to have yours looked at. (p. 154)

I don’t share this modal intuition, so my intuitions need looking at. Fodor is denying that jade is a functional kind, while claiming that it is a (closed)

disjunctive kind, and insisting on the difference. What, then, are closed disjunctive kinds? They are supposed to have all the metaphysical necessity of *water is H_2O* , but admit of more members. The comparison is apt—but perhaps not for the reason Fodor thinks—insofar as the putatively closed disjunctive kind (jade as jadeite or nephrite) seems to have members with particular chemical identities. That is, jadeite is $NaAlSi_2O_6$ or $Na(Al,Fe^{3+})Si_2O_6$, and nephrite is $Ca_2(Mg,Fe)_5Si_8O_{22}(OH)_2$. The comparison is apt, but the problem is that water is not H_2O (see, for example, Bursten, 2014, pp. 640–641). Rather,

water is composed by oxygen and hydrogen in various polymeric forms, such as $(H_2O)_2$, $(H_2O)_3$, and so on, that are constantly forming, dissipating, and reforming over short time periods in such a way as to give rise to the familiar properties of the macroscopic kind water. (Ross and Ladyman, 2010, p. 160)

Likewise, the notation $NaAlSi_2O_6$ or $Na(Al,Fe^{3+})Si_2O_6$ doesn't sufficiently pick out jadeite (note, the chemical definition is itself a disjunction): it only gives a rough idea of the ratios of the constituents and what some of the main bond arrangements are. Bursten (2014) points out that,

compositional ratios are insufficient to individuate fundamental chemical kinds. The drug thalidomide was prescribed as a sedative until it was found responsible for a number of radical birth defects... It was later discovered that thalidomide has... two different geometric arrangements that are not superimposable. ... The difference between the two isomers is not one that can be accounted for by counting the components of the molecules nor the components of any part of the molecule... But that difference makes a difference: one isomer of thalidomide was shown to be responsible for the birth defects and the other for the sedative effect the drug was intended to produce... (p. 640)

Likewise jadeite and nephrite are additionally individuated by chemical and macrophysical properties. Even the latter are a sort of reactivity broadly construed. Now, if melted bottle glass is to be “indistinguishable” from jade, that will be because its macrophysical properties fall within the range that jadeite and nephrite's properties fall within, for example, hardness and density. Notably, jadeite and nephrite don't have the same hardness and density, either

(they each have a range of densities and hardnesses that do not even overlap). Historically, then, their being both considered “jade” suggests that the cultural switches—stone-carving practices—that pick out jade were simply not so fine-grained as x-ray crystallography methods. So, if you cook up what Fodor says looks like jade, walks and talks like jade, yes, on your Nelly, you’ve got jade. Perhaps because jade with a particular provenance has some *additional social function* in relation to some tendency to value rare or original things, we might explain Fodor’s distaste for artificial jade.⁹ A range of macroscopic properties (recalling that Fodor’s concoction falls within this range), plus provenance might make “authentic jade” the multi-functional kind that it is. Still, that would make “authentic” or “natural” jade a more constrained kind (a less open disjunction?), but not a closed disjunction.

Or consider the example of the noble gases, plausibly a robust kind, sharing, as they do, the properties of being gaseous at standard temperature and pressure, and of being relatively inert. Suppose that, pottering around in the cellar one day, you succeed in isolating a substance that is, for all macroscopic purposes, a relatively inert gas at standard temperature and pressure. Have you discovered a noble gas? Yes. Indeed argon, krypton, neon, xenon and radon were discovered around the turn of the 20th Century, and grouped in virtue of their properties more than a decade before anything was known about the reason for their inertness. Suppose that, pottering around with a particle

⁹ Of course, one can see how provenance might be a take-it-or-leave it criterion. Indeed, Canadian nephrite is imported to New Zealand and used in carving traditionally-inspired designs. *Pounamu* is the Māori name for native greenstone traditionally used to make tools and ornaments, but the original New Zealand sources of greenstone are controlled by a charitable trust. Carvings bearing the supplier’s stamp of authenticity command ten times the price of carvings made from imported stone, such that either limited supply or a demand for lower pricing sustains a market for carvings made with Canadian stone. Budget-conscious Canadian tourists might bemoan the two transpacific voyages made by their souvenirs’ materials, but who will cry for them? Unsuspecting buyers (those who didn’t read the label) might later regret their purchase. Again, I don’t harbour Fodor’s sympathies, as though anyone has been cheated. If inefficient, the practice of carving traditionally-inspired designs out of Canadian stones is legal. And if it could be made in the cellar, or a lab, it would still be jade. You’re far more likely to get locked up for poaching greenstone with the more prestigious provenance.

accelerator, you succeed in cooking up a substance with the atomic number 118. Will you predict that, like its lighter cousins, it is relatively chemically unreactive? Yes. Can supernovae cook up new elements, enzymes cook up new polymers, evolution or engineering cook up new organisms and inventions? Yes, yes, yes, and yes. Will many of these fall under existing kinds? Yes—they will be elements, polymers, organisms and inventions, among others. *Most* kinds are open disjunctions. So is jade. “Scientists discover new type of jade,” is not an impossible headline, any more than, “Scientists discover new type of antibiotic.” The only reason you need to “be careful with your inductions,” (Kim, 1992b, p. 26) like “Jade is green,” is that you can’t trust people to fix the extensions of their kind predicates in advance of every discovery. Bursten (2014) points out that the fact that “‘topaz’ refers to all colors of $\text{Al}_2\text{SiO}_4(\text{F},\text{OH})_2$ where ‘ruby’ refers only to red variants of Al_2O_3 ,” “counts against microstructural essentialism” (p. 647). Moreover, as she explains,

the buck simply does not stop there, because color variation in these and all crystals is not a matter of unprincipled chance. Rather, color variation is the result of the presence or absence of a few nuclei of additional elements or of aberrations in internuclear distances in the crystal structure. (p. 647)

For instance, the brilliant red of rubies is due to the presence of chromium instead of aluminum in a few of the aluminum bond sites (and the more chromium, the redder the ruby). The general class of Al_2O_3 crystals, with chromium aberrations or without, is known as “corundum.” The general class might have all been called “ruby,” but it was not. . . . [The ruby] variation has different reactive properties than nonchromiated samples of corundum, because if the aluminum and oxygen in the sample were dissolved, chromium would precipitate out from the samples of ruby and not from the other samples. (p. 647)

Bursten (2014) is advocating a multi-grade causal individuation of chemical kinds. In the case of jade, we might decide either way whether or not “jade” must be green. Less usefully, we might disagree or remain equivocal. But either way, there is a kind of stone that is well suited to certain methods of carving certain kinds of things, and there is sub-kind of stone that is so suited and is mostly green. The whole point of my account of kinds is that what the kinds

are (or, more specifically, what projections we can make) depends not only on the way their members are but on the way the rest of the world is, including us. It is appropriate that our taxonomies should become more complex as more switches are employed to measure similarities and differences. But, as noted in Section 6.2, it would be odd if the discovery of more and more modest multiple realizability or fine variation was taken as counting against the physical identity of higher-level classes or kinds.

Recall, Fodor has to maintain the distinction between closed and open disjunctions, and the prospects for that look grim. He also has to say why closed but not open disjunctions are unfit for laws. Here's Fodor's effort to say why closed disjunctions are unfit:

It's not hard to see why it's so plausible that there can't be laws about *closed* disjunctions. Presumably the nomic properties that a thing has qua F or G are either properties that it has qua F or properties that it has qua G. That's why, if being jade really is a disjunctive property (if being jade is just being jadite [sic] or nephrite) then *of course* there are no laws about being jade "as such"; all the jade laws are ipso facto either jadite laws or nephrite laws. (p. 157)

Notice that by referencing a *plurality* of properties and laws above, Fodor has slipped into thinking about a greater portion of the distinct causal profiles of jadeite and nephrite as sub-kinds of jade. Of course, *all* of F's properties will be ones it has by being an F rather than a G. That does not preclude there being laws about F and G. If we were only ever interested in *all* the properties of a thing, we'd only ever be talking about an individual, which is not itself multiply realizable, and we wouldn't be making generalizations at all. A *particular* realizing basis of a functional kind as an open disjunction, like pain, would have *all* of its other properties and obey all of the laws it does in virtue of being *just that* instance of pain. By the same reasoning, that would mean there are no laws about pain; all the "pain laws" would be ipso facto sets of laws about one or another instance of pain. So Fodor simply gives up too much here to sustain the projectibility of the open disjunctions he wants to defend as projectible in the case of the kinds picked out by psychological predicates. Instead, we're frequently interested in subsets of things' properties.

Interestingly, Kim (1992b) doesn't think open disjunctions are any different than closed ones, but all he says is that, "If the list is 'open-ended', that's all right, too; it will not affect the metaphysics of the situation" (p. 8). So Kim also thinks Fodor is mistaken on this point. That would make for just one burden of showing that disjunctions—whether open or closed—are fit for laws. But I have already said much in the way of alleviating that burden. In Chapter 5, I claimed that arguments against the projectibility of functional kinds fail. Kim says, "we need to go beyond a sense of unease with disjunctions and provide an intelligible rationale for banning them" (p. 8). He provides an argument for their unprojectibility that hinges on the causal individuation of kinds, which under analysis, just highlights the gradedness of kindhood in terms of overlap in causal profiles. The argument fails to demonstrate that single-functional (or "wispy") kinds aren't projectible, even if in fewer predictive dimensions. We can make generalizations about functional kinds, such as, *antibiotics bind with catalytic enzymes essential to cell wall biosynthesis*. And in Section 6.1, I gave a causal individuation account as to why. Similarity is a matter of degree because differing things can sometimes have one or more identical effects in branching causal networks.

Wrapping Up

The upshot is that there are a few reasons to think that multiple realizability is not special. The first is that we have no good reason to think it is, as I argued in Section 5.2. Polger and Shapiro provide no way to distinguish variation and multiple realizability and maintain that realization is an ontological dependence relation. The second reason to think that multiple realizability is not special is that we have good reason to think it isn't: the same causal account of similarity explains variation both great and small, as I argued in Section 6.1. The sameness of effects by which similarity can be non-circularly analyzed makes for a spectrum of robustness of kinds on which both kinds whose members vary only minutely and kinds whose members vary wildly both fall. If that's right, it actually suits the reductionist, who needs in any case to deny something other than *Multiple Realizability*. So the third reason to think multiple realizability

is not special, is that the larger picture in which it is special—antireductionism, including a modest identity theory—is problematic as a form of physicalism, as I argued in Chapter 4. The best way to counter the antireductive argument is instead to deny that variation, great or small, blocks reduction. The identity of kinds as physical kinds does not rely on their one–one relation to other kinds all the way down to, say, fundamental particles. Kinds should be identified with their members. Higher-level kinds are scientific for the same reason that subclasses of those kinds are: they can be expected to have some of the same effects. On that account, reductive physicalists can answer Kim’s challenge and account for the projectibility of multiply realizable functional kinds, as well as defend the status of higher-level kinds which admit of variation, without having to—unsuccessfully—try to eliminate them.

Chapter 7

Sloppy Sufficiency and Restricted Constitutionalism

I have said that the trouble with the argument for antireductionism premised on multiple realizability lies with the further commitment to the *Unity Condition on Kind Identity*, and defended the view that kinds, as disjunctions of physical members, are themselves physical and projectible. Kindhood is a matter of degree, from qualitatively identical members of the most robust kinds which are projectible in every way, to members of very wispy kinds that share in only one effect. How kinds persist or are reproduced is another matter not of present concern. But antireductionist arguments can be run in terms of *individuals* instead of kinds. For example, as we saw in Chapter 4, one might claim that a desk composed of wood could have been composed by a splinter or two less, that a statue constituted by some clay would be destroyed if the clay was squashed, but the clay would not be destroyed, and that by reason of these cases (and uncountably many like them), the world is full of higher-level entities bearing non-categorical properties which are not had by the lower-level things with which they are coincident. Thus, they are distinct, or so the story goes. In this chapter, I argue that this sort of reasoning is flawed.

To expose the flaw, I explore the way in which Bennett (2003, 2008) defends an antireductive physicalism in light of the exclusion problem. She thinks that effects can have more than one sufficient cause without being overdetermined. That sounds like an odd claim, so in Section 7.1, I explain the idea. In Section 7.2, however, I will argue that Bennett's strategy, rather than

making sense of this non-exclusion claim, either fails to sustain the complete coincidence of the relevant mental explanandum and its putative physical explanans, or fails to sustain their distinctness. This amounts to a failure to sustain the distinctness of the mental and physical generally. I use this case to diagnose antireductionism's fallacious inference to distinctness, which I call the "sufficiency error," since it involves an unmotivated notion of sufficiency. In Section 7.3, I show that the sufficiency error is in fact widespread in both antireductive and eliminative approaches to non-categorical properties generally, and to minds specifically.

I take up Bennett's argument in particular because, in it, issues with physicalism clearly cross over with concepts of dependence. The choice between the two ways of being a "building determinist" that we saw in Section 3.2.2, comes back into view here, and Bennett's particular defense of antireductionism is a perfect way to highlight what's wrong with our intuitive but sloppy *Circumstantialism* about building relations as compared with building *Necessitarianism*. The question is whether to be a building "necessitarian", counting only the fully sufficient conditions for an entity or property as its "full" builders, or a "circumstantialist", counting certain physical things or properties as the full builders of a further thing or property for which they are sufficient given some other "background" conditions. As it turns out, the answer to this question has consequences in terms of the ease with which we make the sufficiency error. We should be very cautious of counting less than the complete conditions necessitating an entity as the "full" builders of that entity, as I signalled in Section 3.2.2. My crucial premise that anything on which an entity or property depends synchronically is constitutive of that entity, is a species of necessitarianism. And my crucial premise that entities and properties are identical with everything they depend synchronically on, hinges on antireductionism being a mistake.

I will go on to argue in Chapter 10 that restricted constitutionalism—and the *Circumstantialism* of which it is a species and with which antireductionist arguments begin—involves the sloppy notion of sufficiency that leads us astray with respect to reduction. That is one more reason to prefer one of the two contentious premises of my main argument, *Unrestricted Constitutionalism*,

according to which anything on which an entity depends synchronically is constitutive of that entity. Further, the failure of both closely related arguments for antireductionism—on the basis of multiply realizable kinds seen in Chapter 5, and on the basis of individuals that bear non-categorical properties such as being multiply realizable—recommends that we believe the other contentious premise of my main argument, *Reducibility*. That is, we should think that identity is the relation between an entity or property and all that it depends synchronically on. Accepting *Reducibility*, without requiring that non-categorical properties be eliminated from our ontologies, is crucial to the inference from our partial constitution by technologies to the conclusion that we—as legitimate objects of science—are partly what we make.

7.1 Avoiding Overdetermination

We saw in Chapter 4 that the antireductionist argument rests on claims like those of Baker (1999) with respect to the coincidence and distinctness of entities. Coupled with physicalist commitments, such claims amount to a recipe for a grounding problem and an overdetermination problem (or “exclusion” problem). It is hard, for example, to maintain the distinctness of the mental and the physical, the causal efficacy of both, and not worry about what would apparently be widespread overdetermination of effects by both a physical and a mental cause. In the case of physical and mental entities, Bennett (2008, p. 281) renders the problem as a set of inconsistent claims, one of which ought to be rejected:

[**Distinctness**] Mental properties (and perhaps events) are distinct from physical properties (or events).

[**Completeness**] Every physical occurrence has a sufficient physical cause.

[**Efficacy**] Mental events sometimes cause physical ones, and sometimes do so in virtue of their mental properties.

[**Nonoverdetermination**] The effects of mental causes are not systematically overdetermined.

[**Exclusion**] No effect has more than one sufficient cause unless it is overdetermined.

There is actually one crucial ingredient for the problem which is not explicit

in Bennett’s setup. What would make the effects of mental causes *systematically* and thus problematically overdetermined—rather than contingently and unproblematically overdetermined, as in firing squad cases—is the distinctness and efficacy of the mental, *and* its spatiotemporal coincidence with physical conditions that are also equivalently efficacious for an effect. In order to be worrisome causal competitors, the two causes must be spatiotemporally coincident. Without this coincidence, as we saw in Chapter 4, there is no problem. Of course, the antireductionist is not without recourse, as I signalled in Section 4.4 and promised to say more about here.

Bennett (2003, 2008) defends an antireductive physicalist “compatibilism” against the problem by rejecting *Exclusion* and maintaining that the remaining premises are compatible. She thinks that effects can have more than one sufficient cause without being overdetermined. Bennett (2003) begins with the thought, which I will not dispute, that cases of overdetermination must non-vacuously satisfy *both* of two counterfactual conditionals. In the usual case of physical (p) and mental (m) events that are candidate causes of an effect (e), the relevant counterfactuals are: if p had occurred without m , e would still have occurred, and if m had occurred without p , e would still have occurred. If m and p are separately efficacious for e , it should be possible for p to occur without m , or vice versa, or both, and if they are to be causal competitors, given the occurrence of either p or m , e will still occur. However, one strategy Bennett thinks is open to the compatibilist is to say that p cannot occur without m because p necessitates m .

All the compatibilist would need to do is argue either that it is impossible to have the mental event or property without the physical one, or else that it is impossible to have the physical event or property without the mental one. These claims would render [each counterfactual] vacuous, respectively. (2003, p. 483)

Bennett calls this the “vacuity” strategy. Of course, the compatibilist is not going to claim that m cannot occur without p specifically, since the multiple realizability of m is the antireductionist’s reason to believe its *Distinctness* from p . Bennett says, “It is quite popular for those who deny type and/or token identity to claim that this kind of asymmetric dependency holds—that

the physical necessitates the mental, even though the mental does not return the favor” (2003, p. 484). But the same line of reasoning, at a glance, spells trouble for the vacuity strategy generally, because “*p* can occur without *m* just as easily as *m* can occur without *p*” (p. 485). Bennett offers the example of the dispositions of keys. Without being changed itself, a key can lose the disposition to unlock something if the lock is changed. The key alone does not necessitate the disposition. In other words, “physical events are not less modally flexible than mental ones. They are just modally different” (p. 485).

But Bennett thinks the vacuity strategy might be rescued. She notes that just because *p* might not necessitate *m*, *p* in some particular circumstances will. Returning to dispositions, a key and a particular lock together do seem to instantiate the key’s having the disposition to unlock the lock. Bennett acknowledges two further problems at this point. The first is that “taking this line threatens to undercut the motivations for insisting that the mental and physical are distinct” (p. 486). Thus,

If there are physical events that *do* necessitate mental ones, then those physical events *do* have the relevant relational properties essentially, and *they* are the ones on which the denier of token identity should be focusing. And while there are other arguments against identity—namely, the more popular multiple realization-ish arguments against the downward necessitation from the mental to the physical—the same point presumably applies, *mutatis mutandis*, to them. (p. 486)

In other words, given some circumstances (C) in which *p* necessitates *m*, *p*-in-C also instantiates the non-categorical properties attributed to *m*, and it is not clear why we should not identify *m* with all of *p*-in-C. We shall return to this point shortly.

The second “problem” (although I take it not to be a problem but an important consequence) is that “rescuing the vacuity claim in this way” involves saying that “*p* is not in fact causally sufficient for *e*” (p. 487). The vacuity is that of a different counterfactual. The claim, if *p*-in-C had occurred without *m*, *e* would still have occurred, is vacuous because *p*-in-C cannot occur without *m*. So the vacuity strategy ultimately hinges on the idea that “if a mental cause is efficacious in bringing about some effect, the only physical causes that are also

efficacious in bringing about that effect are ones that necessitate the mental cause” (p. 487).

But this, Bennett thinks, reveals a way that the compatibilist can claim one of the counterfactual tests is not vacuous, but *false*, making worries about overdetermination moot. The compatibilist can say that if p had occurred without m , e would *not* have occurred, because “the conditions that must hold for p to bring about e —physical conditions, note—are basically the same as the conditions in which p necessitates m ” (p. 488). Bennett calls this the “falsehood” strategy.

Together, Bennett thinks the compatibilist can wield these two strategies as a dilemma about causal sufficiency. We might adopt a strict notion of causal sufficiency, “according to which the only thing that counts as causally sufficient for an effect is a whole big package, consisting of what we might have intuitively thought of as a cause *plus* all necessary background conditions and causal intermediaries by means of which it brings about its effects” (p. 490). Call this package p^* . Then the condition that if p^* had occurred without m , e would still have occurred, is vacuous, and overdetermination is not established. On the other hand, we might “adopt a rather more permissive notion of causal sufficiency—a notion on which an event or property can count as causally sufficient for an effect even though it requires some set of background conditions and intermediaries in order to bring that effect about” (p. 490). Then the same counterfactual form, with a narrower set of physical conditions plugged in—just p —will usually be false, and overdetermination is not established.

Bennett notes how similar the strategies are:

both strategies claim, implicitly or explicitly, that the conditions in which the physical event or property occurs with the mental one are the same as the conditions in which the physical event or property manages to bring about the effect. . . . The only real difference between the two, then, is the notion of causal sufficiency on which they rely. The vacuity line relies on a rather strict notion, while the falsity line allows us to continue on with our somewhat sloppy notion of causal sufficiency according to which reasonably normal physical events like patterns of neural firings count as causally sufficient for action. (p. 489)

Relatedly, recall from Section 3.2.2 that Bennett (2017) thinks it is “just a matter of bookkeeping” (p. 54) which of two versions of “building determinism” one accepts. Building determinism is the principle:

[**Building Determinism**] “For all x and y , and all building relations B , if x fully B s y , there is some $z(\neq y)$ such that necessarily, $z \rightarrow y$ ” (p. 51).

Bennett calls her preferred way of being a building determinist, “circumstantialism,” which is the principle:

[**Circumstantialism**] “Let C be some to-be-specified set of background circumstances that includes neither y nor anything that fully builds y . For all x and y and all building relations B , if x fully B ’s y , then, $\Box[(x + C) \rightarrow y]$ ” (p. 52).

Bennett is not the only one to prefer *Circumstantialism*. Many theorists explicitly adopt formulations of dependence relations in this vein. Without worrying about the details of each of the following formulations, notice the reference to circumstances, background conditions, or context (the boldface emphases are mine):

Properties G_1 - G_n instantiated by individual x ’s proper parts y_1 - y_n realize a property F instantiated by that individual x , **under background conditions B** , if and only if, x instantiates F and x ’s proper parts y_1 - y_n instantiate G_1 - G_n ; it is necessary that if y_1 - y_n are proper parts of x , and y_1 - y_n have properties G_1 - G_n , and B holds, then x has F ; and y_1 - y_n having G_1 - G_n , along with their causal powers, provide something within the spectrum from aggregate to composite to mechanistic explanation for x having F , along with its causal powers, under B , but not vice versa. (Endicott, 2016b, p. 3361)

(Realization) Property/relation instance(s) of F_1 - F_n realize an instance of a property G , in an individual s **under conditions $\$$** , *if and only if*, under $\$$, s has powers that are individuating of an instance of G in virtue of the powers together contributed by the instances of F_1 - F_n to s or s ’s constituent(s), but not vice versa. (Aizawa and Gillett, 2009, p. 186)

(Comprising) Powers C_1 – C_n , had by individuals s_1 – s_n (or individual s^*), comprise the power C^* , had by individual s^* **under background condition \$**, *if and only if* the mechanisms grounded by the triggering and manifestation of powers C_1 – C_n , under triggering conditions t_1 – t_n and background condition \$, would together implement the mechanisms grounded by the triggering and manifestation of C^* , under triggering conditions t_1 – t_n and background condition \$, but not vice versa. (Gillett, 2007, p. 202)

(Constitution/Parthood) Individuals s_1 – s_n constitute, and are parts or constituents of, an individual s^* , **under conditions \$**, *if and only if* s_1 – s_n are spatially contained within s^* , many of s_1 – s_n bear spatiotemporal, powerful and/or productive relations to each other, and all the powerful properties/relations of s^* are realized by properties/relations of s_1 – s_n , under \$, but not vice versa. (Gillett, 2013, p. 323)

As Bennett (2017) puts it, “what we call the building base is [frequently] not sufficient-full-stop for what it builds; some kind of appeal to the circumstances in which it exists or obtains is required” (p. 54). But the other way to espouse *Building Determinism* is to say,

[**Necessitarianism**] “For all x and y and all building relations B , if x fully B ’s y , then, necessarily, $x \rightarrow y$ ” (p. 52).

The bookkeeping options here, I submit—and I think Bennett would agree—are exactly analogous to the strict and sloppy options for conceiving of what counts as causal sufficiency. After all, on Bennett’s view, causation is a building relation. The sloppy notion of causal sufficiency is thus a special case of *Circumstantialism*, which counts just some intuitive core—say, a key, or some neural goings on—as the “full” builders, and the strict notion of causal sufficiency is a special case of *Necessitarianism*, which counts only those which completely necessitate the built as the “full” builders. And the reasons to opt for the sloppier *Circumstantialism* over the apparently cleaner *Necessitarianism* is the preservation of the same intuitions and ordinary ways of speaking that are reasons to permit the sloppy notion of causal sufficiency. Notice, too, that *Unrestricted Constitutionalism* is a species of *Necessitarianism*. *Unrestricted Constitutionalism* is the claim that anything on which an entity

depends synchronically is constitutive of that entity, a premise in my main argument introduced in Section 1.2, and a notion of constitution at issue in the debate over externalism, as we saw in Section 2.3.

Likewise, restricted constitutionalism is a species of *Circumstantialism*. These are probably the default notions of constitution and building generally. Bennett (2017), for instance, thinks it “sounds odd” to have a motley of relata in the first place of a building relation—for instance, physical laws, parts, and interrelations among parts—all as builders of some entity. *Circumstantialism* accords with our intuitions about what the core builders are, and bearers of the resultant (in the favourable circumstances) non-categorical properties. *Circumstantialism* respects the way we normally label, locate and *delineate* entities like statues, keys, national flags, minds and people. She says,

I can have intuitive building bases and the somewhat ungainly implementation of the modal requirement, or I can have a cleaner implementation of the modal requirement and uglier, more complex building bases. In the latter case, much ordinary building talk must be treated as invoking mere partial builders. I have opted for the former. . . . But nothing deep turns on this. It at most affects where we sweep various untidinesses. (p. 54)

Actually, what gets swept under the rug is the triviality of the distinctness upon which antireductionism is predicated. We also have it from Bennett that, “no one thinks that mere partial builders necessitate what they build” (2017, p. 49). The idea, then, that full building nonetheless “involves” necessitation, while being allowed to be considered “filled” by some intuitive but mere subset of the actually complete set of physical conditions required for, say, a mental state to be instantiated, means that “full” building only involves necessitation in the circumstances. Or, to put it more pointedly, only “full”-building-in-the-circumstances involves necessitation. But then the “fullness” is unmotivated, except by intuitions. Sheer sufficiency, necessitation, or dependence alone cannot discriminate any core from among the physical conditions that are sufficient for and necessitate an entity, or upon which it depends. Bennett recognizes this when she approvingly adapts the following passage from Lewis (1973) on causation:

We sometimes single out one among all the [builders] of some event and call it “the” [building base], as if there were no others. Or we single out a few as the “[builders],” calling the rest mere “...factors” or “...conditions.” ... I have nothing to say about these principles of invidious discrimination. I am concerned with the prior question of what it is to be [part of the building base] (unselectively speaking). My analysis is meant to capture a broad and nondiscriminatory concept of [building]. (2017, p. 55, square brackets indicate Bennett’s adaptations of Lewis)

Bennett intends the thought here to be a reason to “ignore the problem,” that is, to be neutral between *Necessitarianism* and *Circumstantialism*, because, she claims, the “same issues arise on either account.” The explanation:

There will be, on either account, a difficulty distinguishing what are intuitively the “real” builders from what are intuitively something more like background conditions. This is perhaps clearest on the account I favor: if x necessitates y in C , it is equally true that C necessitates y in x . But the very same issue arises on a necessitarian account. On such an account, complete building bases will necessitate what they build, but will be complex and often spatiotemporally scattered. The intuitively “real” or “central” or “salient” builders can of course be recovered by coupling such an account with a notion of partial building: perhaps x partially builds y when it is part of the complete base that necessitates y But no such notion of partial building will distinguish the seemingly central partial builders from the other aspects of the complete building base that intuitively seem like background conditions only. (2017, pp. 54–55)

I agree that there can be no distinguishing between parts of a dependence base on the basis of dependence alone, but that is just a reason *not* to try to “recover” the core (or “central”, or “salient”) builders on *that* basis. In Chapter 10, we’ll consider other ways “cores” might be delineated, for example, in terms of which parts are alterable by us for engineering purposes. Since we engineer technologies, if we depend on them, we engineer ourselves quite directly. But I am giving a taste of what is ahead. The point here is that sheer sufficiency, necessitation, or dependence alone cannot discriminate any core from among the physical conditions that are sufficient for and necessitate an entity, or upon which it depends.

7.2 The Sufficiency Error

This unavoidable need for circumstances is a problem for compatibilism. Consider the sloppy notion of causal sufficiency and the corresponding falsity strategy. When the physicalist holds the sloppy view of causal sufficiency, “he can think that some physical event p is causally sufficient for effect e , that some mental event m is as well, and that p fails to necessitate m ” (Bennett, 2008, p. 290). But he can only think so if he wants to be *extra* sloppy, and he cannot think he has dealt with overdetermination worries *while maintaining any substantive Distinctness*. We can, and often do, delineate some intuitive core p that is merely sufficient in the circumstances for some m , and thus insufficient for m in other circumstances in which p is also causally insufficient for e . This much entails the distinctness of m and p . The fact that certain neural firings that might (partly) realize human pain, *ceteris paribus*, but that occur instead in a petri dish, do not realize pain, or cause wincings, etc., only shows the distinctness of pain and such neural events. But it is trivial to claim that mental states are not identical to mere subsets of their realization bases. The dispute between reductionists and antireductionists ought not to be whether, analogously, wholes are identical with subsets of their parts. The thought that “ p is causally sufficient for effect e , that some mental event m is as well, and that p fails to necessitate m ” (Bennett, 2008, p. 290) is extra sloppy because there is no basis for claiming that m is sufficient for e in the same sense of “sufficient” that p is. If m is necessitated by p -in- C , and p -in- C is fully sufficient for e , then m is *strictly* sufficient for e , whereas p is only sloppily sufficient for e . The distinctness of p and m is utterly silent on the distinctness of m and p -in- C , or of pain and neural-firings-in-context. The relevant question is indeed whether m is distinct from p -in- C , again, *because* the circumstances in which p does not cause e are the same as those in which m does not occur, and thus does not cause e either. In other words, m causes e whenever p -in- C causes e and necessitates m . So, and this is crucial, if and only if they are distinct, m and p -in- C are the *relevant causal competitors*. They are the entities whose putative distinctness would be non-trivial, no matter what sloppy notion of causal sufficiency momentarily obscures that point. The falsity of the counterfactuals involving just the intuitive core, p , is the falsity

of the wrong counterfactual. The relevant tests for overdetermination are thus, “If m had occurred without p -in- C , e would still have occurred,” and “If p -in- C had occurred without m , e would still have occurred.” Since p -in- C and m are the causal competitors, it is their distinctness and equal causal efficacy that are at issue, and the compatibilist claim must be that these counterfactuals are vacuous because p -in- C necessitates m .

Another way of putting all this is that the falsity strategy collapses back into the vacuity strategy. What then, is the status of the vacuity strategy? Recall that the vacuity and falsity strategies can supposedly be run as a dilemma according to the notion of sufficiency in play. On either the strict or sloppy notion of causal sufficiency, the compatibilist is supposed to have an answer to the overdetermination problem. But what gets lost somewhere in Bennett’s dialectic is the point that the only substantive option of the two—the vacuity strategy—doesn’t work for the reason Bennett (2003) had noted when she moved on from it to defend—unsuccessfully—the falsity strategy. “If there are physical events that *do* necessitate mental ones, then those physical events *do* have the relevant relational properties essentially, and *they* are the ones on which the denier of token identity should be focusing” (p. 486). Again, the usual reason for claiming that higher-level entities are distinct from their lower-level realizers is that they supposedly have different non-categorical properties. The mental is supposed to be distinct from the physical because of differing non-categorical properties, while the very same conditions in which the narrow physical realizers necessitate the mental are the ones which ground the non-categorical properties of the mental. Any part of the extra-neural world and the relations to these that are necessary to make such and such neural firings realize pain are the ones which ground its being a “pain.” Indeed, these are two ways of saying the same thing. So the core-realizers-in-the-circumstances instantiate those properties, and this precludes the line of reasoning by which higher and lower-level entities are distinct by having distinct properties because the fully sufficient physical base, not the insufficient subset, can be said to *have* the properties it instantiates. After all, physicalism—for example, formulated in terms of global supervenience—entails that there can be no properties had by a mental entity (event, property, object, state, or process) which are not

instantiated by *some* physical entity (event, property, object, state, or process), since physical conditions are sufficient and ultimately necessary for the mental. The central premise in my counter-argument is thus Bennett's own thought that, "If there are physical events that *do* necessitate mental ones, then those physical events *do* have the relevant relational properties essentially" (p. 486). The compatibilist must affirm the antecedent, live with the consequent, and so undercut her own argument for distinctness.

The upshot is that the cheapened sufficiency brings a cheap distinctness in its wake, and the inevitable strict sufficiency brings no distinctness at all. Of course pain is not identical with c-fibres firing if the latter are only partly sufficient for pain. The question of what the fully sufficient realization base *is* for a given mental entity is an empirical question once the mental explanandum and its properties are stated, although equivocation, ambiguity, or diverging agendas with respect to the explanandum may give rise to apparently philosophical debates about what are the total realizers of one's preferred explanandum taken to be paradigmatically "mental." I shall have more to say on this last point in Chapter 10. But what the total realizers *are* is a different question than whether a mental entity is identical with its total realizers. If antireductionism is to be substantive, the compatibilist must pursue the claim that *m* is distinct from *p*-in-C, or whatever conditions completely necessitate that mental event. It is not clear that Bennett makes the substantive claim.

There is an analogous objection to her compatibilism that Bennett (2008) herself articulates. This is the objection that, as a compatibilist, you are

only getting out of the problem—if you are—by giving up on mental causation. You haven't said anything about how the mental can really be causally efficacious, and it is starting to feel as though its efficacy is only derivative. Isn't this at best a Pyrrhic victory? (p. 301)

Bennett's response is to say, rightly, that "no one can say that mental and physical causes are completely independent of each other, and yet do not overdetermine their mutual effects. That is the truth at the heart of the exclusion problem." She continues,

Perhaps doing without independent efficacy is a disturbing thought. But the fact is that it is a mistake to think that a *physicalist* can say

anything else. Physicalists need to bite this bullet for reasons having nothing to do with the exclusion problem. It is a direct consequence of their physicalism. Kim is surely right that physicalists need to accept something like his “causal inheritance principle” . . . (2008, p. 301)

In other words, Bennett thinks the exclusion argument “*is* a good argument for the claim that the mental is nothing over and above the physical” (2008, p. 282). Indeed, Bennett’s point in that chapter is to say that the compatibilist strategy is not open to the dualist and that the compatibilist means something “much weaker” than what the dualist means by endorsing *Distinctness*, namely, “that mental events and properties are not truly distinct existences that can be snipped away from their physical bases” (2008, p. 284). The thought is that, “despite not being identical, the mental and physical causes are not exactly *distinct*, either” (2003, p. 480). “The nonreductive physicalist is simply asserting that mental and physical properties are numerically distinct” (2008, p. 286).

But what does this double standard for distinctness come to? The thinking is elaborated somewhat by Bennett in a slightly different context. The principle that “built,” or “nonfundamental,” entities posited by a theory “do not count against the simplicity of a theory” explains why the physicalist says that these entities are *nothing over and above* their realizers; they are “‘ontologically innocent’ (Lewis 1991, 81), ‘no addition to being’ (Armstrong 1997, 12), and even an ‘ontological free lunch’ (Armstrong 1989, 56)” (Bennett, 2017, p. 220). Bennett says, “Nonfundamentalia are no addition to fundamental being, which is all parsimony cares about” (2017, p. 220). On the other hand, the very same principle supposedly justifies the antireductive physicalist’s contention that minds, for example, and other nonfundamentalia,

are an addition to being full stop. They are numerically distinct from—and thus in that sense “something over and above”—whatever builds them. What I am suggesting is that . . . nonfundamentalia do “count ontologically” toward the total number of things. If you are counting up how many entities or kinds of entity exist according to a theory, you have to count them all, nonfundamental and fundamental alike. (2017, p. 220)

I grant that fully built entities are ontologically innocent with respect to their synchronic builders. But that's because the liberal way of counting in question *is not*, contra Bennett, a matter of having "extra ontological commitments" (2008, p. 223). Considering Alexander's dictum (that to exist is to be causally efficacious), we should say that the "distinctness" of the mental and physical is as genuine as the distinctness of their causal efficacies. To deny the latter *is* to deny the former, and we've seen that mental causes are not efficacious for their effects independently of the physical conditions necessitating the mental: "the mental and physical causes are not exactly *distinct*, either" (2003, p. 480). So they are not genuinely distinct existences. What other distinctness could there be? None. I deny, then, that the "not truly distinct existences" (2008, p. 284) can amount to any "ontological" distinctness.

My objection, then, is that the compatibilist is only getting out of the overdetermination problem by giving up or seriously compromising *Distinctness*, and that such a Pyrrhic victory is not a victory for antireductionism at all. The distinctness of antireductionism *as a theory* rests on preserving both *Distinctness* and genuine *Efficacy*.¹ And if these are not preserved, there is no overdetermination worry, and thus no need to deny *Exclusion*—the idea that no effect has more than one sufficient cause unless it is overdetermined.

When it comes to putatively non-categorical properties, there is an untidiness that should not be swept *anywhere*, and we should remain cognizant of the problem, not ignore it. The epistemic and pragmatic quandary (discovering and listing all of what something depends on) is not escaped by the *logical* or

¹ Bennett (2003) calls herself a recent convert to compatibilism, and says at the outset that she is "not entirely convinced by it, but I do think that it is more compelling than anything else that has been suggested" (p. 473). I take her to mean other antireductionist strategies, so, on her own view, she is the reductionist's most worthy opponent, which is good for me, if I'm right that there's a problem here. And when Bennett (2008) reminds us that "Nonreductive physicalists... typically endorse Distinctness for a combination of reasons having to do with the purported multiple realizability of mental state-types, and with the semantics of mental terms," she also notes that, "I myself think that there are many interesting complexities here, and suspect—somewhat heretically—that the distinction between reductive and nonreductive physicalism is probably not metaphysically very deep" (p. 285), and "whether mental and physical properties are numerically distinct matters less than whether physicalism is true" (p. 286).

conceptual sloppiness of allowing a partial list to “count” as complete. More importantly, I submit, the main problem with *Circumstantialism* is that our drive to preserve certain intuitions is complicit in causing deep and widespread confusion about higher-level entities’ relations to lower-level ones, for example mind–body or mind–physical world relations. I have been piecing together the view that the extended mind debate, a theory of mind–body relations, and nothing less than a coherent notion of physicalism, hang on a crucial point that is obscured by one of these bookkeeping techniques. *Circumstantialism* gives up the co-extension, and thus any reason to believe in the complete coincidence of, the mental with some core physical conditions or “full” builders plugged into the formulation. But this point is easy to miss, obscuring the fact that the distinctness claim then being made is not the relevant one. After all, the mental remains co-extensive (and thus completely coincident) with some wider physical conditions, and so the latter conditions are the relevant would-be causal competitors with the mental entity they necessitate, and they are the ones from which the distinctness of the mental would amount to a substantive thesis. *Necessitarianism*, on the other hand, reinstates the co-extension (and coincidence). In doing so, it makes more obvious what the relevant distinctness claim is that a substantive antireductionism would have to make. And that claim is at once less plausible, given that fully sufficient physical conditions necessitating the mental are the same ones that instantiate the properties by which it was supposed to be distinct from the physical.

In short, one ought not to premise one’s antireductionism on the claim that some higher-level entity is not identical to some physical conditions that are not sufficient for the properties of that higher-level entity. Or, if that’s the substance of “antireductionism,” the view is trivial, and not at all at odds with a physicalism according to which either the same higher-level entity with the relevant properties is identical with an extended set of physical conditions, or according to which the narrow physical conditions that are not sufficient for their non-categorical properties simply do not “have” such properties. If physicalism is true at all, everything is physical, so apparently “non-categorical” properties just are distributed physical properties, and the physical bases which instantiate them are *by definition* not the narrowly delineated entities we

intuitively attribute them to. Rather, those properties are instantiated by distributed physical entities. This is the substance of the extended identity theory I am espousing. Non-categorical properties are ruled out as appropriate or coherent grounds for substantive distinctness claims about entities to which they are attributed, because whenever such properties are non-categorical, they are, by definition, *not* instantiated by the intuitive “core” of the realization base about which the non-identity claim is being made, and the question of their identity is no more at issue than the question of my identity with my left toe.

The choice between sloppy and strict notions of sufficiency, then, is no innocent matter of bookkeeping. The sloppy notion produces unwarranted distinctness claims and results in puzzles that still plague our thinking, especially about minds. In slogan form, we might say the problem comes from trying to squeeze a big property into too small a physical base. Call this the “sufficiency error.” The error is that we tend to take as an explanandum some property or entity with certain properties, then intuit—or “invidiously discriminate,” to borrow Lewis’s (1973) phrase—some physical core that we take to be “sufficient” (though only in the circumstances) for that explanandum. Next we note that the physical core is not fully sufficient for the properties we nonetheless attribute to it and on which we base the explanandum’s distinctness from its physical core. Finally, we take that distinctness, whose triviality is overlooked, to tell us something general about, for example, mind–physical world relations. To put all of this another way, the error is to attribute non-categorical properties to entities, use those properties as a basis for claiming that such entities are distinct from physical conditions not sufficient for those properties, overlook the fact that non-categorical properties are by definition not coincident (or co-extensive) with the narrow physical conditions they have been attributed to, and, finally, to think that the distinctness of such entities from mere subsets of the conditions sufficient for them has anything to do with the distinctness of those entities from physical conditions generally.

7.3 The Ubiquity of the Sufficiency Error

Ordinary boundary-drawing and property attribution may not be a problem on its own. It is the further inferences we might draw from such ordinary practices that are problematic. Once we no longer draw such inferences from the trivial distinctness of entities bearing non-categorical properties from mere subsets of the physical conditions fully sufficient for those properties, there is no strong pull to eliminate such entities or properties from our theories, since they are ontologically innocent. (Instead the question is whether there is any regularity worth tracking). So the scope of the sufficiency error is potentially as wide or wider than the problem of special composition, but the conclusion is not that there are no ordinary objects. I think, rather, that ordinary objects and indeed all non-categorical property-bearers are either not quite *what* we ordinarily think they are, or not quite *where* we ordinarily think they are.

Chronicling the role the sufficiency error plays in philosophical confusion would be a project unto itself. But let me mention a few other cases. First, I had to make the sufficiency error repeatedly—in claims usually signalled as questionable by labelling them “putative”—in Section 4.1 and Section 4.3, just to get the grounding and overdetermination problems off the ground. Indeed, the sufficiency error is the recipe for these problems. I risked making the sufficiency error when claiming that one or more multiply-actuatable switches are what make something a member of the kind that it is. I appealed to the switches the thing *could* or *would* operate. But this just means that a thing’s functional and kind properties are themselves modal properties, as are dispositional properties. Equivalently, its modal properties are its kind properties.² The point is just that the narrow builders or core realizers, or what

² Quine (1969) makes a similar point when he says,

what qualifies a thing as soluble though it never gets into water is that it is of the same kind as the things that actually did or will dissolve; it is similar to them. Strictly we cannot simply say ‘the *same* kind’, nor simply ‘similar’, when we have wider and narrower kinds, less and more similarity. Let us then mend our definition by saying that the soluble things are the common members of *all* such kinds. A thing is soluble if each kind that is broad enough to embrace all actual victims of solution embraces it too. (p. 16)

we consider to be the kind members, alone are not sufficient to instantiate these properties. Switches are part of the circumstances that instantiate entities' non-categorical properties.

Consider now the way the sufficiency error can motivate epiphenomenalism or eliminativism instead of antireductionism in the face of the exclusion problem. Prior et al. (1982) famously generalized the exclusion-style considerations against the causal efficacy of all dispositional properties. On my account, they should have given up *Distinctness* instead of *Efficacy*. We can affirm the causal efficacy of dispositions by identifying them with causally efficacious physical conditions, since causal impotence just spells non-existence for dispositions (and any such "epiphenomena"). Instead, in upholding distinctness, Prior et al. (1982) make the sufficiency error. First they say,

For each disposition we can specify a pair of antecedent circumstances and manifestation which together determine the disposition under discussion. In the case of fragility, the pair is (roughly) < knocking, breaking >, in the case of water solubility the pair is < putting in water, dissolving > and so on for the other familiar cases. By "a causal basis" we mean the property or property-complex of the object that, together with the first member of the pair—the antecedent circumstances—is the causally operative sufficient condition for the manifestation [of a disposition]. (p. 251)

Notice here the narrow delineation of the "causal basis". In case it wasn't clear, they say explicitly, "The basis is rather responsible for the manifestation in the appropriate circumstances" (p. 253). So the causal basis, on their usage, does not include those circumstances. However, they argue that "these bases are distinct from their attendant dispositions..." (p. 251), which, given physicalist commitments like *Completeness*, *Nonoverdetermination*, and *Exclusion*, would entail the impotence of dispositions. The reason given for the distinctness is, once again, the idea that nothing about the object—say, its having "molecular bonding α " (p. 253)—is essential to its being fragile. But this is too hasty. After all, the disposition is not completely coincident with the causal basis from which it is being claimed to be distinct. That distinctness is trivial, and the partial sufficiency bases for a property are not causal competitors with the complete sufficiency base; only entities that are distinct and completely

coincident are causal competitors. And there is no reason given to think a disposition is distinct from its causal-basis-in-the-circumstances. But, owing to the distinctness from the causal basis alone, the idea is:

This causal basis is a sufficient causal explanation of the breaking *as far as the properties of the object are concerned*. But then there is nothing left for any other properties of the object to do. By the Distinctness Thesis the disposition is one of these other properties, ergo the disposition does nothing. (p. 255)

There may be nothing left for any other properties *of the object* to do. But the distinctness of the property from the causal basis in the object is *not* equivalent with the disposition's being a property of the object alone. Quite the opposite. The trivial distinctness is predicated on the causal basis being inessential for the property. The sufficiency error is even more clear in the following inference:

the causal basis is part of the explanation and part of the cause of the disposition's manifestation; hence the disposition is not. (p. 255)

The "explanation" here, of which the "causal explanation" via the "causal basis" is only a part, is what might be called metaphysical explanation. The "causal basis" is just *part* of the total explanation of the manifestation—that is, the rest of the explanation of the manifestation (a glass breaking) involves circumstances such as the object's absorption of some appropriately inertial energy, such as its own, that of a rock, or a mechanical wave like that of sound. The two senses of explanation at work—partial and complete—also underlie an equivocation on "responsibility." We're told, "If dispositions must have causal bases, then any object with a disposition must have a property—the basis—responsible," in this case, for a glass's being at a time such that if it were dropped at that time, it would break (p. 252). But as Prior et al. (1982) themselves say, "The basis is rather responsible for the manifestation in the appropriate circumstances" (p. 253). In a world with nothing else, a glass with α is not fragile. So the fragility of the glass is not a competitor with α in the manifestation of breakage, because α is not strictly sufficient for anything's fragility. For this reason, it is not relevant to deny the identity of fragility with such a partial metaphysical explanation of breaking as α .

In short, if we go in for *Circumstantialism* about the intuitively central builders, explanation, most responsible basis, etc, we must be careful, when attributing properties to such cores, not to draw certain inferences about the irreducibility (antireductionism) or causal impotence (eliminativism) of properties or their bearers. If we instead adhere to the stricter *Necessitarianism*, in which an entity depends on, is constituted by, realized by, or composed of (and, on my view, is identical to), *all* of some intuitively-core-constituents-in-the-circumstances, then it is far clearer that neither the distinctness nor the efficacy of the fully sufficient physical basis as compared to its intuitive core is at issue, since they are not completely coincident. The substantive issue is instead whether the entity is distinct from the total realizers; but none of the arguments for antireductionism target this substantive distinctness.

And, crucially, if one wants to claim that an entity is not identical with the wider set of physical conditions strictly sufficient for it, reasoning that *those* physical conditions amount to an entity that have further non-categorical properties, for example, of the property of being otherwise realizable, then one simply repeats the sufficiency error on a bigger scale.

Thus, there is no tension between the conjunction of *Unrestricted Constitutionalism* together with *Reducibility* and the claim that what it is to be a member of a kind is to share in one or more of the same effects. My defense of multiple realizability is actually compatible with the mereological essentialism I recommended in Section 3.3.3 when I questioned Bennett’s assertion that she “could probably be in exactly the same qualitative state of pain even if one or two fewer neurons were firing” (2017, ch. 3, note 19). Once the explanandum is fixed, any superficial tension disappears. That is, if the question is whether the *kind* of state, as singly or multi-functionally characterized, is multiply realizable, the answer is *yes*. If the question is whether the state’s complete causal profile could be the same if its realization base was altered (such that *any* of its properties changed), the answer is *no*. When Bennett says “exactly the same” state—qualitative or otherwise—I take that as a reference to an instance, the identity criterion for which is its complete causal profile, rather than a functional kind, the membership criterion for which may be only its partial causal profile.

Nor, then, does antireductionism follow from something like the subset strategy recommended by Jessica Wilson (2011). The idea is basically that the conditions (putatively) sufficient for the causal powers characteristic of a mental state are also (putatively) sufficient for additional causal powers, such that the instance of the mental causal power is both numerically identical with an instance of a physical power, and not identical with the realization base. But I have been arguing that whatever the disposition, property or causal power is here, it is instantiated not only by the intuitive base—say, a corkscrew, a key, or some neurons—but by the bottles, locks, and, in the neural case, extra-neural circumstances. So the relevant question, to which differing answers would make a substantive difference between antireductionism and reductionism, is not whether a single property is identical to an entity with a set of properties. The relevant identity question for properties is whether the property is identical with a physical property, and the only satisfactory answer is that, being physically instantiated, the property itself is physical. Indeed, this means the identity question with which Putnam (1967) begins his discussion, “Is the property of having a pain at time t a brain state?” (p. 158), if it is not ill-formed, is *too* easily answered, “no.” The relevant identity question for a mental state, object, process or event is whether it has all the same properties as a physical state, object, process or event. Only physical conditions that are fully sufficient for a property instantiate and thus “have” that property, strictly speaking, despite our ordinary attributions of non-categorical properties to subsets of those physical conditions. So, for example, if we claim that a corkscrew is not identical with the chunk of metal we call a corkscrew because that same chunk of metal “has” other causal powers, namely, the property of being a paperweight under certain *other* circumstances, we’re simply being doubly sloppy: once when we attribute the property of being a corkscrew, and again when we attribute the property of being a paperweight, to a core which is actually sufficient for neither property. And again, it is nothing against the physical identity, nor even against the reducibility, of corkscrews or paperweights or mental states that they are distinct from mere subsets of the physical conditions necessitating them, and distinct from supersets of physical conditions sufficient for other properties, too.

Most importantly, it is a sufficiency error that sets up the mind–body problem. It goes like this. Begin with the question of how the mind could possibly be explained by physical conditions. Then offer up a woefully insufficient set of physical conditions—whether synapses, or, by analogy, the English-speaker in Searle’s (1980) “Chinese Room,” or gears in Leibniz’s mill (see Rescher, 1991)—as an example of an incomplete and thus failed explanation. Then, take that incomplete explanation as evidence for the distinctness of the mental and physical generally. Our intuitions about the boundaries of the higher-level things we want to explain, their central (narrow) constituents or “core” (insufficient) realizers, are complicit in a simple and very widespread fallacy about the non-identity of higher-level entities and properties with their constituents or realizers.

Wrapping Up

In Chapters 5 to 6, I have argued that antireductionism premised on multiple realizability fails. The multiple realizers of a kind, or variation among its members, are compatible with a kind’s being physical and projectible. In this chapter, I’ve argued that the non-categorical properties that we attribute to entities—whether modal, sortal, dispositional or functional properties—are no basis on which to premise any substantive distinctness of those entities from their physical constituents or the realizers of those properties. Doing so involves a fallacy I call the sufficiency error. And I have briefly sketched the breadth of the sufficiency error; it is very common.

Having been alerted to the sufficiency error, we should be very cautious of counting less than the complete conditions necessitating an entity as the “full” builders of that entity. We should expect that any projectible non-categorical properties we attribute to ourselves and to our experiences—properties that *something* might have—depend on more than the core physical parts we might ordinarily take ourselves to be constituted or realized by. Having cleared the conceptual terrain in this way, we can begin to see why we are partly constituted or realized by what we make. As I will go on to argue in Chapter 10, these issues are at the heart of the question of extended minds, since the sufficiency

error is baked into the internalist's position. Restricted constitutionalism—and the *Circumstantialism* of which it is a species and to which antireductionist arguments fallaciously appeal—involves an unmotivated notion of sufficiency that leads us astray with respect to reduction. That is one more reason to prefer *Necessitarianism*, and with it, one of the two contentious premises of my main argument, *Unrestricted Constitutionalism*, according to which anything on which an entity depends synchronically is constitutive of that entity. Further, the failure of both closely related arguments for antireductionism recommends that we believe the other contentious premise of my main argument, *Reducibility*. That is, we should think that identity is the relation between an entity or property and all that it depends synchronically on. Accepting *Reducibility*, without requiring that non-categorical properties be eliminated from our ontologies, is crucial to the inference from our partial constitution by technologies to the conclusion that we—as legitimate objects of science—are partly what we make.

Chapter 8

Causal and Synchronic Dependence

In Chapter 3, I surveyed a few dependence relations and introduced what Bennett (2017) thinks they have in common. There are similarities between dependence relations and they form a resemblance class. In Chapter 4, I showed that similar puzzles and problems arise in the case of each synchronic dependence relation, and the crux of the issues and their solutions is independent of the specific synchronic dependence relation terminology in which the problems are couched. The problem arises when we claim that an entity or property is distinct from its parts, constituents, or realizers. In Chapter 5, I examined the role that multiple realizability has played in motivating antireductionism. In Section 5.2, I argued that multiple realization is indeed ubiquitous, and in Chapter 6, I argued that that it blocks reduction if and only if supposedly garden-variety variation blocks reduction, which it doesn't. The disjunctiveness of a kind is no barrier to its being physical, nor to its reduction. In Chapter 7, I argued that a widespread fallacy I call the sufficiency error involves an equivocation—between sloppy sufficiency-in-the-circumstances and sufficiency full stop—when inferring the non-identity of an entity with physical bases that are merely sufficient-in-the-circumstances for it. We sloppily say that an entity “has” certain non-categorical properties while claiming that a mere part of its physical base doesn't, thus applying a strict sense of sufficiency to the base. The resulting non-identity of an entity with a mere subset of its dependence base does not secure antireductionism as a substantive metaphysical thesis.

But if entities are to be identified with their complete synchronic dependence bases, then partial synchronic dependence is partial identity. I take these to be good reasons to think synchronic dependence relations are identical.

I return now to Bennett's (2017)¹ class of building relations and offer, in this chapter and the next, an original argument to the effect that synchronic dependence relations are an important subclass of building relations, and that these relations are in fact identical with one another. Though I depart from Bennett on antireductionism, and although my claim that synchronic dependence relations are identical goes much further than what Bennett is willing to say about the unity of building relations more generally, I am once again engaged with Bennett because she is the main philosopher to have worked on this latter issue in depth. She clearly asks, about building relations, the question I have about synchronic dependence relations: just how similar are they, and how many are there, really? Bennett argues building relations are similar, but not one thing. I argue synchronic dependence relations are very similar, and actually one thing. I'll mention a few other somewhat relevant views, but they are only in the rough vicinity of the view I reject. Bennett, on the other hand, gives detailed arguments against monism about building relations, and since those arguments are made strictly in terms of synchronic dependence relations, if they work, they would work against the view I endorse. So they are *the* extant considerations to grapple with, although I come up with some others myself.

If I'm right, then it doesn't matter whether claims about externalism are made in terms of constitution, composition, realization, or identity. To claim that we have technological parts is to claim that we are partly constituted by technologies, which is to claim that properties attributable to us are partly realized by technologies. And both are to say we *are* partly technological. In this chapter and the next, I explain and defend that view, as I said I would in Section 2.3 when showing that externalists drift among synchronic dependence terminologies and attending underdeveloped concepts, but that this drift is not a fatal flaw for externalism.

In Section 8.1, I begin by discussing the way in which building relations are,

¹ Every reference to Bennett in this chapter is to Bennett (2017).

as Bennett puts it, “tainted” by causation. But in Section 8.2, I argue that causal taint does not render causation and synchronic dependence inseparable as subclasses of building or dependence, and that, contra an alternative view offered by Ross and Ladyman (2010) in their analysis of debates over EM, the question of whether something is a cause or constituent of another thing is indeed quite sensible, as both sides of the EM debate assume. I agree with Bennett that the members of a class of *building* relations, including causation, are not identical with one another, although for different reasons than Bennett. In Section 8.3, I discuss what it means for relations to be identical, and provide a general defense of pursuing the question of whether synchronic dependence relations are identical. What is at issue is whether or not there are any real disagreements to be had about which synchronic dependence relations obtain (or not) between, for example, the cognitive or mental and their physical bases.

In the next chapter, I discuss Bennett’s main reasons for rejecting monism about building relations. In Section 9.2 and Section 9.3, I argue that those reasons do not count against the identity of specifically synchronic dependence relations, making that thesis a live option. In fact, some of the same considerations can be turned into an argument *for* the identity of synchronic building relations.

8.1 Causal Taint

Bennett argues that, “the standard distinction between causal and noncausal determination is both less clear and less useful than typically assumed” (p. 67). She reasons that building is “causally tainted,” in two ways. Building is causally tainted, first, in the sense that the class of building relations includes causation; on Bennett’s view, causation is necessitating, asymmetric, and generative. It should be obvious that I agree there is a resemblance; both causal dependence and synchronic dependence are dependence relations. And although “resemblance classes are cheap” (p. 70), Bennett convincingly argues that the resemblance between causation and other forms of dependence is theoretically useful in that there are instructive analogies between the two. Indeed, I have employed these analogies when defending the counterfactual analysis of

dependence in Chapter 3, and, in Chapter 7, when attacking inferences involving equivocations between sloppy and strict notions of sufficiency, be it causal or synchronic sufficiency. However, similarities between causal dependence and synchronic dependence simply do not entail that the distinction is less clear and less useful than typically thought, even if it's common to overlook the similarity.

But what about Bennett's second reason to believe the distinction between causal and non-causal determination, or between causal and synchronic dependence, is unclear? Building relations, Bennett argues, are causally tainted in a second way, because many building relations obtain in virtue of causal facts. According to Bennett, "causation taints the building family by infecting specific other building relations" (p. 99). There are three grades of this second kind of causal taint, or "involvement." These grades of causal involvement lead Bennett to what appears to be a stronger conclusion, that, "causation cannot be separated out from vertical building without leaving an incomplete class that fails to carve at the joints" (p. 83). This claim appears to be stronger than the claim that the distinction between causal dependence and synchronic dependence is "less clear and less useful than typically assumed" (p. 67).

I'll revisit the inference to either conclusion in a moment, but first, what are the grades of causal involvement in synchronic building? The first way in which synchronic dependence holds in virtue of causal facts is due to the causal individuation of, for example, functional properties. A first-order property is said to realize a second-order functional property just in case the former plays the causal role individuating the latter. Not only do I endorse a causal individuation of functional properties, I have endorsed causal individuation of everything—individuals, kinds, and properties of all sorts. Things are what they do. The fact that we sloppily attribute properties to mere subsets of the physical conditions strictly sufficient for those properties presents us with an apparent choice, which I have until now left aside, as to which conditions *are* strictly sufficient for those properties. I'll say more about this apparent choice in Chapter 10 in relation to mental states. But briefly, we might say that what is actually sufficient for what we call a corkscrew to be a corkscrew is, at any given moment, human beverage-storing practices. Or we might say, a thing has the

proper function (see Millikan, 1984, 2017; Kingsbury, 2004, 2006, 2008) of being a corkscrew in virtue of having successfully functioning corkscrew ancestors. In that case, having a certain ancestry is perhaps a property instantiated by a four-dimensional entity. I'm ambivalent here, as I don't think the choice is a dilemma. Rather, I think each set of physical conditions instantiates different properties. The shorter-duration and wider physical base consisting of unactuated switches perhaps instantiates a disposition. The longer-duration set of physical conditions instantiates a thing's "having" a certain etiology. Then, whether *being a corkscrew*—or a belief, for that matter—is having a disposition or having a certain etiology is a terminological decision. Less plausibly, we might think that being a corkscrew is *either*. I say "less plausibly" because, while the extensions of these two properties are likely to overlap, that doesn't mean they aren't individuating of theoretically distinct kinds. Of course there is also a kind that corresponds to the conjunction of these two properties, being both a thing with a certain etiology and with a certain disposition. What to name these kinds, I don't care. Suppose there is a kind that admits of "swamp"—spontaneously and accidentally generated—corkscrews by requiring just the disposition, and another that admits of broken corkscrews by requiring just the etiology, and yet another that admits of neither (requiring both the etiology and the disposition). But in no case is anything temporally or spatially extrinsic to the physical conditions strictly sufficient for the property or properties in question. It's just that the spatiotemporal dimensions of the physical conditions instantiating the disposition are spatially protracted, and the ones instantiating the etiology are temporally protracted. Those conditions instantiating a thing's "having" a certain etiology *and* disposition are protracted in both spatial and temporal dimensions.

I think the first grade of causal involvement is just the same as the third, so I'll discuss the third grade now, coming back to the second grade afterwards. Bennett argues that building relations can hold over an interval, and I agree. She says,

consider pretty much any ordinary nonfundamental thing, and any ordinary nonfundamental property it instantiates. Consider, again, your chair. Its solidity and ability to bear your weight is microbased in ongoing causal processes at the microlevel, not in properties

statically possessed by stationary particles. The fact that it exists arguably does as well. (p. 99)

The explanation is that

Atoms come together to form molecules as their electrons are attracted to and repelled by each other. All the little bits—not that it is really little *bits*—are constantly moving and causally interacting. When atoms bond it is not as though one sticks a hook into the other and stays that way. (p. 99)

And this is not a special case, nor is it particular to, say, composition.

Pretty much all the kinds of building that counted as canonical back in [Bennett's] Chapter 2 often hold in a temporally extended fashion, in virtue of the causal activity of the builders. Certainly composition, grounding, and microbasing do. (p. 99)

Indeed, all of the relations I introduced in Section 3.1 hold in a temporally extended fashion. Realization is causally infected no differently. The physical conditions instantiating the functional properties realized (partly) by first-order properties obtain either in virtue of a more broadly temporally extended state of affairs—an interval that includes causal interactions—or a more broadly spatially extended state of affairs. But even the latter is not momentary or temporally unextended. For example, the first-order properties that realize second-order properties are themselves not properties instantiated instantaneously, but by physical conditions extended through time. There would be no realization relation in a world stripped of causal interactions.

As for the second grade of causal involvement, Bennett introduces it by asking us to imagine

pulling ingredients from the pantry and baking a cake. In such cases, we start with some things, and use them to make something else. . . . The cake is made from the flour, sugar, eggs, and whatnot. As that sentence shows, we very naturally use various building locutions when we describe cases like these. Yet they involve causal processes that occur over time. . . . It is so natural to use building talk to describe these kind of diachronic causal processes. . . . (p. 86)

Bennett calls this “diagonal” building. Now, Bennett actually agrees with the first stage of a possible objection, that “such talk is made true by a combination of building relations that are not diachronic and diachronic [causal] relations that are not [vertical] building relations.” She just doesn’t think that this entails that a diagonal building relation between eggs and cake doesn’t exist. Bennett thinks an untenable appeal to parsimony would be required, without which the fact that such truth conditions can be given is otherwise “compatible with either of two quite different attitudes toward or views about diachronic building”:

One view is that there are no diachronic building relations, even though there are true tensed sentences about building. The other view [Bennett’s] is that there are diachronic building relations, *but they are not fundamental*. (p. 92)

Bennett says the latter, because she thinks building relations are themselves built (ch. 7). Now if diagonal building is built, note that it is built both causally and from synchronic dependence that obtains in virtue of causal facts (the cake’s atoms partly compose it over its interval, rather than instantaneously). But in that case, diagonal building can thus be distinguished from causation and from building relations tainted by the third grade of causal involvement, since, *ex hypothesi*, it is built from both. On the other hand, I’m not sure the second grade of causal involvement is clearly distinct from the third grade of causal involvement, for the same reason that the first grade of causal involvement isn’t: temporal properties are instantiated by conditions which are temporally protracted compared to the conditions to which they are attributed. To see this, note that Bennett characterizes the first stage of the objection to diagonal building as saying that our ordinary diagonal “made from” talk is “simply stating that things with temporal properties [egg proteins that came from eggs] stand in a purely vertical, synchronic building relation to something else [the cake]” (p. 87). Ignoring that the synchronic relation isn’t actually pure in this way, the point about a cake or its parts having the temporal property of being made from eggs highlights what I want to suggest about the time interval of the conditions sufficient for such a property. To put it simply, it takes eggs to make a cake that is made out of eggs. The property of being made out

of eggs which existed at t_1 that we might attribute to a region of spacetime, “cake”, at t_2 is a non-categorical property for which the latter region is not sufficient. It takes actual eggs to sufficiently instantiate such a property. The property of being made of proteins with certain dispositions (which might have been synthesized by clever chemists to sell to vegans), on the other hand, is a different property, protracted spatially and/or temporally elsewhere.

In any case, I agree that synchronic dependence relations involve causal relations in the sense that causal relations are parts of composed, constituents of constituted, or realizers of realized things. But what of it? Bennett acknowledges:

The claim that some building relations are partially defined in causal terms, or hold partially in virtue of various causal facts, does not entail that causation itself is a building relation. . . . So there is room for an opponent to agree with me that building is causally tainted in the second sense while denying that causation itself counts as a kind of building.

What such an opponent needs to do is provide a reason for ruling out causation proper that does not also rule out the causally tainted building relations that she does accept. This requires a bit of care. For example, it would be of no use to argue that causation is not a kind of building because causation is not synchronic—neither are the relations invoked at the second and third grades of causal involvement. Probably what such an opponent wants to do instead is deny that causation is a member of the class of relations that make for relative fundamentality, an issue that I have postponed discussing. (pp. 100-101)

Now, note that Bennett thinks relations involving causation by obtaining over an interval are thus not “synchronic.” I disagree, but this particular disagreement is merely terminological. Of course, we need some name for these relations. “Diachronic” has sometimes been used (Kirchhoff, 2015), but that is confusing, because “diachronic” is usually reserved for relations between causes and their effects (see Horgan, 1993; Kim, 1999, 2003; Leuridan, 2012; O’Connor and Wong, 2005; Pereboom, 2002; Rupert, 2010; Sober, 1999; Wilson, 2014, 2015). And Bennett uses “diachronic” to refer to causal, as well as diagonal building relations. “Non-causal” won’t work if it implies what both Bennett

and I reject—that these relations do not, in some sense, involve causation. I distinguish between causal and another subclass of dependence relations, and yet I do not claim that the latter is “non-causal,” precisely because I think what I’ve been calling synchronic dependence holds over intervals of time, and that causal interactions can be constitutive of entities. Calling synchronic dependence “non-causal” would obscure that. On the other hand, I do think it is a mistake to call the causal interactions that are constitutive of entities *causes* of those entities, because I take it that they do not occur before the entity they constitute, but they occur *as* the entity they constitute occurs, and I maintain that this is an important difference in explanatory practice, a point I’ll further defend shortly. So I’m going to stick with my usage of “synchronic,” if only because, as far as I can tell, while the literature is settled on how to use “diachronic,” it is not settled on Bennett’s use of synchronic to mean instantaneous as opposed to covarying over an interval, and the latter is at least consistent with some philosophical uses, and more in line with ordinary use. Ylikoski (2013) refers to constitution as “synchronous,” for example. What is synchronous needn’t be instantaneous. For an ordinary example, synchronized swimming is an activity and is not instantaneous at all, let alone essentially so. The movements of the swimmers covary not at a single time, but through time. By contrast, effects at a time vary in relation to causes at a previous time. The particles and the chair, if they are distinct, stand to one another more like synchronized swimmers stand to one another than they do like paradigmatic causes—such as stone-throwings—stand to an effect—such as a bottle breaking.

Terminological choices aside, there is room to agree that what I call synchronic dependence relations involve causal interactions, *and* let Bennett define building as the entire class of dependence relations, including causation, and yet still hold that there are two *more unified* subclasses of dependence relations: causation on one hand, and the rest, on the other. To see this possibility quite clearly, take Bennett’s own map of the conceptual terrain:

Let *A* be the class of purely noncausal vertical building relations.
 Let *B* be the class of causally tainted building relations—those that hold partially in virtue of the obtaining of some causal facts. Let *C* be a class only including causation itself. The claim that building is causally tainted in the first way is the claim that the union of *A*

and B and C is unified. The claim that building is causally tainted in the second way is the claim that the union of A and B is more natural than either A or B alone. (Indeed, a large part of the point of the second kind of causal taint is simply that class B exists.) (p. 100)

Granting all that, there is room to think, as I do, that B , and perhaps also the union of A and B , is clearly and usefully distinct from C . I have nothing more to say about what belongs in A (Bennett mentions set formation (p. 100)), and thus nothing more to say about the A – B class. The synchronic dependence relations I have focused on—realization, constitution, and composition—the ones primarily employed in the EM literature, all involve causation, and are thus members of B .

Recall that there were actually two flavours, shall we say, of the conclusion drawn by Bennett. The conclusion presented at the outset is that the usual distinction between causal and other building relations is less clear and useful than ordinarily assumed. Soundness here hinges on exactly what is typically assumed about that distinction. However, Bennett’s second grade of causal involvement appeals to *ordinary* diagonal building talk, such as “made from”. For philosophical examples, new mechanists clearly have a four-dimensional notion of constitution when they talk about the constitution of dynamic processes and systems (see Craver, 2014; Baumgartner and Casini, 2016; Gillett, 2013; Kirchhoff, 2015; Ylikoski, 2013). And as an example of building relations holding over an interval, Bennett cites the well-known four-dimensional view about persistence through time. She also cites van Inwagen (1995), who thinks some x ’s compose a y just in case their (temporally extended, of course) *activities* constitute a life. Thus polling both the everyday and philosophical notions of dependence, it’s not clear what ordinary assumptions are being made about the clarity and usefulness of the distinction between causation and other building relations, such that the distinction is less clear and useful than ordinarily assumed. The conclusion encountered later in the same chapter, has a stronger flavour. It is that the class of purely non-causal building relations fails to carve at the joints, because it excludes a whole bunch of ordinary ways of thinking about building. That might be true of the A class. But, for the reasons just given, it’s not clear that our ordinary distinction carves out just

the A class, and even if it did, that point would be moot as to the theoretic utility of the B class or the A – B union.

Of course, if there is *no* distinction that can be made between causal dependence and other dependence relations, then the stronger of Bennett’s conclusions is all the more warranted. But I maintain that there is a clear distinction between causation and the rest, *despite* their similarity, and despite causal involvement in the latter. Indeed, the fact that all of the putative synchronic dependence relations relevant to the EM debate involve causation goes to show that they are unified even further, because they are all *tainted in the same way*, while causation isn’t. But let me now defend the view that causation and synchronic dependence are distinct.

8.2 A Plurality of Building Relations

Even the above suggestion that any purely non-causal A class of relations is not useful says nothing about the clarity or usefulness of a distinct class of B relations, or the union of the B and A classes (if the latter has any members). The worthwhile point about causal involvement in synchronic dependence relations, or, as I shall now put it, the observation that causal interactions among parts or constituents can constitute or realize higher-level entities or higher-order properties, is meant to illuminate the existence of that class, and certainly does not show that *that* class is not clear or useful. Strangely, then, the only thing that would show that the B class of relations, the class I am interested in, is not clear or useful, would be if causal and constitutive species of dependence could not be distinguished at all. But surely the very existence of the B class depends on there being a distinction between it and the C class—the class that isolates causation.

However, something in the vicinity of the view that there is no useful distinction between causal dependence and constitutive dependence specifically, arises within the EM literature. In *The Extended Mind* collection edited by Menary (2010), Ross and Ladyman (2010)² defend the view that mature sciences make no use of such a distinction. They suggest the distinction

² Every reference to Ross and Ladyman in this chapter is to Ross and Ladyman (2010).

belongs to a scientifically naïve “metaphysics of domestication” that takes our ordinary spatial metaphors too seriously. They claim, “The causal–constitutive distinction thus is not drawn in physics” (p. 162). As they see it,

Chemistry explains recurring properties of water samples by reference to underlying properties of hydrogen and oxygen atoms and the way they combine to form molecules of H_2O . For example, the question of what makes water a solvent of common salt is answered, in broadly causal terms, by appealing to the chemical properties of the oxygen and hydrogen molecules. However, the metaphysician insists that the question of what constitutes the water is altogether different. This is held to concern what substance or matter makes up the water, and appeal is to be made only to the existence, rather than the causal powers, of the hydrogen and oxygen atoms. However, this special sort of explanation is foreign to sciences, especially as they mature. (p. 158)

According to any metaphysician I’m familiar with, the existence and the causal powers of hydrogen and oxygen atoms are not dichotomized this way. We’re then given a somewhat more plausible characterization of the view the authors oppose:

Domesticating metaphysicians seek to account for the world as “made of” myriad “little things”. . . . Their causal powers are usually understood as manifest in the effects they have on each other when they collide. Thus the causal structure of the world is imagined to be based on emergent or reducible consequences of reverberating networks of what we call “microbangings”—the types of ultimate causal relations that prevail among the basic types of little things, whatever exactly those turn out to be. (p. 159)

Of course, the interactions between the particles manifestly posited by contemporary physics are not just collisions. I’m not sure who they are, but *these* metaphysicians, unlike the metaphysician of H_2O encountered the page before, do seem to acknowledge the causal powers, and not just the existence, of what they consider to be constituents of the world. Still about these metaphysicians, Ross and Ladyman say,

This picture, familiar as it is, finds absolutely no corresponding image in contemporary fundamental physics. The types of particles which physical theory describes do not have spatiotemporal

boundaries in anything like what common sense takes for granted in conceptualizing everyday objects. . . . (p. 159)

Again, it is hard to judge whether or not contemporary physics corresponds to a view which is not itself clear. When Ross and Ladyman say *the* (singular) “metaphysical notion of constitution or composition is an abstraction that does not correspond to any general idea that figures nonmetaphorically in science,” since they do not further characterize that notion, it is hard to know whether their criticism sticks to *some* target. Adams and Aizawa are the only “metaphysicians” Ross and Ladyman cite, saying that their “discussions of constitution appear to be wholly based on naive objectification of everyday containment metaphors” (p. 161). Adams and Aizawa (2008) say,

Consider the expansion of a bimetallic strip in a thermostat. This process is causally linked to the motion of atoms of the air in the room the thermostat is in. The expansion of the strip does not, thereby, become a process that extends into the atoms of the air in the room. It is still restricted to the bimetallic strip in the thermostat. (p. 91)

Ross and Ladyman respond that such claims do not admit of “any possible translation into the terms of a physical theory; the claims are irreducibly metaphorical” (p. 161), whereas I don’t see Adams and Aizawa’s claim as requiring translation: they are already speaking in the same terms as much of our physical theorizing. Now, if Adams and Aizawa would deny Ross and Ladyman’s claim that, “even a fully general theory of cognition—as information processing by relatively autonomous goal-driven systems—need incorporate no single overarching account of limits on the boundaries of cognitive systems” (p. 156), then it would appear that they take themselves to be asking the (naïve) question of “where cognition takes place against a fixed background arrangement of objects” (p. 155). Ross and Ladyman “don’t think there is any such matter of fact, as a special case of there being no fact about where minds are located at all. To talk about the location of the mind is simply to resort to metaphor” (p. 156). Adams and Aizawa, however, don’t take themselves to be answering any such question about “fixed background arrangements of objects” when they talk about the location of cognitive processes, as a generalization

about the *locations* of cognitive processes, simply because whatever it is that is not “fixed,” if it is physical, needn’t be unlocated. To make claims about *the* (singular) location of *the* (singular) mind may be to make a certain kind of factual mistake, but it has nothing to do with metaphor.

But suppose *someone* thinks that water is constituted by the existence of H_2O at an instant, *rather* than by the causal powers of hydrogen and oxygen. Or suppose someone thinks that particle *collisions* generally explain macro-level phenomena. Even so, a few bad explanations aren’t enough to motivate the general discontinuity Ross and Ladyman see between what they variously call common sense, folk wisdom (p. 157), natural language, everyday objects, containment metaphors, and metaphysics on the one hand, and mature science on the other. Moreover, the uselessness of the notion of constitution, making-up, and containment does not straightforwardly follow from misuses of those notions. But that seems to be what Ross and Ladyman think follows from the fact of the bad explanations they have in mind. After dismissing microbangings, they claim,

Indeed, it is no longer helpful to conceive of either the world, or particular systems of the world that we study in partial isolation, as “made of” anything at all. The basic structure of reality as described by fundamental physics can only be accurately rendered in mathematics; to attempt to translate it into the terms of natural language is at once to resort to metaphor, which no amount of elaboration in a metaphysical tone of voice can render plausible as a literal account.

The causal–constitutive distinction is part of the metaphysics of domestication. (p. 160)

It simply doesn’t follow from the failed explanatory power of “microbangings” that the world is not made of anything at all, or that particles have no spatiotemporal boundaries. Even the mathematics which, in Ross and Ladyman’s view, literally describes fundamental physical processes (p. 163), employs variables for time and spatial dimensions in terms of which spin, trajectory, velocity, acceleration can be rendered.

At another point, Ross and Ladyman seem to take the uncontainedness of phenomena and the dissolution of the causal–constitutive distinction to follow

from the very same point we saw Bennett make above, that, in my words, causal interactions can be constitutive of entities. Ross and Ladyman say,

water is composed by oxygen and hydrogen in various polymeric forms, such as $(\text{H}_2\text{O})_2$, $(\text{H}_2\text{O})_3$, and so on, that are constantly forming, dissipating, and reforming over short time periods in such a way as to give rise to the familiar properties of the macroscopic kind water. . . . The causal–constitutive distinction dissolves because the kind water is an emergent feature of a complex dynamical system. It makes no sense to imagine it having its familiar properties synchronically. Rather, the water’s wetness, conductivity, and so on all arise because of equilibria in the dynamics of processes happening over short but nonnegligible time scales at the atomic scale. (p. 160)

Indeed, except as an abstraction, it makes little sense to imagine water having its familiar properties at an instant, or “synchronically” so defined. That is not how I’ve been using the term, of course. That aside, the reasoning about the causal–constitutive distinction doesn’t go through. I deny any premise to the effect that if the properties of water arise from a complex dynamical system, then the causal–constitutive distinction dissolves. I affirm something more like the opposite, that if anything is a property of a complex dynamical system, then those causal dynamics are constitutive of that property. Imagine Smithers, the boss at a nuclear plant. Smithers instructs an operator, Homer, saying, “if the temperature rises above this threshold indicated on the dial, then push this red button to stop the meltdown.” Imagine that the temperature does indeed rise, and Homer, thinking he has understood the instructions, pushes the button, and the meltdown proceeds anyway. At the inquiry, Smithers says something like the following. My instruction in the form of a conditional statement expressed a dependence which is neither causal nor constitutive. As it happens, the red button actually needed to be pushed prior to the temperature rising above that threshold, but as a student of mature science, I’m unpracticed at using the naïve containment metaphors enshrined in tensed speech about the interdependence of phenomena. But that is as it should be. It’s up to everyone else to figure out the temporal relation between the relata of dependence talk or between the antecedent and consequent of the conditionals that express such dependence.

Of course, Smithers is a straw man. But the point of knocking him down is to exemplify the way in which the causal–constitutive distinction tracks the difference between the dependence I have on my parents’ having met, and the dependence I have on my kidneys. Knowing how something came about versus how something works is the difference between being poised to engineer preventions as opposed to cures, and not because these are “folk” or otherwise unscientific entities. In fact, Ross and Ladyman help themselves to alternative but prototypically “vertical” or non-causal turns of phrase such as *giving rise to* or *arising from*. And they do so when claiming that, “Chemistry explains recurring properties of water samples by reference to *underlying* properties of hydrogen and oxygen atoms and the way they *combine to form* molecules of H_2O ” (p. 158, my emphases). It would be an obvious (in both the ordinary and scientific sense) mistake to replace these synchronic dependence locutions with talk of water being an effect of these dynamics as long as there’s a difference between the way water “emerges” from—or just is—these dynamics, and the way H_2O results subsequently from hydrocarbons that have undergone oxidative reactions. It is perhaps confusing, but unimportant, that the former sort of emergence is sometimes referred to in terms of synchronic “effects.” That is a terminological issue I set aside except to note that on the majority of uses, causes precede effects. Either way, we need ways of referring to those things which determine something future to them, and those which underlie something contemporaneous with them. The reasoning is simple. I take it that we in fact deal successfully with higher-level entities, properties, and kinds, and that we need to refer to them to do so. We do so both scientifically and “ordinarily.”³ I take it that these higher-level things, *A*, often come apart, and can be explained in terms of lower-level states of affairs, *B*—entities, properties, kinds, and interactions thereof. I take it that both *A* and *B* have also been determined or otherwise brought about by past states of affairs, *C*. And I

³ I will not further defend herein what I see as the continuity between “folk” and “scientific” knowledge. Suffice it to say that I think scientific vocabularies and mathematics extend natural language and scientific practices codify otherwise ubiquitous cognitive processes of hypothesis-making and belief falsification. I thus imagine the mostly-shared project of growing human knowledge to be best served by inclusive dissemination of scientific history and of contemporary research projects.

take it that for any given A , the B states of affairs are distinct from the C states of affairs. And if they are distinct, then it matters, in everyday and scientific explanatory practice, whether the red button is among the C states of affairs or the B states of affairs. And if so, then it can matter to employ dependence terminology which does not obscure this difference, depending on what one's audience already knows. The most self-contained (pardon the pun), explanations and hypotheses employ a clear distinction.

The lesson we should draw is that the difference between whether a given property, phenomenon, event, object, process or state is a cause or a constituent of some A depends on what A is. The question is thus relative to the explanatory frame, but not insignificant. Moreover, the difference is not between what sorts of relata can be determinative of other sorts of relata, but whether the relata are contemporaneous. The question of whether they are contemporaneous, of course, involves asking just what is sufficient to instantiate the property to be explained, or the properties individuating the entity to be explained. And, as is by now well-rehearsed, whenever properties look non-categorical to the entity to which they are being attributed, we should expect to suck up more of the world in a complete explanation of those properties. Finally, the fact that synchronic dependence relations are tainted by some degree of causal involvement means that it won't do for the internalist to simply point out that the brain is in causal interaction with its surrounding in order to show that externalists commit a coupling–constitution fallacy. The internalist must either show that causal interactions never constitute anything, or show that most externalist explananda occur *after* the putative constituent on which they depend. Either that, or provide the yet-to-be specified criteria of the constitutive over and above synchronic dependence, an issue to which I shall return in Chapter 10. Of course, such a criterion, as I pointed out in Chapter 2, would make the EM theorist guilty of something more like a fallacy of composition than a coupling–constitution fallacy.

8.3 Parsimony Versus Pluralism

Despite their similarities, causal and synchronic dependence relations are distinct. But the question remains whether synchronic dependence relations are distinct. Before turning to that question in the next chapter, first let me say more about what it means to call synchronic dependence relations identical, and why it matters. Recall from Section 3.2 that Bennett (2017) thinks a class should answer to some ordinary concept and be theoretically useful. Aside from whether or not I'm right to suggest that dependence alone captures the class of building relations, let's grant that building construed either way answers to our ordinary concept of what it is for one thing to account for another, and is theoretically useful in some way. Bennett argues that building can be leveraged in a deflationary account of relative fundamentality (ch. 6), and that the shared features of causation and synchronic forms of building make for instructive analogies between the two (ch. 4). Still, a more detailed taxonomy could be even more theoretically useful. For example, suppose building relations can be further classified as causal and synchronic relations. Then it would be open to us to ask how many members there are in each of those classes. I will defend a restricted version of the view that synchronic dependence relations are identical. I claim that at least composition, constitution, and realization have the same features and obtain between the same relata. The view, as a result, is deflationary about the differences between these relations insofar as it does not take them to be differently ontologically committed in contrast to one another. Discovering that synchronic relations are identical would make for further theoretic utility. It would cut short the sort of dispute we saw in Section 5.2 about the prevalence of multiple realizability, unite philosophical discussions as they have fractured into their preferred terminologies and versions of all-too-similar puzzles, and preempt the possibility of fracturing the EM debate in a similar way.

By contrast, the view I reject is that composition, constitution, and realization essentially describe different states of affairs, or that when used to express a dependence between the same relata, these differing dependence locutions would describe different worlds. This view is one according to which it is *mistaken* to say that some wood realizes a table; *mistaken* to say that a

brain constitutes a mind; *mistaken* to say that a piece of marble composes a statue. Or if an entity, M (a mental state), is both composed by some things and realized by something, M will not stand in these relations to the same entities; M may be composed of a (neurons, perhaps) and realized by P (the property of being such that it stands in certain input–output relations with other states), and a and P are distinct ontological commitments. *Synchronic Dependence Relation Pluralism* can thus be characterized as the conjunction of distinctness theses about as many relations as there are that are in fact distinct from one another:

[Synchronic Dependence Relation Pluralism] If y is composed of xs , y is neither realized nor constituted by anything identical with the xs , and the xs , if they constitute or realize anything, do not realize or constitute anything identical with y . Likewise, if M is realized by P , M is neither composed of nor constituted by anything identical with P , and P , if it composes or constitutes anything, does not compose or constitute anything identical with M . Likewise, if A is constituted by B , A is neither composed of nor realized by anything identical with B , and B , if it composes or realizes anything at all, does not compose or realize anything identical with A .⁴

While I don't know of any rigorous defense of *Synchronic Dependence Relation Pluralism* about composition, constitution, and realization, I know of no defense of the thesis that they are identical, and almost certainly not so detailed as this undertaking. Moreover, *Synchronic Dependence Relation Pluralism* about those relations is a rather natural position—almost, I suspect, the *de facto* one. I suspect many would be inclined on intuitive grounds to think I'm mistaken to lump together those relations I do. One interlocutor's reaction was to say that it *sounds* “messed up” to say that some table is realized by some wood. Now, one can run afoul of best practice for separate normative or epistemic reasons—as Bennett (p. 61) points out, explanations can be better or worse on the basis of “Gricean” considerations such as what the audience of

⁴ Like others working in this area, I can be understood as using singular terms for convenience, but P and B could stand for a plurality of realizers and constituents respectively.

an explanation already knows. One can also be wrong for purely grammatical reasons, the way it is mistaken to say, “ya done good,” in English, or “le chaise,” in French. Indeed, I do think synchronic dependence locutions have tended to vary somewhat in their uses. But these practices have not settled into even a rough convention such that they can be trusted to do any explanatory work in the epistemic sense noted above, and I don’t take myself to be in the business of stipulating how these terms might be used more precisely. On the other hand, more importantly, the general picture of parsimony I have in mind may be permissive about terminology while being deflationary about the ontological commitments of dependence claims made with different synchronic dependence locutions. Note, then, that as I refer to *Synchronic Dependence Relation Pluralism*, I will assume, but omit, the “ontological” qualification. It is nevertheless important to keep in mind, because the denial of a plurality of ontological commitments made in various synchronic dependence terms is compatible with a permissive stance toward the differing roles the terminology might play in certain yet-to-be-settled epistemic dimensions.

Despite a dearth of explicit adherents, *Synchronic Dependence Relation Pluralism* is a possible position in the logical space. For one, Bennett does not endorse monism about her class of building relations. Now, Bennett’s pluralist view about building, plausible as it is given the inclusion of causation, is not *explicitly* pluralistic about composition, constitution, and realization. On the other hand, her reasons to reject the identity of building relations are generated by issues about synchronic dependence relations, specifically, set formation, composition, and grounding. So she is certainly a pluralist about any class of relations that includes those relations, even if there was such a class that did not include causation. But I do not quite attribute to her the view that I reject, since she says nothing explicitly for or against the identity of the three relations my positive claim is about.

That *Synchronic Dependence Relation Pluralism* is a possible position in the logical space is further illustrated by Jessica Wilson’s (2014) and Koslicki’s (2015) rejections of the view that instances of *grounding* are identical. Of course, although these latter positions are explicitly pluralistic about grounding, it is not clear as to whether they are the target view. Insofar as Wilson (2014)

and Koslicki (2015) each reject the view that there is a single grounding relation encompassing composition, constitution, realization, set formation, the determinate–determinable relation and the species–genus relation, it is unclear whether either philosopher rejects the view that there could be a single synchronic dependence relation encompassing some subset of those relations.

Similarly, van Inwagen (1995) has rejected, if somewhat passingly, the view that instances of the parthood relation are identical. But I draw the opposite lesson from the same example. The deflationary move I want to make with synchronic dependence relations generally is even more obviously plausible among parthood relations insofar as we certainly needn’t multiply the terms for “part.” Since one can apparently say true things about all sorts of relata using part–whole terminology, it is not clear that the differences in the *relation* borne between all sorts of wholes and their parts is in every case something other than the differences in the parts.

Finally, Polger and Shapiro (2016) can be understood as holding the view I reject. We saw their view of realization in Section 5.2. Their way of denying that realization obtains as often as we might think suggests that they are at least implicitly committed to *Synchronic Dependence Relation Pluralism*. They say,

Gillett’s (2002, 2003) account [of realization] lumps together several distinct dependence relations. If an account of realization is not discriminate, then learning that *x* is realized by *y* would be uninformative. It would just tell us what we already knew—viz., that *x* depends on something else. (ch. 2, n. 14)

The distinct dependence relations Polger and Shapiro (2016) have in mind are composition and constitution. The idea seems to be that if the composition relation holds between two relata, the realization relation does not, and vice versa. On their account, this is not because of some linguistic rules governing which relata can enter into each relation, and probably not even because of considerations as to what would or would not be informative in one *epistemic* context or another. Rather, on their view, it’s mistaken to say that aluminum, for example, realizes a corkscrew when in fact it composes the realizing mechanisms of that corkscrew, because of the way the world is, or

perhaps because of the way nature is jointed or the way kinds are related, such that each dependence locution serves aptly and exclusively to describe differing states of affairs. The differences are metaphysically deep. For example, the realizing mechanisms of the corkscrew may be reducible to the materials that compose them, while the corkscrew is not reducible to the particular realizing mechanisms, or so it seems, when the realization base is narrowly construed according to our circumstantialist intuitions.

Again, the point here is that *Synchronic Dependence Relation Pluralism* is a possible position in the logical space. The position may indeed be the default view; it is *prima facie* plausible, and likely defensible. So if I'm right that it's mistaken, it is a worthy target. The monism I recommend would also unite philosophical discussions as they have fractured into their preferred terminologies, while pre-empting the possibility of fracturing the EM debate in a similar way.

Wrapping Up

I'm in the midst of arguing that it doesn't matter whether claims about externalism are made in terms of constitution, composition, realization, or identity. In this chapter I have begun to explain that view. In Section 8.1, I began by discussing the way in which building relations are causally tainted. But in Section 8.2, I argued that causal taint does not render causation and synchronic dependence inseparable as subclasses of building or dependence. The question of whether something is a cause or constituent of another thing is indeed quite sensible. In Section 8.3, I discussed what it means for relations to be identical, and provide a general defense of pursuing the question of whether synchronic dependence relations are identical. In the next chapter, I discuss Bennett's main reasons for not embracing monism about building relations. I argue that those reasons do not count against the identity of specifically synchronic dependence relations; in fact, some of the same considerations can be turned into an argument *for* the identity of synchronic building relations.

Chapter 9

The Identity of Synchronic Dependence Relations

There are quite obvious similarities among the synchronic dependence relations set out in Section 3.1. I've argued further in Chapter 4 that very similar puzzles arise when one takes higher-level things to be distinct from their lower-level physical bases, no matter whether these are thought to be realizers, constituents, or parts. But as I argued in Chapter 6, the irreducibility of higher-level kinds to physics is premised on the *Unity Condition on Kind Identity*, which is a mistaken requirement that each higher-level kind be identical to a lower-level one. As I argued in Chapter 6, multiple realizability is no more a problem for a kind's being physical than variation. As I argued in Chapter 7, the distinctness of higher-level individuals from their lower-level constituents is otherwise premised on an equivocation between strict and sloppy notions of sufficiency. The failure of arguments for antireductionism suggest that complete synchronic dependence is identity. Further, I've argued in Chapter 8 that synchronic dependence relations, though causally "tainted," are tainted in the same way, and that the distinction between causal and synchronic dependence is nonetheless intelligible and explanatorily ineliminable. In this chapter, I argue that the synchronic dependence relations most encountered in the philosophy of mind and the EM debate (constitution, composition, and realization) are identical with one another.

Interestingly, the opposite thesis seems to be commonly assumed but not explicitly or rigorously defended. The flip side of the same situation is that

whether or not my view is entirely unique, my argument for it is. So there is a significant unexplored feature of the philosophical landscape—a sort of unclimbed peak. The reason to explore these issues is that, if I’m right, then the literature on synchronic dependence relations is needlessly and detrimentally fractured and extensive, and, in contrast, the EM debate can remain unified in an important way. As much as it would be clearer if EM theorists stuck to just one way of referring to the synchronic dependence they claim often holds between cognitive processes or mental states and extraneural entities normally considered to be part of the environment, this apparent waffling is merely apparent. It is not a deep problem in the sense that whether cognition is constituted by extraneural entities or whether it is realized by them are not different questions. They are one and the same question. While I think Clark and others are actually safe not to worry much about their chosen dependence locutions, I will explicitly argue for the identity of these relations. That enables inferences from premises stated in one synchronic dependence locution to conclusions stated in another. Indeed, the corollary view that partial constitution is partial identity is crucial in my inference from our partial constitution by technology to our partial identity with technology.

First, in Section 9.1, I’ll describe two theses Bennett (2017) rejects about building, and contrast these with the thesis I endorse about synchronic dependence relations. Then, in Sections 9.2 and 9.3, I’m going to explore the two reasons Bennett thinks that although building relations, including causation, are unified, they are not versions of a single relation that is in some sense more natural or fundamental than the rest. I argue that these considerations don’t count against the possibility that synchronic dependence relations are identical. In Section 9.4, I consider two other possible ways in which synchronic dependence relations might differ. Although Bennett (2017) seems not to make much of these differences, and Bennett (2011) rejected their significance, I take one of them seriously: the idea that synchronic dependence relations differ in virtue of the relata they take. Indeed, it is surprising to claim that composition, constitution and realization have the same features and are co-extensive insofar as realization is prototypically a one–one relation between properties, constitution is a one–one relation between entities, and composition is a many–one

relation between parts and wholes. Actually, none of these terms are restricted to those uses. But even so, I argue that these differences among relata are superficial and fail to underwrite any deep difference between the sorts of synchronic dependence that obtain between them.

9.1 Characterizing Dependence Relation Monism

Bennett (2017) cashes out the thought that “there is a most general building relation of which the others are versions” (p. 22) in two ways. First Bennett characterizes what she calls “generalist existence monism” (p. 23)—which I have been and will continue to simply call “monism”—about a class of relations. Since I advocate monism about a different class, I’ll paraphrase Bennett’s (p. 23) formulation in order to abstract away from the case of building:

[**Monism**] Some single relation is really the only relation there is.

The second way to think “there is a most general building relation of which the others are versions” (p. 22), is a view about the class which Bennett calls “generalist priority monism”—but which I’ll just call “generalism.” Paraphrasing Bennett’s (p. 23) formulation again, generalism about a class of relations is the view that:

[**Generalism**] There is a most general relation that is more natural or fundamental than the specific class members.

For example, think of age-relatedness, and the more specific relations, “a minute older than” and “a century older than.” These two relations cannot hold between the same specific relata, or, again, these two expressions cannot be true of the same individuals (at least, not within the same frame of reference). But if individual *a* is a minute older than individual *b*, it is also true that *a* is older than *b*, and if individual *c* is a century older than individual *d*, it is also true that *c* is older than *d*. The truth of the same more general expression is ontologically innocent in respect of any of its species; we get that truth “for free” from either state of affairs. So, for example, we might

think that causation and composition are species of a more general relation, dependence, such that whenever a either composes or causes b (but not both), b depends on a . Again, if composition is a species of dependence, it would be a setback to theoretical utility if we thought we were disagreeing were one of us to say that a certain table partly depends on its legs and another were to say that it is partly composed of legs. Obviously, we'll accept some form of generalism about age-relatedness; the point of using the clear examples just given is to demonstrate what is at issue, even while the relation among synchronic dependence relations is not so clear.

Bennett (2011) defended the possibility that *Generalism* might be true of building relations. She provided only a hint of how one might defend *Monism* about building, which we'll see in Section 9.4. However, Bennett (2017) thinks *Monism*, and even the weaker *Generalism*, are likely false of building relations. Again, of course, the conclusion I endorse—that synchronic dependence relations are identical—is not the one Bennett rejects. Bennett rejects *Monism* and *Generalism* about *building*, including causation. I reject those theses about building for some of the same reasons as Bennett does, but neither mine nor Bennett's reasons are reasons to reject *Monism* about synchronic dependence. I will nonetheless discuss Bennett's reasons to reject *Monism* and *Generalism* about building, since the same reasons might be turned into an argument against the view I do endorse. That is, if one rejects *Monism* and *Generalism* about a class of things on the basis of deep differences between the members, one would reject their identity with one another for the same reasons. My main purpose in what follows, then, is to argue that there are no deep differences between synchronic dependence relations.

I should also note at this point that neither Bennett's characterization of *Monism* nor of *Generalism* is quite the claim I defend about synchronic dependence. Her official formulation of building relation monism says that, "There is *really* only one highly general building relation that *somehow underwrites* all the more specific talk" (p. 23, my emphases). I don't feel the need to posit anything further or more real than, for example, constituents or parts to do this underwriting. Neither would Bennett, presumably; she argues (2017, ch. 7) that building relations are themselves built by the relatum in the first place of the

relation. But that’s precisely why I’m unwilling to say some general dependence relation is more natural or fundamental than the specific. And I’m ambivalent in such contexts about what there “really is” (2017, pp. 13, 23, 25), as a consequence of the permissive form of physicalism I endorse. That physicalism is at once “reductive” in the sense that it takes higher-level things to be nothing over and above their parts, constituents, or realizers. But because they are nothing over and above finer-grained descriptions of the same states of affairs, coarse-grained things are *welcome* for pragmatic reasons in a perfectly scientific ontology, without any necessary sacrifice to explanatory or predictive accuracy. Again, we needn’t be eliminativists about higher-level things. Likewise for sets, classes, kinds, properties, and relations. So the claim that synchronic dependence relations are identical is somewhat different than Bennett’s official formulation of building relation monism. There really is composition and there really is constitution and there really is realization—indeed they’re the same and these terms are co-extensive.

9.2 An Objection from Failed Asymmetry

In this section and Section 9.3, I will resist the two objections to *Monism* and *Generalism* about building that Bennett canvases. Despite the differences between *Monism* about *building* relations and the thesis that synchronic dependence relations are identical, the arguments we’ll see against the former are relevant to the latter because those arguments are made in terms of synchronic dependence relations. Specifically, these objections involve composition, set formation, and grounding. As in Section 3.3.2, I feel compelled to face the objections in the terms in which they have been made, even though I have restricted myself to discussing those dependence relations most invoked in the EM literature. I am compelled for two related reasons. The first is that I take it to be an open question whether my arguments have any bearing on set formation and grounding, and I suspect they do. Secondly, if they do, considerations in those areas would have analogues with respect to composition, constitution and realization. Of course, it’s also possible that this is precisely where the analogy might break down. For example, even if set formation and composition

aren't versions of one relation, that wouldn't tell against the identity of the three relations my positive claim is about. One *might* just dismiss the examples to follow as irrelevant to my conclusion, given the tentatively more limited scope of my claims. On the other hand, I'm not convinced the examples work in their own right, and if they don't, I don't need to retreat to the disanalogy of set formation or grounding with the relations I have focused on, and I can continue to think that my positive claims might extend to further relations. So let's hear the arguments.

One objection to *Monism* and *Generalism* Bennett offers is motivated by the worry that two building relations might hold in opposite directions between the same two distinct relata. The example Bennett is concerned with is Schaffer's (2010) "priority monism":

[Schaffer] thinks that the entirety of the universe grounds the existence and properties of, say, this cup. The cup is built out of the One, rather than the other way around. Yet it is also the case that the cup is part of the entirety of the universe—that the One is composed of the cup, the desk on which it sits, my car, and everything else in the universe. So the cup (along with a lot of other things) bears the composition relation to the One, but the One bears the grounding relation to the cup and its properties. (2017, p. 27)

Then, the concern is that, "any highly abstract, very general relation that holds whenever a more specific building relation [holds] will fail to be asymmetric... if it is possible for two different specific building relations to hold in opposite directions between the same entities" (2017, p. 26). She illustrates this with respect to disjunctions:

every disjunction of any asymmetric relation and its (asymmetric) converse fails to be asymmetric. If *a* is older-or-younger than *b*, *b* is unsurprisingly also older-or-younger than *a*. In short: because disjunctions of asymmetric relations can fail to be asymmetric, and all building relations are asymmetric, disjunctions of building relations can fail to be building relations. (2017, p. 27)

Indeed, suppose that *older than* and *younger than* are versions of a common relation "different in age to." If *a* is older than *b* (and not vice versa), then *b* is

younger than *a*, (and not vice versa), and *a* is different in age to *b*, and, crucially, *vice versa*. So the relation of which the asymmetric relations are versions is itself symmetric. Note that even being *asymmetrically related to* fails to be asymmetric: if *a* is asymmetrically related to *b*, then *b* is asymmetrically related to *a*. Now, supposing that *Generalism* about building is to be understood as committed to the general's sharing all of the features of the specific, then such a special (as I'll call it) generalism is false in these cases. So be it. However, notice that the same line of reasoning above should be taken to show that such a special generalism is false of the relations "older than" and "younger than." That's an odd result which is not entirely consistent with the broader idea of things' being varieties of some general thing.¹

Now, the examples of building relations holding in opposite directions between the same two distinct relata could be resisted just because they "involve substantive and controversial claims that are endorsed by only a handful of people" (2017, p. 28). But Bennett does not dismiss the examples in this way,

¹ I leave this point aside, except to note the following. Bennett thinks that *Generalism* is plausible of the unity of species of a common genus, whereby some features are fundamental to all the species of the genus (2017, p. 24). So it's inconsistent to hold that *Generalism* is true in those cases, but that the reasoning to failed *Generalism* on the basis of the symmetry of the general relation, "goes through regardless of whether the general relation is taken to be a determinable, a genus, or even a mere disjunction" (2017, p. 26). After all,

Determinables need not share the formal properties of their determinates;
genuses need not share the formal properties of their species; disjunctions need
not share the formal properties of their disjuncts. (2017, p. 26)

The idea that *Generalism* is plausible of genera must rest on a somewhat different gloss on "generalism" than the idea that generalism is false of building. I've marked one of these generalisms "special generalism." The very example of asymmetric relations is supposed to show both that the disjuncts or species cannot be "versions of a more general relation," in the special sense, while presupposing that they *are* versions of a relation, in some other sense. So the point can only be that whatever *Generalism* is plausible in cases of species-genus unity ought not involve the claim that a genus has all the features of its species. So when Bennett says, "if all building relations are asymmetric, and the general relation that holds whenever a more specific one does is not asymmetric, then it is not a building relation," on her definition of building, I can agree, but deny that this is a problem. Again, that special sort of generalism is also false of younger-than and older-than, which are quite obviously "species" or "varieties," in *some* sense, of a common relation, if they are not the same relation.

and nor do I. Bennett also admits that,

in order for these examples to tell against generalism [and *Monism*] about building, it has to not only be claimed that the relevant relations hold in the relevant directions, but also that they are both building relations. So it may be possible to avoid the cases by denying that one of the relevant relations really counts as a version of building. (2017, p. 28)

For the sake of argument, I have been supposing that Schaffer's grounding and composition do share formal features. I think they are both synchronic dependence relations expressible in true counterfactuals. However, the asymmetric relation in the cup's partly composing the One is not necessitating (only full builders necessitate what they build). As I pointed out in Section 3.3, when building is partial, it is obviously asymmetric, but not necessitating. But when building is complete, it is necessitating, but not obviously asymmetric. For instance, if x fully composes y , then if not x , then not y , is true. But it is also true that if not y , then not x . By contrast, consider partial dependence. Take the brain without which I would not exist. It is not necessary that if I do not exist, that brain does not. It might be pickled. But take *all* those conditions without which I would not exist. Now it is true that if I did not exist, necessarily, not all those conditions would exist. This highlights the real reason to resist the example. As Bennett says,

in neither case is it quite true that y is B_1 -built from x and x is B_2 -built from y . Rather, it is only true that y is B_1 -built from x and x is *partially* B_2 -built from y . In the Schaffer case, for example, the One grounds the existence and nature of the cup, while the cup *plus everything else in the universe* composes the One. The One is only partially composed of the cup. (2017, p. 28)

On the other hand, since everything fully composes the One, and the One fully grounds everything, we still have apparently asymmetric, and necessitating relations that would be versions of a single or general relation which is symmetric. But—and this is the crux of the matter—the symmetry of *that* general relation is the right result. *If* we want to say two things mutually fully build each other, we *should* say that any general relation of which both building relations are species

is symmetric. That's the right thing to say because if everything composes the One, and the One grounds everything, we should also say that everything *is* the One, and the One *is* everything. Full synchronic building, as I argued in Chapter 7, is identity. In Section 3.1.5, I pointed out that the identity relation supports counterfactuals that make identity look like a sort of dependence: if a bucket of water is a bucket of H_2O , then delete that H_2O and you will have deleted that water, and vice versa. As for fundamentality, we should say the water and the H_2O are equifundamental. The equifundamentality just follows from the indiscernibility of identicals, and the symmetric dependence is just a trivial self-dependence. Whether to call it dependence is much like the question of whether to call the general relation of which two asymmetric building relations are species or versions, "Building". I recommend that we do, simply to highlight the feature shared by partial and complete dependence relations—they make appropriately fragilized counterfactuals true. Further, since nothing can be part of a whole that does not exist, wherever there is partial dependence, there is complete dependence. Or, to put the same thing another way, if something partly depends on another thing, then the former also partly depends on more.

Still, Bennett doesn't want to call such a relation "Building" because a symmetric identical or general relation would lack the theoretical utility of indicating relative fundamentality. Recall from Section 3.2.1 that Bennett takes building to explain relative fundamentality; the "input" relatum of all building relations is more fundamental than the "output" relatum, and that's all there is to fundamentality. However, notice that if it could both be true that the cup is more fundamental than the One, and that the One is somehow more fundamental than the cup, these claims cannot involve the same sort of fundamentality. Bennett is aware of this, saying that—"relative fundamentality claims are, in the first instance, indexed to particular building relations" (2017, p. 163). Thus she says, "Strictly speaking, there is more than one kind of relative fundamentality" (2017, p. 163). So the sense in which the One is more fundamental than the cup is different than the sense in which the cup is more fundamental than the One. But note, too, that special generalism about relative fundamentality, as indexed to such a variety of building relations, is

suspect just in case special generalism about her class of building relations is suspect. That is, a general *more fundamental than* relation will fail to be asymmetric if building relations can hold in two directions between the same individuals. There can be no general relation of fundamentality that *a* bears to *b* by building *b*, if *b* ever builds *a*. On the other hand, the very relativity, we might say, is at odds with the notion of fundamentality. Or, it is a very deflationary notion of fundamentality if all there is to some *a*'s being more fundamental than *b* is its building₁ *b*, and yet *b* can build₂ *a*. If building relations can hold in opposite directions between the same two distinct relata, the result that the general relation of which they are versions would not share the formal features of its species means that building relations are unable to index any notion of fundamentality that possesses any more unity or theoretical utility than the genus of building relations itself. After all, notice that relative fundamentality is to be understood in terms of building, and building is always asymmetric because relative fundamentality is asymmetric.

So the possibility of distinct building relations holding in opposite directions between two relata does not tell against their being species of a general relation. But would it preclude their being the same relation? That is, if grounding and composition are both species of synchronic dependence, and hold in opposite directions between the same two distinct relata, then it looks, at first blush, that there is more than one relation here. However, the problem for being a pluralist about the relations between the cup-and-everything-else (that is, everything) and the One is that they always go together, they are sufficient for and necessitate one another, and that makes them very hard to tell apart. And if they are identical, then everything and the One do all the same building as well as stand in all the same relations. On the other hand, if we are to insist on the distinction between the One's grounding everything and the cup's composing the One, then we must insist on the distinction between everything and the One. But if we do, we have a grounding problem. What makes the difference between everything and the One? Surely everything withholds nothing from the One, and there's nothing more to the One than everything. Moreover, we now have an unavoidable choice that looks difficult to motivate either way.

To see this, suppose the cup depends on the One, by being grounded in it. The cup also depends on what we usually think of as the cup's parts and interrelations thereof—a hunk of material impermeable to water and shaped roughly in a hollow cylinder enclosed but on one end. Now, because the cup is part of the One, the One seems to help compose itself by grounding its own parts, and everything seems to help ground itself by composing its own grounds. Now, Bennett argues (ch. 7), successfully in my view, that building relations are themselves built by the builders. Bennett's first defends "anti-primitivism" about building—the view that building relations are themselves built. She then defends "upwards" anti-primitivism, which is a view about *what* builds building relations. She thinks that if *a* builds *b*, then *a* also builds the fact that *a* builds *b*. So, what builds the fact that the hunk and the rest compose the One? Bennett wants to say, the hunk and the rest. What builds the fact that the hunk composes the cup? Upwards anti-primitivism says, the hunk alone. But what builds the fact that the One grounds the cup? I suppose that if building obtains here, which it does *ex hypothesi* to launch the objection in the first place, the One is relatively more fundamental than what it grounds and thus it alone builds that grounding relation. At a first pass, then, it looks like we need the hunk and the rest, as well as the One, to build the relevant building facts. But that seems to multiply conditions unparsimoniously, since the hunk and the rest fully compose the One and thus necessitate it, and the One fully grounds the hunk and the rest, thus necessitating them. Suppose, then, that we prefer to say that the hunk and the rest do all the building work: they compose the One, they build the fact that they compose the One, and in building all that, they build the fact that the One grounds the cup, and the fact that the hunk composes the cup. But the choice of basic building blocks is actually not so clear. That's because the claim that *a* builds the fact that *a* builds *b* is clearly a claim about *full* building. What then, builds the asymmetric but non-necessitating relation that obtains between mere partial builders and what they help build? For example, suppose some *xxs* compose *y*. What builds the fact that one of the *xs* partly builds *y*? Obviously, *x* *partly* builds that fact, but not fully. The full builders of that fact are actually the full builders of *y*. In other words, the fact that *x* builds *y* is not fully

built—does not obtain—until y is fully built. So there's a clear sense in which the partial building relation is only fully built by the full builders. But that arguably undermines the “upwardness” of building building relations; it runs up and back down again. So it looks like we could equally say the One does all the work, and makes all the building facts, including the partial ones, obtain. The difficulty of deciding the priority is a *prima facie* reason to think we've taken a wrong turn somewhere. But even leaving that aside, the point is that the need to give priority to one of the putatively distinct relata as far as building the building relations that hold in opposite directions between them tells against the very worry that building relations might hold between the same two distinct relata. And if the One and everything that composes it are distinct, we must give priority to one or the other as far as building the building relations between them, on pain of double-building all the building relations (an overdetermination problem). Either way, to give priority suggests one or the other of the relata, but not both, does all the building. In other words, anti-primitivism (upwards or downwards) about building is incompatible with the possibility of two asymmetric, necessitating, and generative relations holding in opposite directions between the same two distinct relata. This harkens back to the point about relative fundamentality being no notion of fundamentality at all. Any case in which these building relations appear to obtain will be one in which, facing the aforementioned grounding problem head on, and in the absence of anything to ground a distinction between parts and the whole they are strictly sufficient to compose, we should deny the distinctness between them.

All this time we've been talking about all of the One grounding the cup (and the rest). On the other hand, suppose that we could stipulate mutually exclusive criteria for parts and grounds that would exhaustively and non-redundantly divide the aforementioned hunk of material from the rest of the One, and have the hunk composing the cup, but make a different grounding claim, namely that *the rest* grounds the cup. But now we no longer have the same worry about a general relation that holds between the same two distinct relata. The rest grounds the cup while the hunk composes it, while both compose the One. Now the cup's “inner” composition and “outer” grounding look much like the

constitution of an entity according to restricted constitutionalism as a form of *Circumstantialism*. The question then becomes something else entirely. Recall from Section 4.3 that the special composition question can also be phrased as a grounding problem: in virtue of what do some x 's compose a y ? In the context of the present considerations, it takes certain background conditions for some things to compose a cup. After all, if a cup's disposition to hold water depends on the nature of water, it's not as if a world devoid of all but the aforementioned hunk of material to which that disposition-in-the-circumstances is attributed instantiates the relevant disposition. Dependence alone does not mark the boundary between the hunk and the rest, since the cup's being what it is depends counterfactually on both. As I will argue in more detail in Chapter 10 in the case of constitution, it's not clear what else could be added.

The upshot so far is that if grounding and composition hold in opposite directions between two relata, and if such complete composition and complete grounding are versions of a general relation, that general relation isn't always asymmetric. But this isn't a problem. Counterfactual dependence is antisymmetric and reflexive, that is, symmetric in the case of identity. And the possibility of apparent mutual synchronic dependence doesn't tell against the identity of the dependence relations in question, since it is implausible to simultaneously hold the relata as distinct (such that the relations could be), hold that building relations are built by one relatum (such that the whole lot isn't infinitely overbuilt), and hold that both relations are building relations. In other words, one finds an asymmetry of the dependence between distinct relata, or a symmetrical dependence between mutually necessitating relata which are not distinct, but no evidence of any deep distinction between composition and grounding.

9.3 An Objection from Failed Extensionality

Another objection to *Monism* and *Generalism* about building relies on a pair of examples.

a and b together compose the fusion $a + b$ and set-form the set $\{a, b\}$. (Bennett, 2017, p. 25)

The fact that a figure f is equilateral and the fact that f is a rectangle together set-form {the fact that f is equilateral, the fact that f is a rectangle}. But the same two facts ground the fact that f is square. In such cases, the same input builds two distinct outputs... via a single building relation. (p. 25)

These examples are a problem for both *Monism* and *Generalism*, when combined with a plausible principle of extensionality. In mereology, extensionality is the principle that wholes with all the same parts are identical. An analogous principle here would be that,

[Extensionality] “two distinct things cannot be built of the very same builders” (p. 25).

In other words, given something like *Extensionality*, what we have is an inconsistent set of claims:

- (1) a and b build c (completely);
- (2) a and b build d (completely);
- (3) c and d are distinct.

Of course, in case one of the building relations above is causation, it is not plausible to deny that c and d are distinct. So *Monism* and *Generalism* about a general relation that includes both causation and synchronic species will not obey *Extensionality*. Dependence, we can now see, doesn't obey *Extensionality*. That is certainly a reason to deny that causation and synchronic dependence relations are identical, and that is plausibly why causal and synchronic species of dependence are explanatorily ineliminable, as I argued in Section 8.2. But in the case of synchronic dependence relations, c and d are co-located by definition, since both are co-located with a -and- b . In such cases, we might plausibly deny (3), and say that the set of two facts, {the fact that f is equilateral, the fact that f is a rectangle} is the fact that f is square. Or consider a play that has as its parts exactly two acts, and an audition that has as its constituents exactly the same two acts. According to *Extensionality*, we must simply say that the play is the audition and the audition is the play. Otherwise, we have distinct but coincident entities whose differences are not accounted for by their

parts. Alternatively, if the intuitive differences between facts and sets or plays and auditions has any merit, by *Extensionality*, we should question whether we are dealing with entities with all the same parts. Of course, the problem for *Monism* and *Generalism* disappears if they have just some of the same parts. And it isn't obvious that the entities in the examples above do have *all* of the same parts. For example, Bennett passingly cites Lewis (1991), who thinks that the set $\{a, b\}$, has as its parts not a and b , but the singleton subsets $\{a\}$ and $\{b\}$. In that case, if the fusion $a + b$ is distinct from the set $\{a, b\}$, we can say that is because they are built from different things, namely, mereological parts on the one hand, and subsets on the other, not necessarily because mereological parts stand to mereological sums in a different relation than subsets stand to larger sets. Then the question becomes whether $\{a\}$ and $\{b\}$ have as parts only a and b . My answer, as already indicated in Section 3.3.3, is that for any notion of set formation that resembles building, a is not sufficient to set-form $\{a\}$, unless the empty set itself is an ontological free lunch; one way or the other, you need the empty set. If a is sufficient for $\{a\}$ because the empty set is free, then $\{a\}$ is nothing over and above a . If $\{a\}$ is something over and above a , then a is not sufficient for $\{a\}$. Likewise, if there's a difference between the play and the audition, it will be because of differences in parts, for example, because the audition includes the casting director, whereas the play includes the whole audience (or none of it), or is only what it is due to some circumstances which differ from those which make the audition what it is.

The upshot is that *Monism* about synchronic dependence relations is not off the table for reasons of *Extensionality* if one is prepared to deny that completely co-instantiated entities can ever be distinct, which I am, as I defended at length in Chapters 4 to 7. But all this points to something stronger than a defense against the objection. The same considerations can be turned into a positive argument for *Monism* about synchronic dependence relations. The problem, recall, is that

it is at least somewhat odd to think that the same input, plugged into the same nondisjunctive relation, can generate two very different outputs. Oughtn't there be something else that makes the difference? (Bennett, 2017, pp. 25–26)

Indeed, there ought to be, or there is no difference. But now the very same question can be pressed to the pluralist, who insists on the difference. The pluralist has a new grounding-like problem (see Sections 4.2 and 4.3). Admitting the complete synchronic co-instantiation of nonetheless somehow distinct entities, the question arises, in virtue of what are entities with all the same parts (or building inputs) distinct? To accept *Synchronic Dependence Relation Pluralism* and to think the same builders build more than one thing is to think the relation itself makes the difference between the built things. Standing in putatively different relations to their parts, such as composition in one case, and set formation in another, must apparently make the difference between distinct co-instantiated entities—one composed and the other set-formed from all the same parts. The differences between these completely co-instantiated entities must be grounded, not in their grounds, but in their *groundings*—in their relations to their completely shared parts. A special grounding relation must hold between an entity on the one hand, and its relation to its parts on the other. Thus, in light of *Extensionality*, the burden falls to the pluralist to make sense of the idea that differences between entities could depend on which dependence relation they have to their parts, rather than on their parts alone. But what do the realization, constitution, or composition (or grounding, or set formation) relations *themselves* add to being that the realizers, constituents, or parts (or grounds, or members) do not? Any positive answer would be at odds with anti-primitivism about building relations. The best answer is nothing, since, after all, full builders are supposed to necessitate what they build. The positive argument for *Monism* about synchronic dependence relations, then, is this. Since “no one thinks that mere partial builders necessitate what they build” (p. 49), and since “two distinct [synchronic] things cannot be built of the very same builders” (p. 25), then, if “ a and b together [necessarily] compose the fusion $a + b$ and [necessarily] set-form the set $\{a, b\}$ ” (p. 25), then $a + b$ is not distinct from $\{a, b\}$. So a and b together putatively both compose and set-form just one thing. Being identical, $a + b$ and $\{a, b\}$ stand in all the same relations to a and b . If they have the same features, which they do, and are co-extensive, which they are, then composition and set formation are identical.

Now, if the pluralist precludes the above argument by retreating from the

claim that *a* and *b* necessitate the fusion *and* necessitate the set—claiming that they necessitate the fusion or the set, or neither, but not both—they get out of this grounding-like problem by giving up the *Extensionality*-based line of reasoning against *Monism*. As in the case of the play and the interview, the idea that the latter are distinct involves appealing to different things making them what they are—for example, an interviewer partly makes the interview what it is, but not the play what it is. So although the objection from *Extensionality* backfires, the pluralist might retreat to the claim that a set of builders which fully compose, constitute or realize something, at most only partly compose, constitute or realize something else. Or the pluralist might claim that certain builders are only involved in one kind of building.

9.4 An Objection from Differing Relata

This brings us to the last putative differences between the synchronic dependence relations most relevant to the externalism. Bennett (2011, 2017) explores two putative axes of difference between building relations. One is what she calls the *relata axis*. She explains,

as I have characterized them, the relations take different kinds of relata. Composition and constitution operate on objects. But realization and microbased determination operate on properties, property instances, or states of affairs. Grounding perhaps operates only upon facts. And set formation can operate on anything: there are sets of sets, sets of events, and sets of spoons. (2017, p. 16)

This difference between the relata that various relations take strikes me as the most deeply held intuitive difference between building relations. In Section 8.3, I pointed out that it can sound “messed up” to speak of realization as a relation among entities, for example, to say that some wooden legs (partly) realize a table. In other contexts, on the other hand, philosophers seem to do this all time. It is common to talk of physical states or events realizing mental states or events, and it is common to talk of a plurality of realizers of a single realized thing. For example, Shoemaker (2003) introduces the idea of realization, saying,

One way of characterizing physicalism is as the view that mental states and events are “realized in” physical states and events. What does “realized” mean here? One common view, often advanced by functionalists, says that a mental property is a “higher-order property,” the property of having one or another of the first-order properties that satisfy a certain condition, and that realization is the relation the first-order properties satisfying that condition stand in to the higher-order property. (p. 1)

Robert Wilson (2001) also allows states to be realized (p. 5). To add further permutations, Wilson speaks of “physical states constituting a total realization” (p. 9), and of “properties usually taken to physically constitute an individual” (p. 16). On my view, of course, none of these uses are inherently problematic. The point is just that, as Bennett puts it,

It’s not so clear these relations take such different relata. Certainly, it’s not so clear how to use the English words. I have already mentioned that some people use “realization” in a quite general way. . . . We—philosophers as well as ordinary people in ordinary contexts—use verbs like “compose” and “part” and “constitute” and “ground” in a lot of ways that permit a lot of relata. (2017, p. 17)

Additionally, what Bennett calls the *unification axis* of difference between building relations is their logical form—whether or not they take a plurality in the first position. Those that do, Bennett calls unifying in the sense that they wrap many things up into one. Proper composition, for instance, is a many–one relation, while constitution and realization are (often, though not always) one–one relations. Again, it is common, perhaps even equally common, to hear talk of constituents (see Section 3.1.2), and not unheard of to speak of realizers, as we saw above. Again, for example, Gillett’s (2002a) account is directly targeted against the “flat” view of the “standard account”, that realization is a one–one relation between properties.

Now, Bennett (2017) is a pluralist about building for the separate reasons we saw in Sections 9.2 and 9.3 to do with the asymmetry and extensionality of building she thinks ought to be preserved. She is a pluralist about building, but not because of the relata and unification axes. Bennett (2017) actually

says, “I continue to think that the first few arguments I discussed [on the basis of the two axes of difference] are not successful...” (2017, p. 24). Still, the idea is that,

it is surely necessary, though perhaps not sufficient, that identical relations be extensionally equivalent and share all formal properties. So, quite generally, there are two easy strategies for arguing that two relations are distinct. First, we can look for ordered pairs in the extension of one relation that are not in the extension of the other. Second, we can launch what amount to Leibniz’s Law arguments by pointing out that one relation is, say, transitive and the other is not. (2011, p. 96)

Applied to the case of building relations, the reasoning is this:

The relations have different logical form, and take different kinds of relata. For example, recall that composition is a many–one relation on objects, and realization a one–one relation on properties or property instances. So they are not extensionally equivalent, and violate Leibniz’s Law in other ways as well. Presumably it follows that they are distinct relations. (2011, p. 96)

Bennett (2011) is quick to point out that this does not entail that *Generalism* is false. After all, the idea that building relations are species of a common genus or determinates of a common determinable which is somehow more fundamental or natural than its species or determinates is perfectly compatible with the distinctness of its species or determinates. To reject *Generalism*, one would have to additionally think that “the differences between the standard building relations show that they cannot be determinates of a common determinable that is more fundamental than they are” (2011, p. 96). But that is not plausible. Other quite natural relations take relata from a variety of ontological categories, and can be plural or singular. Bennett says,

consider the identity relation. Everyone thinks that the identity relation holds between relata of all sorts of ontological categories. This stapler is this stapler, the number four is the number four... and so on and so forth for events, states of affairs, properties, tropes, relations, and whatever else your ontology contains. It is not the case that there is one identity relation on material objects, and a distinct one on, say, states of affairs; it is the same relation. (2011, p. 98)

So *Generalism* about building isn't refuted because of the axes of apparent difference between building relations, and neither is *Generalism* about synchronic dependence relations.

I don't think *Monism* about synchronic dependence is refuted by the apparent differences either. But because Bennett accepts different arguments against *Monism*, she only hints at how the counterargument on the basis of these differences might go. She says in passing that, "Matters are actually more complicated than I am letting on, but let us grant for the sake of argument that such reasoning successfully shows that the relations are distinct" (2011, p. 96). She then adds in a note that, "It is actually left open that all of the different relational predicates refer to the single fundamental determinable relation that I explore. . ." (2011, n. 35 on p. 96).

This is the view I advocate about differing synchronic dependence locutions. I've done quite a bit of groundwork on this point in the last two sections. In Section 9.2, I argued that when complete mutual dependence holds between two relata, they are identical. In Section 9.3, I argued that the *Extensionality* principle—that two distinct synchronic things cannot be built from the very same builders—together with the thought that full builders necessitate what they build, and the thought that building relations are themselves built by the builders, entails that anything completely built by the same set of builders is identical.

In both cases, the fact that nothing could make for distinctness between the relata without adding to one of them makes them identical. Of course, if we have mere partial builders, the story is different, but the objections do not arise. The cup's partly composing the One and part of the One grounding the cup doesn't show that composition and grounding are distinct. Likewise, a and b fully composing $a + b$ and *partly* set-forming $\{\{a, b\}\}$ doesn't show that composition and set formation are distinct.

So, synchronic necessitation doesn't *circle up and down* between a distinct input and output, and synchronic necessitation doesn't *fork* upward into distinct outputs. A similar strategy works for denying any deep difference between building relations on the basis of having different logical forms and taking different relata. I will argue that synchronic building doesn't *string up and*

up between distinct relata, and it doesn't *run up in parallel* between distinct relata. Let me explain.

The unification axis might be taken to show that chains of differing synchronic dependencies “string up” in series. The idea that constitution is a one-one relation between entities, while composition is unifying—a many-one relation between entities—results in the idea that, for example, some atoms compose a piece of marble which constitutes a statue. Hence the string: the xx 's compose a y , and y constitutes a z . Even so, here's an argument to the effect that the xx 's both compose and constitute z , and that composition and constitution are identical as a result.

- (1) a composes b .
- (2) b constitutes c .
- (3) If a composes b , and if b constitutes c , and if $a = b$, then a constitutes c .
- (4) If a composes b , and if b constitutes c , and if $b = c$, then a composes c .
- (5) $a = b$.
- (6) $b = c$.

So,

- (7) $a = c$, and a constitutes c , and a composes c .
- (8) That is, composition and constitution obtain between the same relata.
- (9) That is, composition and constitution do not differ in the relata they take.

Further,

- (10) Composition and constitution, as synchronic dependence relations, have their other features in common.
- (11) If relations obtain between the same relata and have the same features, then they are identical.

So,

- (12) Composition and constitution are identical.

The first two premises belong to the pluralist, so I assume them for the sake of argument. The third and fourth premises follow from the indiscernibility of identicals. The argument for (5) and (7)—the identity of wholes with their complete parts and the identity of constituted things with their complete constituents—was the subject of Chapters 4 to 7. Recall from Chapter 4 that distinctness results in grounding and overdetermination problems, and recall from Chapter 7 that arguments for their distinctness trade on equivocations between strict and sloppy notions of sufficiency for the properties that putatively distinguish them. Premises (8) and (9) are restatements of the intermediate conclusion (7), but I state them this way to make it obvious how they and (10) jointly satisfy the conditions in (11). Premise (10) is a statement of my view that synchronic dependence relations share features, and is similar to Bennett’s own view that the class of building relations is unified. Premise (11) is a version of Bennett’s claim that “it is surely necessary, though perhaps not sufficient, that identical relations be extensionally equivalent and share all formal properties” (2011, p. 96). I treat extensional equivalence and the sharing of formal properties as sufficient for the identity of relations, simply because I cannot imagine what else would differentiate relations, so I must leave it to anyone that still wants to insist on *Synchronic Dependence Relation Pluralism* to fill that in. By my lights, then, the argument is sound: constitution and composition are identical.

In the case of realization, the idea is that realization is a relation among properties, rather than entities. The first thing to note, of course, is that realizing and realized properties don’t float free of entities; they are instantiated by entities and individuated by them. So it’s as if we have *parallel* synchronic dependence relations. Realization runs alongside constitution (or composition) insofar as properties of constituents (or parts) realize those of the constituted (or composed) entities. Again, however, pluralism here rests on the distinction between properties and entities. Of course, a given entity may have many properties, and thus not be identical with just one of them, so denying distinctness between entities and properties is not quite as straightforward as denying the distinctness between a whole and its complete set of parts. And yet, unless we want grounding and overdetermination problems (or the causal impotence

of properties), physicalists should think that real properties are efficacious and ontologically innocent with respect to their instantiations. Thus, if some set of physical conditions is strictly sufficient for many properties, then the instantiating entity should be identified with all of its properties, such that partial identity obtains between the entity and each of its properties, just as partial identity obtains between a part and the whole it composes. Again, there is no reason to think the partial identity relation is distinct when it obtains between a set of properties and a member of that set, as opposed to between an entity and one of its properties, or as opposed to between a whole and one of its parts.

Now, isn't it a corollary of all this that complete and partial dependence have different extensions and features? They do. And yet, given all the xx 's that necessitate some y , that is, given y (xx 's = y), the fact that an x takes part in composing y is no addition to being; that relation is an ontological free lunch. And the relation does not obtain without the existence of the dependent relatum. In short, without existing, no thing can depend on anything else. So the part-whole relation itself depends on the whole, and there's a sense in which the particular features of partial necessitation, partial synchronic dependence, or partial identity relation—features such as the asymmetry of the one-way counterfactual conditional that express the relation—are made by the whole set of parts. It is part of the identity of wholes which come apart that they cannot exist without their parts. Still, a whole instantiates both the identity relation between itself and its complete set of parts, as well as partial identity with—or partial synchronic dependence on—each of its parts. The identity relation is distinct from the partial identity relation, just as any proper part is distinct from the whole. In that case, there are two qualitatively distinct synchronic dependence relations, partial and complete. But notice that this distinction cross-cuts realization, constitution, and composition. The relations of complete realization, complete constitution, and complete composition are the identity relation, and are thus identical. The relations of partial realization, partial constitution, and partial composition are the partial identity relation, and thus identical.

Finally, a quite general way to see the trouble with *Synchronic Dependence*

Relation Pluralism is to note that in each case that was supposed to show the difference between the relations, we find more than one putative synchronic dependence relation necessitated by the same state of affairs. It's not the case that there's some realization over here, constitution over there, and composition over yonder. The relations are instead spatiotemporally coincident.

But let me spell that out just a little further. First, again, no one posits realization as a relation between uninstantiated properties. Second, it would not be sensible to think that an entity is either composed or constituted, but not both, such that composition and constitution were never spatiotemporally coincident. If a constituent was partless, it would be fundamental. The constituted thing would be partless, and arguably fundamental as well (being fully necessitated by the partless constituent). For example, a statue constituted by some marble that was not itself composed of anything would make the statue fundamental. Third, the non-coincidence of composition and constitution would make composed things unconstituted, and thus arguably either non-existent (as the compositional nihilist thinks), or, again, equifundamental with the parts. If a fusion doesn't constitute anything, it would just be the parts, which are fundamental, and so would the fusion be.

Thus, attempting to separate the obtaining of realization, constitution and composition in space and time would yield things without properties and properties attached to nothing, the impossibility of building statues, and tables about which we'd have difficulty saying whether they were non-existent or fundamental.

At any rate, I won't pursue this possible strand of pluralism further, because no one actually appears to posit constitution as a relation between partless entities, and only very few people think parts never constitute a new thing. Those that do, of course, either think the composition relation never obtains such that composition could be a relation distinct from other synchronic dependence relations (compositional nihilists), or they think, as I do, that complete parts/constituents compose/constitute things, but that those things are nothing over and above those parts/constituents.

Again, however, on the standard approaches we've already seen, the pluralist can go in for the distinctness of the relations predicated on the distinctness

of spatiotemporally coincident entities and properties—the picture in which relations circle back down, fork upward, string up and up, or run up in parallel between levels. But holding those relations distinct shoulders the same burdens of holding the relata distinct. That is, it faces grounding and overdetermination problems. In short, there are no synchronic dependence circles, forks, strings, or train tracks that do not collapse because of the ontological innocence of the relata involved. So the relations aren't distinct.

Finally, there's no upward *convergence* from distinct parts of a physical base to a single entity. This last picture of the difference between synchronic dependence relations has already arisen briefly in Section 2.3 and the end of Section 9.2. I will save most of my denial that this picture is plausible for Section 10.1, however, I want to introduce it while the other pictures are fresh and say why, even if it is plausible, it doesn't amount to a distinction between the three relations I claim are identical. On this picture, relations converge upward from distinct relata in the first place of the relations to the same relatum in the second place of the relations. For example, suppose that a brain state realizes a mental state in some circumstances which, let's suppose, "ground" the mental state's being the state that it is. The first problem with this is that it doesn't amount to a difference between constitution, composition, and realization. No one seems to think that a statue is *composed* of marble while being *constituted* by the circumstances in which the marble exists. Realizers, constituents and parts are instead all held to be among the core, central, or most salient builders of an entity, while *Circumstantialism* about these relations names the equally necessary background conditions just that. Suppose we called the job of the circumstances "grounding." Insisting on *Circumstantialism* about composition, constitution, and realization *might* be a way to insist on a real difference between those three together on the one hand, and, say, grounding on the other. Still, even if it's somehow wrong to believe *Unrestricted Constitutionalism*—that anything on which an entity depends synchronically is constitutive of that entity—it wouldn't be mistaken because some of the things on which an entity depends are not constituting it but actually composing or realizing it, but because some were grounding it.

Moreover, it's not obvious that synchronic dependence relations are really

divisible this way. For one thing, that would make *Circumstantialism* more than a matter of bookkeeping, and hinges on *Circumstantialism* about those relations being plausible at all—an issue I have covered in detail in Chapter 7. The second problem with the upward convergence picture, then, is that it becomes difficult to see how grounding could be defined so as to exclude constituents and parts. The problem is stipulating or discovering where the boundary between parts/constituents/realizers and grounds lies, hence the classic EM question, “Where does the mind stop and the rest of the world begin?” (Clark and Chalmers, 1998, p. 7). Understandably, constitution and realization have explicitly been used across this boundary, as in, respectively, semantic externalism (see Section 2.2), and the notion of “total realizers” of a mental state (see Shoemaker, 2003). Finally, note a methodological principle that Bennett (2017) recommends, given unity in the class of building relations: “One should not break the analogy without argument” (p. 78). So the burden to justify a double-standard lies with anyone who would motivate the difference between two subsets of synchronic dependence relations by way of being a circumstantialist about some while being a strict necessitarian about one or more others.

Wrapping Up

In short, we have plenty of reasons to think composition, constitution, and realization are identical in that they obtain between relata that are ontologically innocent with respect to one another and are spatiotemporally coincident, and no good argument for *Synchronic Dependence Relation Pluralism*. That means it doesn’t matter what synchronic dependence terminology externalist claims are made using. Together with the *Reducibility* claim I have been defending, the interchangeability of these terms means that from our partial constitution or realization by technologies, we can infer our partial identity with them. But I only hinted at why we should deny the “convergence” picture of the difference between, for example, constituents and grounds. More needs to be said about this problem of knowing where the boundary lies between “core” and “total” realizers, or the real constituents and mere circumstances. And that brings us

squarely back to the EM debate.

Chapter 10

The Metaphysics of Extended Minds

In Chapter 2, I argued that the debate over externalism is a special case of diverging notions of constitutive dependence, kinds, and the proper objects of science. I have argued since that various dependence locutions used in physicalist explanatory projects look related, and give rise to the same puzzles when paired with antireductionism. A solution to one problem promises to be a solution to many, independently of the synchronic dependence locution in which those problems are stated. That is also a reason to believe the synchronic dependence relations are closely related. Since dependence, coincidence, and distinctness are the common ingredients for these problems, those ingredients should be inspected. Given that physicalism is a well-motivated thesis, we should not drop the dependence of higher-level things on physical matters. One of the usual arguments for distinctness is predicated on the multiple realizability of higher-level kinds—the idea that since functional kinds admit of diverse realizers, the kind is not identical with any single physical base, and that each member is distinct from its physical base because it has the property of being of that kind while its physical base supposedly does not. But similarity and kinds can be explained in terms of multiple-causability, switches in nature. Even though switches, and thus multiple realizability, are ubiquitous, kinds don't fail to be physical; they can be identified with their members or perhaps with the switches that individuate them. Functionalism, and even the possibility of “wispy” one-dimensionally projectible functional

kinds, is thus compatible with a coherent physicalism that does not result in overdetermination by insisting on the distinctness of higher-level entities, kinds and properties from lower-level physical conditions. Further, there is no valid inference to antireductive physicalism from the failure of narrow physical bases to have the non-categorical properties we ordinarily say the entities they (partly) instantiate “have.” The inference is what I call the sufficiency error. So we have no reason to believe that higher-level entities are not identical with their lower-level parts, constituents, or realizers, while the idea that everything is physical, that everything real is causally efficacious, and the causal closure of the physical world, suggests their identity. The solution to various puzzles of material constitution is thus not to distinguish between higher-level things and their *complete* parts, constituents, or realizers. Now, a further feature uniting these relations is that they are causally tainted. But even though causal relations can be constitutive of higher-level entities, there is an important distinction between how given states of affairs over their duration are brought about prior to their beginning, and what makes them up for their duration. The causal/constitutive distinction is not meaningless. The distinction between constitution, composition, and realization, on the other hand, is much less clear. Synchronic dependence relations are identical insofar as wherever there are properties there are entities, and vice versa, and so wherever there is constitution there is realization, and vice versa. The relations are coincident and instantiated by relata which are ontologically innocent with respect to one another. So whether or not externalism is true doesn’t hinge on whether it is put in terms of parts, constituents, or realizers, implementation, physical base, etc. Rather, the truth of externalism hinges on whether members of conscious kinds—whether mental states, cognitive processes—have properties whose instantiations are partly extraneural and about which fruitful generalizations can be made. And they do; traditionally, intentional states are characterized by non-categorical properties, and there is no reason to believe that such properties should be disqualified in advance as objects of scientific enquiry. Indeed, we already use them in both ordinary and scientific explanatory and predictive practice.

It remains to apply these arguments in more detail to the questions left over

from Chapter 2’s initial pass at an argument for externalism. In Section 10.1, I argue that there is no clear competitor to *Unrestricted Constitutionalism* as I introduced it in Chapter 1 and Chapter 2. In Section 10.2, I further defend this notion of constitution against a certain intuition and a certain assumption that together tend to promote internalism. The intuition is that we could be brains in vats, which can be taken to show that we could be wrong about everything, such that what explains offline and illusory experiences is what explains apparently veridical experiences. Then, assuming that brains primarily explain offline and illusory experience, brains explain experience generally. I deny that the intuition shows that we could be wrong about everything, and thus deny the analogy between offline/illusory experiences and veridical ones. I then challenge the assumption that even offline/illusory experiences are *completely* explained by the brain. As a result, I argue in Section 10.3 that externalism need not claim that paradigmatically cognitive processes and mental states occur in the brain, which concedes too much to internalism. In case the resulting cognitive kind is thought to be too slippery, I apply the defense of kinds characterized by non-categorical properties to the mental, the cognitive, and minds, in contrast to eliminativism about the latter. In Section 10.4, I conclude by arguing that externalists should bite the bullet with respect to the “bloat” objection—roughly, that anything might be cognitive—and account for the counter-intuitiveness of my view and the intuitiveness of internalism in terms of differing epistemic salience within the ontological dependence bases of higher-level things.

10.1 Revisiting the *Unrestricted Constitutionalism* Premise

I begin by re-assessing, in light of the preceding chapters’ exploration of synchronic dependence relations, the claim that anything on which an entity depends synchronically is constitutive of that entity. By, “entity,” I intend to include objects, states, processes, events, phenomena and experiences. By “synchronically,” I mean, for a thing’s duration. Here is the argument for that claim, which I introduced in Section 1.2.

[Differentiability Condition] If constitution is best analyzed simply as synchronic ontological dependence, then anything on which an entity depends synchronically is constitutive of that entity.

[Non-Differentiability of Constituents and Background Conditions]

Constitution is best analyzed simply as synchronic ontological dependence.

So,

[Unrestricted Constitutionalism] Anything on which an entity depends synchronically is constitutive of that entity.

Strange as it might appear, *Unrestricted Constitutionalism* is not without precedent. About causation, Lewis (1973) says,

We sometimes single out one among all the causes of some event and call it “the” cause, as if there were no others. Or we single out a few as the “causes,” calling the rest mere “causal factors” or “causal conditions.” Or we speak of the “decisive” or “real” or “principal” cause. We may select the abnormal or extraordinary causes, or those under human control, or those we deem good or bad, or just those we want to talk about. I have nothing to say about these principles of invidious discrimination. I am concerned with the prior question of what it is to be one of the causes (unselectively speaking). My analysis is meant to capture a broad and nondiscriminatory concept of causation. (1973, pp. 558–559)

And Bennett thinks Lewis’s point can be adapted with respect to building generally. If Lewis accepted the analogy, he would presumably be a building necessitarian (see Section 7.1) and thus a necessitarian about constitution. That is, he would endorse *Unrestricted Constitutionalism*. On the other hand, my view that nothing differentiates constituents from other members among a set of physical conditions on which an entity depends at some point in its duration is by no means the ordinary one. Consider Gillett’s formalization of the constitution relation:

(Constitution/Parthood) Individuals s_1 – s_n constitute, and are parts or constituents of, an individual s^* , under conditions $\$$, *if and only if* s_1 – s_n are spatially contained within s^* , many of s_1 – s_n bear

spatiotemporal, powerful and/or productive relations to each other, and all the powerful properties/relations of s^* are realized by properties/relations of s_1-s_n , under $\$$, but not vice versa. (2013, p. 323)

Here, the extent of the realization base—of those properties individuating the entity—is the measure of the spatial (and, we might add, temporal) extent of the individual. Now, one might think the spatiotemporal containment condition does the required work. But it actually does nothing to justify our intuitions about entities' narrow boundaries as against mere background conditions. I accept the spatiotemporal containment condition, and I prefer strict *Necessitarianism*. The strict necessitarian can accept Gillett's condition; the necessitarian just thinks the spatiotemporal boundary of the constituents is much broader than we intuitively think it is. Ostensibly, the extent of cognitive processes and mental states are in question in debates about externalism. The extent of entities is precisely the question, not an answer to how to differentiate constituents from other parts of the dependence base. Since Gillett's definition requires that the constituents' properties must realize *all* of the constituted entity's properties (under certain conditions), it assumes the admittedly common-sense difference rather than motivates it. So the extent of constituted things is generally at issue insofar as there is an important choice to be made between *Circumstantialism* and strict *Necessitarianism* about the full builders of entities whose non-categorical properties are ordinarily attributed to physical conditions manifestly not sufficient for those properties.

I know of no other formalizations to which the same considerations do not apply. In particular I have in mind Craver's (2007) "mutual manipulability" account of constitution, and Baumgartner and Casini's (2016) abductive theory of constitution, both of which require, like Gillett's account, that constituents be spatiotemporally contained within the constituted. Or consider Baker's definition of constitution:

certain kinds of intentional states, certain kinds of social and political entities and certain conventions. . . are essential to national flags: nothing is a flag without them. For any primary-kind property G , such as the property of being a national flag, call the milieu required for something to have G , " G -favorable" circumstances. G -favorable

circumstances are the total background conditions that must obtain for something to have G . For any particular place and time, the presence of G -favorable circumstances is necessary for the property G to be instantiated then and there; but the presence of G -favorable circumstances by itself is not sufficient for G to be instantiated then and there. An informal idea of material constitution is this: where F and G are distinct primary-kind properties, it is possible that an F exists without there being any spatially coincident G . However, if an F is in G -favorable circumstances, then there is a new entity, a G , that is spatially coincident with the F but not identical to it. (Baker, 1999, p. 147)

Again, the spatial coincidence of G with F is merely stipulated. Baker thus assumes, rather than does anything to motivate *Circumstantialism* about the dependence relation being defined. Note, too, that there's little more at the heart of these definitions than a notion of ontological dependence. Baker says roughly that a thing in some circumstances is *sufficient for the existence* of another thing (and that the circumstances are necessary—note how close we are to providing necessary and sufficient conditions for the second thing). Gillett (2013) interdefines constitution and realization, and Aizawa and Gillett (2009) defines realization using the *in virtue of* locution, for which they are criticized by Endicott (2016a), who suggests that realizers *explain* the realized (see Section 7.1). But this doesn't help. As Bennett says,

The problem is that such talk [of dependence being explanatory] is ambiguous. In one sense, adding an explanatory requirement to my characterization [of building relations, including constitution and realization] would add nothing; in the other sense, adding an explanatory requirement is neither plausible nor well-motivated. (2017, p. 61)

Understood in a metaphysical sense, “explains” works just the way “in virtue of” does. Understood in an epistemic sense, the characterization is not plausible, because, as we saw in Section 5.2, such explanation is context sensitive. Explanation, in the epistemic sense, and dependence are related, of course. Suppose ontological dependence grounds the truth of such explanations. In that case, dependence does the explaining; dependence explains, in *both*

senses, “explanation,” in the epistemic sense. And Wilson and Craver (2007) make the same point about realization that I’m making about constitution:

A second possible restriction on the realization relation would be to make it constitutive—that is, to require that the realizer of O’s having P be wholly contained within O’s spatial boundaries. (p. 95)

But, they continue,

it is plausible to argue that if realizers are to be metaphysically sufficient for the properties and activities they realize, then they must extend beyond the boundary of O and beyond the relationships that exist among the things inside O. (p. 96)

So we don’t, so far, have a notion of constitution on offer that doesn’t boil down, in one way or another, to synchronic ontological dependence. Certainly, the realization relation can’t non-trivially constrain the extent of the constituents while the constitution relation constrains the extent of the realizers. In the formalizations we’ve seen, without something *more* to constitution than dependence, nothing rules on the boundary between constituents and the required circumstances. The intuitive boundary can be drawn differently, or the constituents and circumstances relabelled, and the definition of constitution would equally apply. That’s to be expected, as far as the present view is concerned. That is, there isn’t much that *can* be added, except, as I have argued, the synchronicity of constituents with what they constitute. And without something more to constitution, we’re left with a notion that applies equally to what we intuitively think of as background conditions.

Here’s a further strategy I can think of for attempting to carve out the core realizers or constituents of a higher-level thing. I said in Section 7.3 that non-categorical property-bearers are either not quite *what* we ordinarily think they are, or not quite *where* we ordinarily think they are. We can either take a spatiotemporal region as an entity and attribute to it just those properties for which it is strictly sufficient, or take a property as our explanandum and end up with all those conditions sufficient for it. But what about taking multiple properties usually attributed to some intuitive core, and, despite knowing that the entity individuated by those properties is not wholly coincident with that core, asking whether there is overlap at the core with the sets of fully sufficient

conditions for each property. The picture here is of a Venn-like diagram where a shape for each property represents the spatiotemporal extent of its realizing base. Might the core be the overlap of all these shapes?

Probably not, for at least two separate reasons. First, it's conceivable that some properties attributed to the core may require different parts of it and so not overlap within the wanted core (even if they overlap outside the wanted core). In case of non-overlap, however, according to this strategy, we would have to choose among "cores" and lose certain properties which we could not even sloppily attribute to the chosen core. Non-overlap looks especially plausible in the case of rather more spatiotemporally strange entities, such as persons. Consider my having a particular eye colour and my having a certain intelligence whose realizations presumably have little to do with one another. Or consider my shirt's being partly grey and partly maroon. On the thought that it takes some light to realize a thing's being coloured, the overlap is in the wrong place; the overlap of the physical bases of these two properties is beyond what we think of as the shirt. Thus, notwithstanding the possibility of non-overlap, the second and more damning issue with the Venn strategy is that it is doubtful whether, for every given set of properties, their physical bases will overlap such as to guarantee just the core we want. It's plausible to think that, frequently, certain conditions which are intuitively background conditions and not among the core parts, constituents, or realizers of an entity will be commonly required for all of the properties of that entity. Bennett (2017), for example, says, "it might seem natural to say that being taller than is an internal relation, the obtaining of which is fully grounded in the heights of the relata. But in fact, its obtaining also depends on the curvature of spacetime" (p. 54). If that's right, then the curvature of spacetime is among the realizers of a whole lot of properties, and thus among the core realizers of some entities, even on the picture of the core as the overlap in the conditions necessary for a set of properties.

Although it would not obviate the point that the entity individuated by those properties is not strictly contained within the region of overlap, still, if this strategy worked even roughly, it might go some way to explaining our *intuitions* about the boundaries of entities, because conditions like the curvature

of spacetime would be ones we could take for granted. While the curvature of spacetime is ontologically salient to the “metaphysical” explanation of size relations, it won’t be salient to most explanations in practice. Likewise, there may be other, *non-dependence* based ways to differentiate among the complete dependence base or total realizers of an entity, for example, Otto’s belief about the museum. The belief depends on biological Otto, on his notebook, and on the museum, among other things. There are dozens of ways we might carve up these physical conditions. There are the parts of a thing that are more or less closely interacting, moving about together, and so forth. In particular, I’ll revisit the morally salient subset of the total realizers in Chapter 11.

What about the boundaries of a *system*? I’ve spent quite a bit of time talking about entities and their properties in general, and while it’s true that much of the extended mind debate is transacted in terms of entities such as cognitive processes and mental states, some of the debate has turned on the question of the boundaries of cognitive *systems*. Is there a difference among sorts of entities as far as considerations of sufficiency for their properties—for example, between systems on the one hand, and events, processes, objects and states on the other? One thought might be that the proper parts of our cognitive systems are those which are portable and thus available when needed, as well as trustworthy—criteria we saw in Section 2.3 suggested by Clark (2010c, p. 79). Could some portable system be the core we intuit? As Ross and Ladyman (2010) put it,

the ubiquity of the containment metaphor derives from the fact that as tool builders we humans are naturally interested in isolating systems in such a way that we can transport them around without significantly changing the kinds of processes we can use them to effect (Cartwright 1989). But modeling the world in terms of the affordances it offers for human manipulation at the special spacetime scales in which we plan and execute actions is not the way to construct objective accounts of the universe. (pp. 161-162)

Actually, the ubiquity of the containment metaphor derives from the fact that entities and properties are spatiotemporally located. But the idea that we isolate systems, transport them, and use them to effect particular processes has some apparent merit as a basis for individuating a core cognitive system.

Unfortunately, this doesn't secure the desired ontological result. Consider portability to be a disposition. According to Prior et al. (1982), "For each disposition we can specify a pair of antecedent circumstances and manifestation which together determine the disposition under discussion. In the case of fragility, the pair is (roughly) < knocking, breaking >" (p. 251). In the case of portability the pair is (roughly), < applying a relevant force, moving relative to some context of interest >. That the movement in question is relative to a context of interest and that the force be relevant is crucial of course. Otherwise, Mount Everest is portable, given its disposition to move when a force is applied (eg, the inertial and gravitational forces acting on the Earth as it rotates, revolves around its barycenter with the Moon, and orbits the Sun). Our cognitive systems don't exactly "have"—all on their own—the property of being portable.

To repeat, the overlap strategy doesn't help us say what the "true" ontological constituents of Otto's belief are, as compared to the mere circumstances it depends on, absent some other stipulation about constitution, and we've seen that one such stipulation—a portability criterion—fails because it invokes a property that depends on wider circumstances than the thing to which we attribute that portability.

All this might seem to suggest that a definition of constitution one way or the other *is merely stipulative*. In that case, the circumstantialist might dig in her heels and complain that the necessitarian begs the question about constitution as much as the circumstantialist would by adding (yet to be seen) non-dependence based criteria to her definition of full building, proper constitution, or core realization. I seem to be saying that what are normally considered background conditions are actually constituents if constitution is just a synchronic dependence relation, and that constitution *just is* a synchronic dependence relation. As I noted in Section 2.3 (Footnote 5), it might seem like the restricted and unrestricted notions of constitution could co-exist if theorists simply made their claims more precisely in terms of either. Wilson (2007), for example, argues that we can account for the truth of the claims, "This water is nothing more than the molecules that constitute it," and, "This statue is something over and above the marble that constitutes it," in terms of

differing notions of constitution. The key difference between these notions of constitution is the sort of “necessitation,” involved. The former “compositional” notion of constitution involves *intrinsic* necessitation. The latter “ampliative” notion of constitution involves *extrinsic* necessitation. A thing or things, x or the x s, compositionally constitute(s) y if,

x is in some intrinsic state(s), or the x s that compose y are arranged, during p such that x itself, or the x s themselves, necessitate the existence of y . (p. 17)

By contrast, x or the x s ampliatively constitute(s) y if,

x (the x s) is (are) in extrinsic conditions during p that themselves necessitate the existence of y . (p. 18)

I said earlier that *Unrestricted Constitutionalism* could be understood as the claim that “anything on which an entity depends synchronically is *compositionally* constitutive of that entity.” So it might seem as though the claim that “not everything on which an entity depends synchronically is *ampliatively* constitutive of that entity” is not in direct conflict with *Unrestricted Constitutionalism*. But, on my view, it remains true that whenever something is constituted at all, *some* physical conditions are compositionally constitutive of that entity. So the ampliative constituents of an entity are also compositional constituents of it. And one who doubts the very existence of the boundary between the ampliative constituents and the remaining conditions on which an entity depends will think that anything on which an entity depends synchronically is, taken by itself, ampliatively constitutive of that entity. According to *Unrestricted Constitutionalism*, we can account for the truth of the claim about the water and the claim about the statue, if they are true, by reading them more precisely in terms of partial and complete constitution. On that view, some water (or any entity) *is nothing more than* the conditions that completely constitute it, and a statue (or any entity) *is something over and above* any conditions that only partly constitute it.

So the disagreement between *Unrestricted Constitutionalism* and its denial is about whether the fully sufficient conditions on which an entity depends synchronically can be differentiated as concerns the relation of synchronic ontological dependence in which they stand to the constituted entity. *Unrestricted*

Constitutionalism is the claim that they cannot, and its rival is the claim that they can. I'm thus not merely widening the definition of constitution to include something *else*; I'm attempting to cast serious doubt on whether there really is anything *else*. That is, I'm questioning whether there is more than one relation of synchronic ontological dependence at all, let alone more than one sort of constitution. If there's only one sort, then there will only be one sensible definition of constitution. And I think what we tend to call background conditions fit the definition of constituents insofar as the latter have been sensibly defined in terms of synchronic ontological dependence, which is that aspect of constitution on which there is some consensus. Even those definitions that assume *Circumstantialism* undermine it by failing to offer a definition that picks out the truly constitutive physical conditions from the apparently background ones. *Circumstantialism* is, again, the principle that,

“Let C be some to-be-specified set of background circumstances that includes neither y nor anything that fully builds y . For all x and y and all building relations B , if x fully B 's y , then, $\Box[(x + C) \rightarrow y]$.”
(Bennett, 2017, p. 52)

But for all *Circumstantialism* about constitution says, the museum (x) might fully constitute a belief about itself (y) in the circumstances in which there is Otto and his notebook (C). So, without specifying the C/x boundary, there is already a point in favour of *Necessitarianism* and *Unrestricted Constitutionalism*.

And there are two additional points working against the circumstantialist's picture. The first arises with attempts to specify the to-be-specified C , which, note, is to correspondingly specify x . Recall from Chapter 2 that it looks like Clark is plausibly a circumstantialist about constitution when he says, “Chalmers and I thus offered an argument (which one may accept or reject: that is, of course, another matter) concerning conditions not of ‘being cognitive’ but for incorporation into a cognitive system” (Clark, 2010a, p. 84). While I have countenanced that view as a kind of *Circumstantialism*, specifically, a restricted constitutionalism, for the sake of argument, it's actually not clear that Clark's criteria work the way he implies when he offers “criteria to be met by nonbiological candidates for inclusion into an individual's cognitive system”

(Clark, 2010c, p. 79). To see this, recall from Section 2.3 that Adams and Aizawa's (2001) mark of the cognitive—being non-derivedly representational—was a criterion of kind-membership. As such, the question of the mark of the cognitive is orthogonal to the question of the physical extent of a given member of the cognitive kind. So, too, with any criteria for kind-membership. Now, why think that Clark's criteria work as marks of the cognitive rather than marks of the constitutive? For one thing, because the notebook's being "reliably available and typically invoked", and its information being "automatically endorsed," deemed "trustworthy", "easily accessible as and when required," and "consciously endorsed at some point in the past," (p. 79) look, on closer scrutiny, more like properties which, *even allowing ourselves to be sloppy*, are attributable to the Otto-notebook coupling rather than to the notebook alone. For another, Clark's are criteria for integration into something *cognitive*, rather than criteria for constitutive relations generally. We would presumably need different criteria for deciding the constituents of cognitive systems as opposed to tennis balls or economies, but these differences shed no light on what is common to each such putatively differing "constitution" relation, which is synchronic ontological dependence expressible in counterfactuals. So, arguably, the property of *being reliable* is one to be explained constitutively rather than explaining constitution. Reliability picks out an instance of a kind-member, rather than the true constituents of that member. Even supposing we take Clark's view as offering criteria, not for being cognitive, but for being part of something cognitive, it is not clear that such criteria are consistent with the way relational predicates work. For example, if one thing is more fundamental than another, to the left of it, younger than it, or an ancestor of it, it's not clear how being reliable, or blue, or large, or roughly spherical, or any additional property a thing might have, would make a difference to its relative fundamentality, age, or genealogical relation to another. Those relations obtain, or do not obtain. Of course, there might be properties that obtain whenever a relation does, for example, being in a relation to something else *is* to "have" (or partly instantiate) the property of being in that relation. So failing to have *that* property is to fail to have that relation. But why should a constitution relation obtain between the notebook and Otto's belief just in case those properties

of the Otto–notebook coupling cited by Clark obtain? It would be as odd to stipulate this ad-hoc species of synchronic ontological dependence as it would be to stipulate a relational predicate for things’ being blue-and-to-the-left-of or large-and-more-fundamental-than. Rather, blue-and-to-the-left-of is a *prima facie* unparsimonious relation that fails to carve at the joints, so to speak.

Further, even allowing such ad-hoc relations and defining cognitive constituents as those which are depended-on-and-reliable-and-available-and-trusted-and-so-forth, certain infrastructural elements of Otto’s world we might intuitively think of as circumstantial—roads, address systems, and gravity—might be reliable, available, trusted, and so forth.

Further still, the apparent intuitiveness of *Circumstantialism* is actually just a restatement of our habit of sloppy property attribution, and this intuitiveness hardly works in favour of circumstantialism if intuitions conflict within broadly circumstantialist positions. Intuitions ostensibly conflict in the extended mind debate over what and how many criteria to add to carve out from the *truly full* builders just the “full” builders we might want. The mere fact of this divergence highlights the ad-hoc and unparsimonious nature of stipulating a constitutive relation picked out by additional properties of that which is constituted. And the ad-hoc and unparsimonious nature of any specific stipulation counts against its plausibility as an account of what is “constitutive” generally. So, the first thing working against the circumstantialist is simply this. As already pointed out, there is no positive account of the difference between the core realizers or salient constituents and mere background within the analytic metaphysics literature. There is nothing further within the extended mind debate to furnish such an account. And the lack of a plausible account makes it at least a little dubious that one can be given.

The other thing counting against *Circumstantialism*, including restricted constitutionalism, is worse. Recall that Bennett (2017) only *slightly* prefers *Circumstantialism* for its apparent intuitiveness (p. 54), but that it is mostly a matter of bookkeeping, and where we sweep various untidiness. By contrast, I don’t think the untidiness of non-categorical property attribution should be swept away at all. What finally tips the balance in favour of the necessitarian is once again the point that, as Wilson (2001) puts it,

core realizations in themselves are not metaphysically sufficient for the properties they realize, but must be part of some larger functional system. This point is of some significance in itself because it is core realizations that are typically invoked in discussions of reductionism, realization, and functionalism, especially in the philosophy of mind, even if no one really believes the sufficiency thesis to be true of them. (p. 14)

Now, Wilson (2001) infers from his similar diagnosis that non-reductive physicalism might be revised. As I understand it, on his view, wide realizations don't reduce to the constituents of the bearer of the realized property, because Wilson counts those constituents in the traditionally more narrow way. On my view, those constituents are only partial ones, in much the same way that "core" realizers are not sufficient for what they realize. I take the same considerations, by contrast, to be reasons to reject anti-reductionism. I have argued that the sloppy notion of sufficiency operant in anti-reductive arguments, once pressed, just shows that entities and properties either cannot be held to be distinct from their physical bases, or cannot be held to be coincident with physical conditions that are merely partly sufficient for them. In other words, where a property has wide realizations, the constituents of the entity that bears the property are as wide as the realizers, and reduction proceeds. That is not a matter of bookkeeping, but has significant ramifications for our total view of the world as populated by physical things.

Most importantly, then, working against *Circumstantialism*, is the fact that in sweeping into the circumstances those conditions it takes to be untidy, *Circumstantialism* actually helps the antireductionist sweep bad inferences—sufficiency errors—under the rug. On the bookkeeping metaphor, circumstantialism cooks the books. It is complicit in fraud. The sufficiency error is to infer, from the distinctness of an entity with a narrow set of physical conditions not strictly sufficient for the non-categorical properties attributed to it, a thesis about the nature of the world which is thought to be substantive. The antireductive physicalist's thesis is that everything is physical and yet (somehow) not every thing is a physical thing, to put it starkly. The antireductionist, of course, would like to put it less starkly, replacing the first "is" with heretofore ambiguous dependence language like "supervenes on the" or "is realized by." I

have argued that if it is substantive, antireductive physicalism is not coherent, and if it can be made coherent, it is not substantively antireductive. Of course, we can go on talking like circumstantialists and refrain from the bad inferences, as long as we recognize the sloppiness of our usual non-categorical property attributions. Being inexact, *Circumstantialism* is not exactly *false*; it is perhaps a shorter-hand way of representing the situation. It's just that the sloppy representation easily leads to the sloppy inference. But I take such a point against the theoretic utility of an ontological dependence relation—gerrymandered so as to include the notebook while excluding the museum as constitutive of Otto's belief—to be enough to show that the burden falls to the circumstantialist to say why the necessitarian is misusing the word “constitution” when claiming that, for example, museums are constitutive of beliefs about museums. That is, the burden falls to the circumstantialist to show that constituents can be differentiated in some way that retains that feature of constitution which we do, as far as I can see, overwhelmingly agree on—that constitution is a synchronic ontological dependence relation—and which is reflected by its typical use in the English language.

It is from these considerations that I derive the *Unrestricted Constitutionality* premise as a conclusion. I argued above that nothing differentiates constituents from other members among a set of physical conditions on which an entity depends for its duration. If so, then anything on which an entity depends synchronically is constitutive of that entity. This definition of constitution means that it won't do for the internalist to simply point out that the brain is in causal interaction with its surroundings in order to show that externalists commit a coupling–constitution fallacy. The internalist must either show that causal interactions never constitute anything, which would be hard to do (see Section 8.1), or show that most externalist explananda occur *after* the putative constituent on which they depend.

10.2 Brains in Vats

An intuitive objection to externalism has been looming for some time. It is the objection that the brain *primarily* explains experience. I'll argue that,

depending on how it is understood, the objection fails in one or more ways as a counterargument to externalism, and the reasons it fails are reasons to recognize the explanatory significance of the world for experience in general, and so to think, first, that *online* cognition and veridical experience are paradigmatic of the cognitive and mental. Second, the same are reasons to think that even offline cognition is partly constituted by the world.

Suppose one grants the definition of constitution as offered. Still, one might complain that even among constituents so defined, when it comes to cognition, there are more and less important constituents, to put it roughly. The intuitive objection is thus premised on the apparent fact that cognitive processes can occur, as it is usually put, “offline.” We can certainly think about things that are absent, abstract, or non-existent. We can reason, plan, imagine, and dream in apparent isolation from the world. Even in apparent contact with the objects of experience, we can hallucinate, confabulate, experience powerful illusions, and fail to perceive what others do. Because the brain alone explains these cases, it is thought, the brain primarily explains cognition in general. Of course, the thought that if the brain completely or primarily explains experience in offline and illusory cases, then it does so in general, is true only on the crucial assumption that our full range of experiences are analogous in some important way. Why think they are? Here, I take the idea to be that if we can hallucinate and experience powerful illusions, then it might be that even apparently normal waking experience is more akin to dreaming than it is to experiencing what the world is really like. Indeed, we might *only* be dreaming. The idea that we might *only* be dreaming is the idea with which Descartes began his *Meditations*, and this strangely powerful intuition, that we could be wrong about *everything* we perceive, is brought to life in the film *The Matrix* (Wachowski and Wachowski, 1999) in which whole human organisms are kept in tanks and are plugged in at the bases of their skulls to a machine which provides all the necessary signals to induce an experience of a life more like the one we have (or think we have). The idea is that, for all I know, my mental life might occur in a very different world than the one in which I think it occurs. This is precisely the sort of radical doubt that is widely recognized as either poor evidence of anything, or incoherent and self-defeating. The idea that we might be wrong

about everything infects the basic thought that my mind could be realized by a brain-in-a-vat—indeed, that the realization of my mental life involves a brain at all. In short, it would be very odd to think that we could be wrong about everything *else* our mental lives depend on, but that we are right about our mental lives' dependence on brains, and use those claims to motivate a further claim about the explanatory primacy of brains.

The externalist doesn't face the same problem as long as the externalist denies either one of the things causing trouble for the internalist. On one hand, the first route would be to deny that our mental duplicates could be realized by brains-in-vats. On the other hand, if one accepts the possibility of our having envatted mental duplicates, two subsequent routes appear. One is to deny the explanatory primacy of brains by doubting that we have brains at all. The last route is to deny the analogy between illusory and veridical experience, *despite* the possibility that we can be mistaken in serious ways about the way the world is.

Perhaps surprisingly, I don't think the externalist should deny that it is possible for our mental lives to be duplicated, perhaps even by brains in vats. After all, if anything is, brains are largely composed of the sort of multiply actuatable switches that I have posited to explain multiply realizable kinds. So at least my argument for externalism should be able to countenance the possibility of mental duplicates. And as much as internalism is easy to grasp in terms of the doctrine that my mental life could be realized by a brain-in-a-vat, the two claims are not equivalent. It is certainly a *consequence* of my mind's being just the activity of my brain that my mental life could be realized by a brain in a vat. But internalism is not a consequence of that possibility. However, nor do I think that possibility should take us down the subsequent second route, casting doubt on internalism by doubting everything including our having brains at all. Since that route secures nothing, it doesn't secure externalism either. Rather, the externalist should take the last route. The externalist need not deny the possibility of our having envatted mental duplicates in order to break the analogy between two ends of what the externalist can argue is a spectrum of cognitive success in dealing with the world; whatever it might "really" be beyond our present ken, we are in analogous epistemic predicaments.

However, this point is subtle. For, even if the consequences of generalized doubt are incoherent, the fact that I make serious errors in perception and judgement of the world are a reason to doubt that any given “veridical” experience of the world corresponds to the way the world is. So the concern about “veridical” experience is not so easily dismissed, and there is a bit more work to do.

Notice that for both us and for our putative envatted mental duplicates, the world seems at least to many of us to be such that our mental duplicates could be realized by brains in vats. And yet it seems to both me and my mental duplicate that we are not envatted. It also seems to us that we sometimes dream and have illusions, and sometimes, as we might ordinarily put it, have better contact with reality. One feature of my dreams is that they can be very disjointed. And, *ex hypothesi*, this must be a feature of the dreams of my envatted mental duplicate. He can tele-transport and fly. People, places and things readily transmogrify. Dreams may seem real to dreamers, but can be unstable relative to waking experience. Likewise, our mental duplicates think they both dream and have waking experience, as we think. And it seems to us, envatted or not, that there is, as we might ordinarily put it, some truth of the matter as to our being right or wrong about our non-envattedness. If we seem to have experiences which are in more or less contact with reality, and there seems to be a truth of the matter as to which is which, then we must admit a meaningful distinction between illusory and veridical experience. So, envatted or not, we have a meaningful distinction between illusory and veridical experience.

But hang on. Might we deny the above conditional? Not easily. What could make the distinction not meaningful? The distinction would not be meaningful only if we were wrong about there being some truth of the matter as to which experiences are illusory and which are veridical. But of course there seems to be a difference. In that case, we can be mistaken. But we can only be mistaken if there is a truth of the matter. This is a *reductio* of the claim that there is no truth of the matter as to whether things are as they seem combined with the plausible thesis that there seems to a truth of the matter. Formally, the *reductio* can be rendered this way:

- (1) There seems to be a truth of the matter about whether things are as they

seem.

- (2) There is no truth of the matter about whether or not things are as they seem.
- (3) If there seems to be a truth of the matter about whether things are as they seem when there is no truth of the matter about whether or not things are as they seem, then we can be mistaken about the way things really are.

So,

- (4) We can be mistaken about the way things really are.
- (5) If we can be mistaken about the way things really are, then there is a truth of the matter about whether things are as they seem.

So,

- (6) There is a truth of the matter as to whether or not things really are as they seem.

The above is valid but the first and final claims are contradictory. The problem with the claim that there is no truth of the matter as to whether or not things really are as they seem is a special case of the problem with the claim that there is no truth. If it is true, then it can't be true. It is incoherent. Likewise, we have no way to sensibly make the claim that there is no truth of the matter as to whether or not things are as they seem. It would be no good to claim that there *does not* seem to be a truth of the matter, for it seems to me that there is, and to call me wrong in this case is to call me right. In this case, I can't be wrong. The very fact that there seems to be a truth of the matter about whether or not things are as they seem is a logical guarantee that there is. Of course, it is more coherent to think the way things are explains the way they seem than vice versa. It's natural to think that something's being must be ontologically prior to anything's seeming to be. So the way things seem must sometimes be a guide to the way they are, and that is so because the way things seem depends on the way they are.

Let's consider why that should be the case. If we have a meaningful distinction between illusory and veridical experience, and I've argued that we

do, then, with respect to the seeming possibility of brains in vats realizing our mental duplicates, we have a nested dilemma about the nature of the world and our epistemic predicament in relation to it. Either the worlds in which my non-envatted and envatted mental duplicates differ, or they do not. Presumably they differ, if the possibility of my having an envatted mental duplicate is to have any force whatsoever in sustaining the analogy between illusory and veridical experience as evidence of the explanatory primacy of brains, since, if the worlds do not differ at all, they are the same, and thus cannot be ruled out as ontologically explanatory of sameness of experience. So the worlds differ, on the whole. If they differ, then it is possible in principle for *someone* to discern the difference between envatted and non-envatted worlds. But if it is possible in principle for *someone* to discern the difference between the envatted and non-envatted worlds, then either it is possible for *us* to discern the difference, or not.

Now consider what has to be the case in order for brains in vats to be unable to know that they are envatted, and hence for it to be possible for us to have envatted mental duplicates—after all, *I* don't know I'm in a vat, so neither does my mental duplicate. It is not simply the case that we might never discover something the way that bats might never discover colour. Our technologically extended discriminatory apparatus, whether imagined or real, are growing. A deceiving demon has to tread very carefully here. Recall also, that it's crucial for the internalist that the brain-in-a-vat be a *brain* in a vat, or internalism is false. Thus suppose remembering is a brain-bound process, as the internalist thinks. Since our mental lives are such as to seem *not* to be envatted, envatted brains' storing memories would heavily constrain the way that an evil demon—or God in the Cartesian picture, the machine in *The Matrix*, or the television producers of *The Truman Show*—must interact with the pineal gland, brain, or caged organism they deceive. The demon-producer-machine must remain *faithfully* “deceitful”, lest cracks in the structure of the “spoofed” world be noticed by the caged minds. On a sort of dualist picture, then, the demon is as faithful as God is in upholding the regularities of our experience. He must even fake occasional benevolence in order to pursue his malicious agenda, which is to us, *ex hypothesi*, indistinguishable from God upholding

her laws, pursuing benevolence but sometimes having to allow evil. Now, elsewhere in the causal network as upheld by the being whose ultimate designs elude us, the difference must reveal itself; angels, for example, might know whether or not we are deceived. On a physicalist picture, the mechanisms of the faithfulness by which deceptive neuroscientists or artificial intelligence deceive us presumably depend on further regularities or laws of nature. Their machines must work in a regular and predictable way just so that the machines can be programmed to stimulate my brain while I think I am talking in just such a way as to make it seem that I am talking. In any case, the agent additionally has to have thoroughly constrained capacities to manipulate or otherwise move through its world. In the film *The Truman Show* (Weir, 1998), Truman, a captive and unwitting television star whose surrounds are curated and whose life is broadcast around the clock, is dissuaded in various ways from making certain exploratory efforts, but Truman behaves in unpredicted ways that make the ruse difficult to secure. Truman becomes suspicious about a number of inconsistencies, and an enlightened angel of sorts (a sympathetic co-star) tells Truman the truth. In *The Matrix*, a pill allows the protagonist, Neo, to wake up in his vat. (Here, since the pill is an illusion, the idea can only be that it is a sort of mental placebo. Neo's thought that the pill could have the effect of revealing things as they really are is enough to produce the exploratory motor-neural activities that actually wake Neo in his tank. Or something like that.) But in the case of our mental duplicates, the deceiving demon-producer-machine can afford no such accidents. The deceiver must faithfully preclude the possibility of our emancipation in order to dupe our mental duplicates. But notice that these regularities are themselves discernible to us, particularly on the assumption that we are *brains* in vats with brain-bound remembering. If the fabric of our "illusory" experience is woven tightly enough, then there's an important sense in which the "mistakes" we have no hope of correcting aren't mistakes. Nothing makes them mistakes except the possibility of correcting them because they fail us in some way. But, again, *ex hypothesi*, our mental duplicates' belief that the earth rotates on its axis doesn't fail them any more than that belief fails us. The fully regular and constrained world is good enough to guarantee our knowledge of just that much reality.

On the other hand, suppose the fact of a difference between worlds with mental duplicates entails its being possible for its subjects to discover which sort of world they inhabit. If they make the discovery, then these worlds would no longer be mental duplicates.

The point is that in each case, our overall epistemic predicament is entirely analogous. To recap, in case the difference between the worlds is utterly indiscernible to anyone, ever, there is no mental difference between those worlds—our mental duplicates' beliefs are as right and wrong as ours. In case we are deceived but could never know as much, we can be sure about all that must be the case in order to prevent us from knowing we are deceived, which is everything we know in the vat, including that our perceptual experiences are in differential contact with the world, so to speak. Our mental duplicates dream in disjointed ways and sometimes briefly mistake shadows of shrubs for shadows of dogs, and so forth, but, by contrast, much of their experience is of and made possible by faithful deception. In case the envatted worlds could diverge mentally because of unfaithful tells of truth, the beliefs of the enlightened mind change, but the epistemic predicament doesn't. In Plato's famous allegory in which some people are chained such as to only be able to see shadow puppets cast by firelight onto a cave wall, the freed cave-dweller doesn't initially trust his bewildering sunlit experience. Having noticed our mistakes, we should think our revised beliefs are falsifiable. But we can be certain of our fallibility, and so can our mental duplicates in vats. Again, that we might discover we are brains in vats just means that much of what we believe we've discovered is open to revision. Having revised our beliefs, we can be certain that we are fallible. Either we're not fallible at all, or we can be certain of being fallible, and in that, we're not completely fallible. We're right about that much, in the most absolute possible way. Descartes' certainty that he was a thinking thing is a special case of this point, but we can be sure of somewhat more with respect to our epistemic predicament. So, whatever the world might really be, we have a spectrum of perceptual success in dealing with it.

There's a problem with the inference from the fact that we can be wrong to the worry that we might always be wrong. It is not coherent to think that we could be wrong about everything. So the conditional, "if we could be wrong at

any moment, then we could be wrong at every moment,” is false. But it was a premise in what I imagined to be the argument defending the analogousness of online and veridical experiences with offline and illusory experiences. And that analogy was crucial to the substance of the objection to externalism—the thought that if the brain primarily or completely explains experience in illusory cases, then the brain primarily or completely explains experience in general.

10.3 The Boundaries of the Paradigmatically Mental

Returning to the objection to externalism with which I began Section 10.2, it can be reconstructed this way:

- (1) In offline and illusory cases, the world is not primarily, if at all, responsible for the character of our experiences.
- (2) But brain activity certainly is; where there is no brain, there is no experience that we know of.
- (3) If the brain, not the world, is completely or primarily responsible for the character of our experiences in offline and illusory cases, then the brain completely or primarily explains experience in those cases.
- (4) If the brain completely or primarily explains experience in offline and illusory cases, then the brain completely or primarily explains experience in general.

So,

- (5) The brain completely or primarily explains experience in general.

The conditional (4) is dubious because of the disanalogy between offline and illusory experience and veridical experience, the denial of which is self-defeating. The falsifiability of our beliefs is evidence of a spectrum of cognitive success in dealing with the world, and the spectrum is evidence of the disanalogy between the two ends of the spectrum.

That problem aside, now notice the ambiguities in such an argument in the case of “responsible” and “explains.” The argument is strongest when

“responsible” is read in the sloppy, rather than the strict sense of sufficiency. The externalist denies the premise read in the strict sense, but will readily admit that human experience partly depends on human brains, and that the world (understood as wholly distinct from brains) is insufficient for experience. But even granting (1) read with a sloppy sense of responsibility, a valid further inference to (5) requires that we read “explanation” in the conditional (3) as either incomplete metaphysical explanation or explanation in the epistemic sense. Then, however, explanation in the conclusion must be read in the same partial or epistemic sense. Hence why I render the objection as I understand it using “primarily,” which involves some notion of salience.

But explanation in the merely epistemic sense yields no ontological conclusion, and internalism is an ontological thesis. Why think externalism and internalism are ontological theses? For one thing, because they are in conflict. If externalism and internalism were just claims about what sorts of things are most epistemically salient in explaining our mental lives, rather than physically necessary and sufficient for minds, then both sorts of explanation, or either, or neither, may be salient to an audience, and their claims needn’t be in conflict. A metaphysician might care about both explanations, those in well-understood simple worlds with poorly-understood complex brains might tend to find taking brains apart illuminating, those in dynamic and poorly understood complex worlds with simple well-understood coordinative discriminatory and goal-seeking circuitry might find taking apart the world illuminating, and neither physics nor neuroscience explains anything to someone all-knowing. Here and now, externalism as a merely epistemic thesis is also true. Because minds depend on wide physical conditions, and because those have largely been ignored in contemporary sciences of mind, it is especially salient to correct that. It’s salient if we’re going to be self-reflective about something we do already; engineer our own experience. (And I’m inclined to think we should be self-reflective, a point I’ll assume without further argument here.) The nature of the world is what makes our explanations in the epistemic sense better or worse, and the nature of the mental according to externalism as an ontological thesis is such that, given the prevalence of internalism as a doctrine, externalist explanations in the epistemic sense are likely to be illuminating.

So if externalism as an ontological thesis is right, externalist explanations in the epistemic sense are guaranteed to be appropriate at times such as this, whereas there is no such guarantee for epistemic internalism unless ontological internalism is true.

So the epistemic version of the objection is less than what the internalist needs to show, and the externalist rejects it anyway, because, if ontological externalism is true, externalist explanations will almost certainly be the better ones in some situations.

But what of an ontological reading of the objection? Now, if “responsibility” is read in the strict sense, the externalist can plausibly deny (1). That is, the externalist denies that even “offline” experiences really do occur completely in the brain. Here one only needs to accept externalism about meaning, and take the line on synchronic dependence that I defended above. Put another way, content externalism and *Necessitarianism* together yield the view that the things our mental states are about are constitutive of our mental states. I take externalism about meaning to be independently plausible and accepted widely enough that I need not rehearse arguments for it: intentional states are about what they are, mental states mean what they do, and cognitive processes have the content they have, at least partly in virtue of the world in which they are situated. That is why, in Section 2.3, I lumped the view, which I called “restricted” externalism—that the notebook, but not the museum, is constitutive of beliefs about the museum— together with internalism as regards their notions of constitution. Both hold that some system—the brain in one case and brain–body–technology couplings in the other case—are properly constitutive of mental states or cognitive processes or systems. But both views either fail to motivate the difference between constituents and other parts of the physical bases on which these entities depend, or they make the sufficiency error.

To see this, consider Menary’s claim that, “Active externalism is to be distinguished from an externalism where the contents of beliefs are dependent on my history. The external features are distal, not having a synchronic effect on the organism.” (2010, p. 2) Menary’s characterization of the difference between externalism and active externalism brings to light the sufficiency

error that the latter makes. No waterless hunk of spacetime is sufficient for beliefs about water. Rather, if we say that an organism's belief about water means what it does in virtue of its past, we posit a *temporally* broader entity, encompassing the organism's life history, thus including water as a constituent, synchronic within the relevant temporal horizon. On the other hand, we can say that the belief means what it does in virtue of a *spatially* wider swathe of the world, including water.

This apparent choice between theories of meaning might look like the choice between perdurance and endurance theories of persistence through time, and, correspondingly, eternalism and presentism as theories of time. But that is actually a separate matter, about which I shall say more in a moment. Rather, I think these two broad sorts of theory of meaning aren't in conflict, but have differing explanatory targets. The first is an answer to the etiological question of how I *came* to have the disposition to, for example, seek out water when thirsty. The second is an answer to the question about what instantiates beliefs about water having the property of being beliefs about water at a given time or over a shorter duration. These two answers track two ways for God to create my present belief that water is thirst-quenching. She can do it the good-old fashioned way, and breathe into being all the necessary conditions starting sometime in my infancy or as far back as back goes. Then, even in the absence of present water (if God was to delete it), if we still wanted to say that there came to be beliefs about water, what makes the belief about water is a temporally extended process of which water is a part. Call this belief an "orphan" disposition. On the other hand, God can create my body and water-including surrounding out of nothing beginning a moment ago and instantiate my beliefs about water as functionally characterized in terms of, for example, my disposition to drink water when thirsty. Call this belief a "test-tube" disposition. In answer to either question about what *makes* an orphan belief or a test-tube belief a belief about water, considerations of sufficiency should lead us to conclude that water is a constituent of both sorts of beliefs about water. Even if we say the orphan belief has the property of *having* a certain causal relation to water, what is sufficient to instantiate that property is nothing less than that causal history. Either way, no utterly waterless hunk

of spacetime instantiates the property of being an intentional state about water. The reason a brain isn't sufficient to make a certain belief the belief that it is is multiple causability—the fact that one could have a certain brain state because of something else entirely. What makes the mental state the belief that it is thus involves more of the causal network of which the brain is a part. What we *can't* say is that a given neural state has the property of meaning something for which it is insufficient, and treat the wider conditions which make the belief what it is as irrelevant to constituting it.

Or, not obviously. I've been speaking as if four-dimensional objects exist. Supposing they don't, and thus supposing the present is all there is, the presentist would do well to have a story about how a present world could instantiate beliefs about presently non-existing things. For example, the presentist thinks that my water-including past no longer exists, so that water in my past constitutes nothing present. In other words, the presentist perhaps thinks that four-dimensional objects are really changing three-dimensional objects supervening on agents and dynamic physical conditions. If the present external world is waterless, the presentist explains my belief that water is thirst-quenching by saying that some present, albeit potentially distributed, property of the world is what makes my belief about water mean what it does. Having water in its past is a property the present world has. Or, the world *is* such that water was, among other things, a liquid at standard temperature and pressure, thirst-quenching, etc. But in that case, what is now constitutive of beliefs about water, what makes them now mean what they do, is still a swathe of the now-waterless but nonetheless well-watered world.

Let's consider another example, since beliefs about water are perhaps odd examples. After all, it may be enough that I am largely constituted by water, in the ordinary sense that water is partly constitutive of the cytosol, interstitial fluids, and blood which are, among other things, constitutive of me. In other words, just my body (in the ordinary sense), might be enough world to make my beliefs about water mean what they do. Instead, imagine Otto-and-his-notebook as survivors of an apocalypse that sees the demise of the museum. Do the same considerations apply? Is Otto-and-his-notebook enough world for Otto's belief to mean what it does about the museum? I think not. Consider

the truth conditions of Otto's belief. Whether or not the truth-conditions of a belief are the same as its meaning is a matter I set aside. But the property of meaning x needs physical instantiation, and truth conditions are physical conditions (I take it that "truth supervenes on being" (Bigelow, 1996, p. 38) in one way or another). Moreover, falsity supervenes on being, so that the absence of a truth condition is itself a physical condition. So it is plausible to think that the truth conditions of a belief and the conditions that instantiate its having the property of being the belief that it is might overlap. And the truth-conditions of post-apocalypse-Otto's belief about the museum's past location are broader than his notebook. Is the claim that the Museum was on 53rd true just in case a notebook makes it? Is it true just in case millions believe it? No. There are obviously many ways to make a notebook say that the museum is on 53rd. But there is no way under determinism to bring about the *entire* post-apocalyptic world without also making it such that the Museum was on 53rd. In the post-apocalyptic world, whatever else might eventuate, there will be effects of the Museum's having been where it was. For example, light reflected from its roof will still be travelling toward Alpha Centauri. So the claim is true, on the presentist's account, just in case *all* of the relevant portion of the present world over which the Museum's location had any influence is such that the Museum was on 53rd. A world in which Otto's notebook is wrong can be brought about under determinism from a world in which the Museum never existed, but no world can be brought about under determinism which is such that the Museum was on 53rd when it never was. If the truth conditions and the instantiation of the belief's meaning what it does are physically overlapped conditions, then, the idea is that some portion of *that larger* world constitutes the belief's having the meaning that it does. I don't think I need to say what portion. Semantic externalism is first and foremost a claim about where meaning ain't, to play on Putnam's famous claim that meanings "just ain't in the head!" (Putnam, 1975b, p. 144) On any conception of how intentional states mean what they do, nothing purely neural will suffice to explain (in the metaphysical and thus also the epistemic sense) how an intentional state means what it does.

Again, *Unrestricted Constitutionalism* as a form of *Necessitarianism* about meaning entails that the constituents of beliefs are more widespread than

brains. Recall now from Section 2.1 the idea presented by Clark and Chalmers (1998) that a process or state should be considered “cognitive” or “mental”, if some functionally equivalent process or state in the brain would be considered cognitive or mental. In fact, an argument for externalism doesn’t require that premise. Although it would depart from Clark and Chalmers (1998), given their appeal to the functional parity of extended cognitive processes with (putatively) brain-bound ones, an argument for externalism might just as well begin with the premise that there are cognitive process and mental states, full stop. In a sense, the parity principle concedes too much to internalism by allowing putatively brain-bound processes to set the standard for what is cognitive (e.g. Gallagher, 2013). Instead, as Menary (2010) summarizes, second-wave arguments have tended to focus less on parity and more on “complementarity”—the idea that extended cognitive processes may be functionally novel. We’ve now seen two compatible ways to deny that paradigmatically cognitive processes occur in the brain. The first, as we saw in Section 10.2, is to deny that processes putatively occurring in the brain are *paradigmatic* of the cognitive. The second is to deny that such processes really do occur completely in the brain, which is a consequence of *Necessitarianism* about meaning.

Let me make a few final notes in defense of this view. The claim that there are cognitive processes and mental states, full stop, without any mention of where they occur, somehow appears to be *less* self-evident than the version of the premise inspired by Clark and Chalmers (1998) which linked cognitive processes and mental states with brain processes and states and set up the parity-based argument for *Extended Mind*. At least, a concern resurfaces about the notion of scientific kindhood in play—the other sticking point in the debate over *Extended Mind* that we saw in Section 2.1. All I have argued is that functional kinds, and *mutatis mutandis*, other non-categorical properties, can be projectible, and that we needn’t take an eliminative stance toward them. Still, eliminativism about *particular* kinds and properties isn’t, on the account of the reducibility of functional properties I have given, in general tension with that account. Rather, whether a given kind or property can be reduced or should be eliminated is a matter of the particular kind or property. But that makes the burden of disqualifying extended members of mental or cognitive kinds

heavier. The physical nature of similarity as I understand and have defended it means that internalists can't make a general argument for eliminativism about extended cognitive systems on the basis of the supposed homogeneity of natural kinds, and then claim that only brain states form a natural kind and as such are properly cognitive and mental, as Adams and Aizawa (2010) try to do. This would eliminate a working concept and recycle the terminology unjustifiably, as if to change our existing concept into something more respectable, and it would fall short of showing that, as the concept now stands, the mental is not projectible, and that what is projectible does not extend. Rather, I think we have reason to see human–technological couplings as partly constitutive of our mental lives and as projectible: after all, wholes with mass-produced or stock-standard parts are at least as similar and behaviourally stable as their biological parts, as I noted in Section 2.1 when challenging Adams and Aizawa's (2001) speculation that human–tool couplings could not figure as objects of science because they would work differently than biological objects of science.

Still, countervailing intuitions persist. Failing to disqualify externalism on the basis of projectibility, we might wonder if all the same considerations could be turned into a *reductio* in favour of eliminativism about any notion, including “mind”, that admits of strangely extended members—as kinds characterized by non-categorical properties do—by adding some premise to the effect that *Necessitarianism* and *Unrestricted Constitutionalism* specifically are just plain ridiculous. The thought here is that *of course* any respectable properties must belong to neatly circumscribable or easily appreciable spacetime regions. I've been tempted by this line of thought, but ultimately, I am not moved by it. I'm compelled to reject the implicit premise that would complete the *reductio*. The idea that respectable properties must belong to neatly circumscribable regions would eliminate all non-categorical properties, and is thus far too damning of what are, if existing scientific practice is anything to go by, darn good objects of scientific enquiry and explanation, from fragility to beliefs. Here, I'm echoing Fodor (see Section 5.3), without thinking it's just a mystery that different things can (partly) be corkscrews or intentional states. Actually, what explains their being so are wider circumstances. *Necessitarianism* is vindicated by its explanatory power in precisely those cases of projectibility that eliminativism

would like to ignore and that antireductionism would invalidly premise itself upon.

10.4 Biting a Little Bloat

Necessitarianism and *Unrestricted Constitutionalism* stand. Failures, like the ones we saw in Section 10.1, to provide a clear basis for isolating the ontologically core constituents or realizers of non-categorical property-bearing higher-level entities, suggest that the complete constituents or total realizers of our mental lives are instead divisible in terms of epistemic and pragmatic salience to us. And this is exactly what the externalist should say in response to the so-called “bloat” objection. That objection is roughly this:

The coupling argument brings out the extent to which the mark of the cognitive bears on the bounds of cognition. If Clark and Chalmers opt for the simplistic view that anything that is causally connected to a cognitive process is part of the cognitive process, then there is the threat of cognition bleeding into everything. This is sometimes called something like “the problem of cognitive bloat” or “cognitive ooze.” These names do justice to the ugliness of the view, but not to its radical nature. The threat is of pancognitivism, where everything is cognitive. (Adams and Aizawa, 2001, p. 57)

The answer is to bite the bullet, though not quite one of the calibre shot above. Note once more that extended mind theorists make no inference from a thing’s being cognitive and interacting with another thing to the latter’s being cognitive (see Section 2.3). So let’s be clear that “everything is cognitive” does not follow from cognitive things’ being identical with, constituted by, or realized by physical conditions more broadly distributed than traditionally supposed. For one thing, being cognitive is a property of a whole for which each of its parts are individually insufficient and thus do not in any sense have. Not every part of a big red thing is big and red. Likewise, pencils are not cognitive even when coupled to the mathematician. Nor are neurons cognitive. Detached brains are not cognitive, and for that reason, even calling brains in general cognitive is just as sloppy as calling pencils cognitive would be. That aside, the argument put forward here—that anything on which an entity depends

synchronically is constitutive of that entity—nonetheless sucks up potentially large swathes of the physical landscape. However, not only is every part of a cognitive thing not cognitive, there are also plenty of things obviously not part of present cognitive processes or mental states, even if only because they are outside the rearward light cones of present mental lives. What’s happening today at the center of Alpha Centauri, 4.37 light years from Earth, is of no present consequence, although in 4.37 years, light from those events might come to partly constitute someone’s (somewhat mistaken) thought that they are small. In fact, the ingredients for just that thought are staggering relative to everyday appreciation, if nonetheless dwarfed by the cosmos. While there is no threat of the pancognitivism described above, there is a sense in which the externalist should bite the bullet. Indeed, the point is that our everyday property attributions involve a sort of inaccuracy which leads to confusion if we take them at face value. No wonder the view on the table is counter-intuitive—that much the externalist should unflinchingly admit. The sufficient physical conditions for Otto’s belief about the museum as functionally individuated include not only the notebook, as Clark and Chalmers argued. For the same reason they include Otto’s notebook, they include the museum. They’re both on the metaphysical ingredient list, so to speak, of the property in question.

We can explain our intuitions about the core realizers of our lives rather differently. Mental constituents may differ in their epistemological status, relative to an enquirer, without differing in their ontological status relative to the thinker. Recalling Lewis’s liberal position on the causes of an event, let me once more adapt something he says about intuitively strange causes, such as absences. Lewis (2000) says,

One reason for an aversion to causation by absences is that, if there is any of it at all, there is a lot of it—far more of it than we would normally want to mention. At this very moment, we are being kept alive by an absence of nerve gas in the air we are breathing. The foe of causation by absences owes us an explanation of why we sometimes do say that an absence caused something. The friend of causation by absences owes us an explanation of why we sometimes refuse to say that an absence caused something, even when we have just the right pattern of dependence. I think the friend is much better able to pay his debt than the foe is to pay his. There are ever

so many reasons why it might be inappropriate to say something true. It might be irrelevant to the conversation, it might convey a false hint, it might be known already to all concerned. . . . Of course, such reasons for refusing to say what is true are not confined to causation by absences. “Counterfactual analysis of causation?—Yeah, yeah, my birth is a cause of my death!” said the scoffer. His birth is indeed a cause of his death; but it is understandable that we seldom say so. The counterfactual dependence of his death on his birth is just too obvious to be worth mentioning. (p. 196)

Interestingly, Lewis’s example of “being kept alive by an absence of nerve gas in the air we are breathing” is arguably an example of synchronic dependence. Or, whether the absence of nerve gas is a cause or constituent of my being alive hinges on clarifying the temporal extent of my being alive as the explanandum. The absence of nerve gas throughout my writing this chapter is a constituent of my writing this chapter. The absence of nerve gas at my birth in 1982 is a cause of my writing this chapter.

Sometimes citing absences sounds odd, sometimes not. We wouldn’t normally say what I just did about this chapter, but we would say someone’s absence from work gets them fired. That absence is quite real—it is just the same as that person’s being at the golf course. In this case, we cite the absence because it doesn’t matter whether the truant was at the golf course or the bar. Citing the absence of nerve gas is easier than listing everything that is here instead. And if doing so sounds odd, it isn’t because absences are unreal, but because so many absences go without mentioning. Where I am, there isn’t normally nerve gas all over the place, whereas employees are expected to go to work. And as with absent causes, we ignore all sorts of synchronic dependencies that are obvious. What’s interesting—what makes externalism both important and controversial—is that this tendency to ignore constituents is so pervasive as to make unfamiliar dependence claims sound mistaken, when by all accounts of constitution as synchronic ontological dependence, they can in fact be true. In some ways, it is the unrecognized brick-and-mortar foundations of our particular mental lives that are the most crucial to bring to light.

Wrapping Up

In this chapter, I began by arguing that there is no clear competitor to *Unrestricted Constitutionalism*. I've defended that view against an intuitive objection which fails to show that the possibility of our being brains in vats means we could be wrong about everything and that the brain thus explains experiences in general. Further, I've deployed *Necessitarianism* about meaning to challenge the assumption that even offline/illusory experiences are *completely* explained by the brain. The upshot is that externalism need not claim that paradigmatically cognitive processes and mental states occur in the brain. While that is counter to existing assumptions, I noted that kinds characterized by non-categorical properties, such as the mental, the cognitive, and minds, in contrast to eliminativism about the latter, are already well established in science. Finally, I've argued that externalists should bite the bullet with respect to the "bloat" objection, and account for the counter-intuitiveness of their view and the intuitiveness of internalism in terms of differing epistemic salience of conditions within the ontological dependence bases of higher-level things. Of course, the brain is justifiably at the bleeding edge of a rather wider epistemic frontier of explaining our minds. Indeed, I think this is what explains our having certain intuitions, which should not be taken as a total guide to the ontological facts, about the core realizers of our minds. It's perfectly sensible to go on studying the neural parts of the mind. It's not, however, sensible to claim that the mind is the brain, nor by contrast to claim that because it's not, that the mind is not identical with any physical conditions. And because individual minds, and mind as a kind, *are* identical with physical conditions, minds needn't be eliminated from our ontology. Likewise, I take it that there are other *interesting* subsets of the complete realizers of our mental lives. As I argued in Section 10.2, internalism must be an ontological thesis. And if externalism, as an ontological thesis, is right, then just about any of the actual physical conditions on which our mental lives depend might be interesting. We'll explore one such subset in the next chapter.

Chapter 11

Core Selves, Dynamic Selves

Clark and Chalmers (1998) say,

Does the extended mind imply an extended self? It seems so. Most of us already accept that the self outstrips the boundaries of consciousness; my dispositional beliefs, for example, constitute in some deep sense part of who I am. If so, then these boundaries may also fall beyond the skin. The information in Otto's notebook, for example, is a central part of his identity as a cognitive agent. What this comes to is that Otto himself is best regarded as an extended system, a coupling of biological organism and external resources. To consistently resist this conclusion, we would have to shrink the self into a mere bundle of occurrent states, severely threatening its deep psychological continuity. Far better to take the broader view, and see agents themselves as spread into the world. As with any reconception of ourselves, this view will have significant consequences. (p. 18)

In other words, if technologies can partly constitute human cognitive processes, Clark and Chalmers (1998) think they can partly constitute human minds, and thus human selves. Similarly, this project began with the thought that human beings are creatures with technological parts. We are technological beings, with mental lives partly constituted or realized by technologies. From the outset it was clear that a way to show that technologies are part of us was to show that technologies are parts of our minds, given the plausibility of a strong relation between our minds and selves. Owing to that plausibility, I have spent the most time on what I argued in Chapter 2 are the underlying issues with the idea that our minds have extraneural constituents.

In Chapter 2, I sought to characterize the argument for *Extended Mind* and externalism generally and show that that disagreements about externalism are cases of deeper philosophical disagreements which infect philosophical debates more widely. Those disagreements involve diverging notions of scientific kinds, and issues around the synchronic dependence relation or relations that obtain between higher and lower-level entities and properties—whether identity, constitution, realization, or composition. I have since been arguing for *Reducibility* and *Unrestricted Constitutionalism*. That is, I’ve advocated the identity of higher-level entities and properties with everything they depend upon, which includes anything without which they would not be the entities that they are or have the properties they have. The dependence of entities and their properties on physical conditions fails to carve out just those conditions we typically attribute many properties to. This view has the counter-intuitive consequence that minds and selves, as special cases of higher-level things, are not bounded within those conditions we typically think, since they are bearers of a great many non-categorical properties. Non-categorical properties can nonetheless be projectible, and we needn’t take an eliminative stance toward them in advance.

I now want to speak about our selves, if only because it is natural to speak about persons or selves when it comes to ethical matters. To that end, this chapter is intended as an epilogue loosely tying the issues I’ve covered in detail to the question of who we are and to those ethical issues that motivated the project at the outset. I say, “loosely,” not because there is no tight connection to be explored, but because I have invested the most time on those issues I thought would most advance the debate. In closing, then, I will quicken the pace at the unavoidable expense of depth. I hope I have moved carefully enough to this point; I know I have moved slowly and I am grateful to my readers for having made it to this point.

We’ll see in Section 11.1 that there must be a very close relation between our minds and our selves if our selves are to be targets of praise and blameworthiness at all. I think we *are* minds, and our minds—whether at any one time or over time—are identical with extensive physical conditions that include technologies. Before turning to consider the ethical ramifications of this view, I make one further argument that the biological body isn’t the core realizer of mind and self.

Ubiquitous and thus easily overlooked phenomenological considerations as to our personal boundaries are at odds with the conservative notion of the human animal as coincident with its biological tissue. I argue in Section 11.2 that phenomenology suggests that our bodies are dynamic—supporting the extended thesis. The embodiment and embeddedness of minds, together with the thesis that our bodies are dynamic, yields a picture of minds as dynamically embodied and embedded. Our particular mental lives would fail to be duplicated if our biological duplicates were to be transplanted into different technical carapaces and wider, partly technological circumstances. In every case of psychological continuity there is also circumstantial continuity constituting a person’s being of a kind with their past and future selves. Tissue-like technologies, as life-supports, may be proper objects of moral concern.

Likewise, in Section 11.3, I suggest that it is plausible and worthwhile to distinguish another morally salient subset of the complete constituents or total realizers of our minds and selves from the rest. Technologies are among those special constituents of our mental lives which we design. That is, technologies are among both the constituents of our mental lives, and among those things we have responsibility for. While this subset of the constituents of our mental lives does not map cleanly onto the conditions necessary and sufficient for it, such subsets may help to explain our circumstantialist intuitions. Though explicable, these are the intuitions whose grip on us I hope to have loosened in the previous chapters, clearing the way for a fresh perspective on our minds and selves. I close with a few ruminations about the moral implications of our sizes as extended (and overlapped) beings.

11.1 Dynamic Selves

We usually say we *have* minds. Thus, assume for the moment that we are not just minds. And it is natural to speak of persons, rather than minds, when it comes to ethical matters. A common theme among theories of personal identity is that selves or persons are the bearers of rights and responsibilities and the objects of praise and blameworthiness. A prevalent “methodological assumption that has been retained by most theorists on identity and ethics

since [Locke],” is that “a theory of [personal] identity’s plausibility depends significantly on how well it accounts for our practical concerns” (Shoemaker, 2016, sec. 1). Our normative practices serve as either a kind of test for the plausibility of any such theory, or perhaps as the very grounds of the “identity” relation in question. Of course, this connection between persons and ethics is why I’m interested in the idea that we are partly what we make: technologies are an ethical matter.

The inference from our minds’ partial constitution by technology to the conclusion that we are partly what we make, would only require that we are at least partly our experiences—that we are at least partly mental. Indeed, there must be a strong mind–self relation if selves are to be the objects of self-concern and of praise and blameworthiness. To see this, note that on most accounts, moral responsibility presupposes conscious agency. Locke (1841) famously linked personal identity with ethical and practical concerns when he defended, according to Shoemaker (2016), “a crucial condition of moral agency, namely, self-reflective consciousness” (sec. 1). And although the debate on moral responsibility, according to Mackenzie (2014), tends to focus on “identifying the necessary and sufficient conditions, with respect to an agent’s current motivational structure, for a particular choice or action to be autonomous, or one for which the agent should be held responsible,” Mackenzie points out that “our actions and choices at a particular time are intelligible only in the context of our personal histories,” and that these “are also directed towards a future that we aim to realize through agential activities like planning. . . .” (p. 154). So, moral agents are those which are autonomous and have some continuity between states or stages of conscious agency. Persons or selves must have—or be—minds if persons or selves are to be the objects of ethical and practical concerns. Otherwise, we should stop speaking about persons in connection with these matters, and just speak about minds. If that were the norm, the conclusion that our minds are partly what we make would be sufficient to highlight the strong connection between technologies and ethical matters. But speaking about minds as the objects of ethical and practical concerns is not the norm.

It thus seems that I should say *something* more about the relation between

minds and selves. However, there is an extensive literature on personal identity. Having focused on the extended mind and issues having to do with multiple realizability and synchronic dependence, it would be difficult to do justice to the personal identity literature, despite the fact that the issues overlap. Any analysis of selves as higher-level things requires the very tools I have attempted to sharpen. But that analysis must be left for another occasion. So, rather than conducting an extensive survey of the literature or constructing a careful account of selves, what I do want to draw out of the literature is the point that the substantiveness of certain debates about what we are presupposes mind–body distinctness. For example, the apparent conflict between the broadly “Animalist” claim that we are human animals and the broadly “Lockean” claim that we are the minds of human animals is greatly softened or dissolved if the minds of human animals are identical with the physical conditions on which they depend, and if these conditions significantly or completely overlap with the conditions on which being a human animal depends. Likewise, as a subset of Lockean views, the broadly “Narrativist” claim, that the self is a sort of mental construct, is at odds with other views only on the assumption that mental content is distinct from its total vehicles—an assumption I hope to have undermined.

To see this undermining at work more specifically, it will be instructive to consider the features of broadly Lockean and Animalist positions that appear to generate tensions between them. On Lockean views, I am the “same” person as I was yesterday and will be tomorrow just in case those persons are uniquely psychologically continuous with one another, “where psychological continuity consists in overlapping chains of strong psychological connectedness, itself consisting in significant numbers of direct psychological connections like memories, intentions, beliefs/goals/desires, and similarity of character” (Shoemaker, 2016, sec. 2.1, summarizing Parfit, 1986, p. 207). Now, most parties to the debate, including Lockeans, think minds depend on physical conditions, brains not least. So one objection to psychological criteria is that, “if Lockeans distinguish between persons and human animals, their view implies that whenever any person thinks some thought, a human animal also thinks this thought,” (Parfit, 2012, p. 7) which generates too many thinkers, too many

persons, attending overdetermination problems, and the problem of knowing which we are. Lockean views supposedly have this problem because,

If persons are, in the Lockean sense, entities that can think about themselves, and whose continued existence essentially involves psychological continuity, a human embryo or fetus is not a person. But this fetus is, or becomes, a human animal. This animal's body, Lockeans claim, later becomes the body of a Lockean person. . . . It would be convenient for Lockeans if this animal retired from the scene, by ceasing to exist, thereby leaving its body under the sole control of the newly existing person. But that is not what happens. Most human animals continue to exist, and start to have thoughts and other experiences. (Parfit, 2012, p. 7)

But, as Parfit (2012) notes, Lockeans can claim that persons are “phase sortals” of human animals in much the way that teenagers are. Sortals provide a criterion for what is to be counted, and phase sortals apply to something during part of its existence (see Grandy, 2016). Alternatively, a certain kind of Lockean can claim that human animals constitute persons under certain circumstances (see Baker, 2000). For Baker (2000), a person “has causal powers that a body would not have if it did not constitute a person” (p. 109). Accordingly, there should be just one set of causal powers here. That phase sortals are not causal competitors with what they are phase sortals *of* is perhaps most easily seen on a four-dimensional model. The partial spatiotemporal coincidence of the parts of a human animal with the parts of a Lockean person for its duration is no more a problem than the coincidence of two streets at their intersection as far as their slippery pavement's being causal competitors in a traffic accident, or between a movie and one of its scenes as far as being causal competitors for a certain outburst of laughter. Parfit (2012) himself thinks Lockeans needn't retreat to this line, however. Instead—although, I would argue, analogously—Parfit (2012) thinks Lockeans can claim that we are *the thinking parts* of human animals. Again, because there aren't too many digesters when an animal digests food by having digestive parts, there aren't too many thinkers when an animal thinks by having thinking parts.

These moves, however, highlight the main objection to Lockean views. The Animalist now objects that we aren't *essentially* Lockean persons. As

Shoemaker (2016) puts it, “I am many things, including an adult, a professor, a driver, a voter, and so forth. None of these is my essence, however, for I either did or could exist without being them,” and,

just as I was once a teenager, and before that an adolescent and a child, wasn’t I also an infant, and ultimately a fetus? Furthermore, suppose I were in a horrible accident and went into a permanent vegetative state (PVS). Wouldn’t *I* then be in a PVS? (sec. 2.2)

The same sort of objection applies to the constitution view. And one might object to Parfit’s (2012) “thinking parts” view in an analogous way, asking, “is it really a serious view” that we are “about four inches tall and weigh less than three pounds”? (Olson, 2007, p. 76). Likewise, to the claim that we are minds, the objection as summarized by Parfit (2012) runs, “we can’t be our minds, since we have arms and legs and our minds don’t” (p. 19).

Animalism, then, is the view that I am the “same” person as I was yesterday and will be tomorrow just in case my past, present, and future biological organisms are continuous with one another. The apparent trouble with Animalism is that it runs counter to intuitions about what would matter to us in imagined transplant, transport, upload, and duplication cases, which variously ask us to imagine our psychologies realized by a body different from the one we have now. Animalism seems to imply that if my brain was to be put into your body, upon waking you would think you were me, but you would be mistaken. “As many Animalists concede, this widely held belief, which some call the *Transplant Intuition*, provides a strong objection to their view” (Parfit, 2012, p. 10). Animalists are forced by their transplant intuitions to concede that identity might not be what matters ethically: “the Biological Approach does have an interesting ethical consequence, namely that those practical relations are not necessarily connected with numerical identity” (Olson, 1999, p. 70).

Notice that I’ve been putting “same” and “identical” in quotes when discussing “identity” over time. That’s because there’s an important distinction between synchronic identity and putative “diachronic identity.” No one thinks a person at various times is the same in the sense of obeying Leibniz’s Law, the indiscernibility of identical things which have all the same properties. But some theorists do think there is a special kind of identity which is a sort

of “numerical” identity such that a person at one time can be “one and the same” as a person at another time. Shoemaker (2016) refers to “a criterion of *diachronic numerical identity*, a criterion of what makes something one and the same thing as itself at different times” (sec. 2.3). And Parfit (1986, p. 201, 2012, p. 5) explicitly holds that numerically identical things may be qualitatively distinct. Gallois (2016) says,

To say that *a* and *b* are qualitatively identical is to say that *a* exactly resembles *b*. To say that *a* and *b* are numerically identical is, at least, to say *a* and *b* are one thing and not two. Whether *a* and *b* can have all their qualities in common without being numerically identical is controversial. Nevertheless, it seems that *a* and *b* can be numerically identical without being qualitatively identical by having different qualities at different times. (sec. 1)

But this last claim is in clear conflict with another definition, according to which, “Numerical identity requires absolute, or total, qualitative identity, and can only hold between a thing and itself” (Noonan and Curtis, 2017, sec. 1). Now, although we can still formulate puzzles about persistence conditions of entities in other terms, this particular issue is merely terminological. The relation that obtains between me now and the baby I was is quite simply something other than numerical identity in the strict sense. For that reason, speaking of personal “identity” can simply be misleading. Strictly speaking, identity is not at issue. And clearing that up alone may help to clear up the apparent tension between views about what we are. For example, when I say, “I was once a baby,” I say something that looks like an identity claim. If it is the claim that, “I presently have the property of having been a baby,” we’ll say it’s true, but it makes no claim that a relation of numerical identity in the strict sense obtains between the person I am now and the baby that the person I am now was. Tenses do the required job of noting the passage of time and change. On the other hand, if “I was once a baby” is the claim that I am strictly the same as the baby I was, then it is false.

Note, too, that it may be plausible that differing practical and ethical concerns would tend to track differing “units”—entities differing in their properties, which is to say, partly in their spatiotemporal extents. For example, certain other-regarding concerns may tend to track something like “person-lives” (see

Schechtman, 2014), conferring rights on fetuses, young babies, and human organisms in permanent vegetative states, as a way of protecting not a present but a possible future Lockean person. Shoemaker (2016) calls the relevant units, “living human beings”, “Lockean persons”, and “selves,” and additionally calls momentary experiencers “atoms” (sec. 3). Many “self-regarding” concerns (working, planning, delaying gratification), tend to track a Lockean person—an entity consisting of states which are psychologically continuous in some way. The latter concerns are sensible, assuming it is sensible to care about other persons at all, and given that I am uniquely epistemically positioned to knowledgeably influence the experience of future participants in a particular person-life (the one of which I am a part). But the explanation for that unique influence is causal. I, now, needn’t be strictly identical with any future self. And, in fact, “self”-regarding concerns seem to differentiate any overarching entity from its smaller temporal parts, such that we distinguish and differentially treat today’s self from tomorrow’s, and each of these from a distant future self. Each instance of any such unit, such as a person-life, or any other whole temporally extended four-dimensional entity, is unproblematically identical with its whole self in the strict sense. But each of that whole’s temporal parts are only part of the whole. So I can’t be strictly identical with a future person in virtue of being part of a whole life of which it is a part. The relation in question—whether between temporal parts, or between phase sortals, or between what is constituted at one time versus another—is not strict identity.

I said above that puzzles about persistence through time can be formulated even once we insist on a particular use of the term “identity.” For example, about putative four dimensional entities, we can ask a question analogous to van Inwagen’s special composition question (see Section 4.3). That is the question of when or why the composition relation obtains between parts and some whole. Why take that particular spacetime worm, Paul Hubble, and not a different spacetime worm, like my left half plus the moon from 1992 to 1994 combined with your pet rat’s yesterday afternoon nap, as an object of explanation, prediction, and practical or ethical concern? We can also ask, what makes distinct phase sortals sortals of *one* other thing? And Olson (2007) raises a grounding problem for the constitution view of persons, asking,

If the same atoms can compose two things at once, what could make those two things qualitatively different? What could give them different mental properties, or different persistence conditions, or different modal properties? (p. 63)¹

Thus, for each way of distinguishing between Lockean persons and animals, there is a grounding problem. If Lockean persons are phases or temporal parts, or spatial parts of animals, then the distinction is clear enough, but we can ask what makes them parts of one thing at all. If Lockean persons are constituted by animals, we can ask what makes them two things at all.

But the grounding answers to each either trivialize or undermine the distinctness. First, in answer to the special four-dimensional composition question, there needn't be any metaphysical restriction on composition such that we insist there is no such entity composed by an assortment of spatiotemporal parts of me, the moon and an event, in order to give an answer as to why we ignore the latter's existence while caring about, for example, a person-life. Rather than strict identity, one of the relations between me now and me yesterday is similarity, which can be cashed out in terms of dispositions to have one or more of the same effects on various multiply actuatable switches (see Section 6.1). Similarly, Millikan (2017), for example, thinks of changing individuals as "historical" kinds, members of which cluster closely in property space. The temporal parts of Paul Hubble, taken as a person-life, fall under a kind and are projectible in various ways. So the four-dimensional object they compose is worth tracking. And some subsets of those temporal parts (say, my teenage years) fall under more than one kind and are even more projectible. These temporal parts also stand in unique causal and epistemic relations to one another: again, I am uniquely epistemically positioned to knowledgeably influence the experience of future participants in the person-life of Paul Hubble.

Grounding answers that can be given about constituted things similarly trivialize or undermine the distinctness of Lockean persons and bodies. Although

¹ Olson (2007) also raises other grounding problems. He asks, "What could give us identity conditions different from those of human animals?" (p. 63). He goes on to characterize two further objections to the constitution view. He says that it has no satisfactory answers to the questions, "under what circumstances does constitution occur?" (p. 66), and, "The question is now what makes it the case that animals constitute us. Why not things a bit larger or smaller than animals?" (p. 73). All of these, I think, are grounding questions.

the constitution view of persons—as a species of *Circumstantialism*—provides no answer, according to the *Necessitarianism* I have been advocating, the same atoms can't compose two things at once without instantiating all the same properties. On one hand, the realizations of differing property instances have differing physical extents. On the other hand, any thing to which we attribute more than one property instance is one and the same as the complete realizers of those properties.

What grounds the similarity among temporal parts, what instantiates the sortalish properties of a human animal's being a person at a time, and what makes it the case that a person and human animal coincide (to the degree that they do), are in each case those conditions usually relegated by *Circumstantialism* to the background.

Putting this together and returning to the apparent disagreements between Animalists and Lockeans, we can now see that when we distinguish between a partly coincident or *putatively* coincident Lockean person and a human animal (recalling that the boundaries of each are in dispute), we are not multiplying entities, but cross-classifying things. The mere fact that the class of Lockean persons is not identical with the class of human animals does not mean that a certain Lockean person is not identical with a particular human animal—that is a question of their properties.

If we make the Animalist move and claim that I am not a Lockean person because I was once a baby, we use “I” to refer to something with the property of having been a baby. But if the animal-in-the-present-conditions is sufficient for such a property and is also sufficient for being a Lockean person, then the “I” refers to something with both properties. On the other hand, if we insist that “I” refers to the temporally extended animal life which has a baby as a temporal part, then we should also say that I am partly a Lockean person. Either way, when we say we were babies, I don't think we make identity claims between our now-selves and the babies we were. I think we use “I” in the former way, to refer to something presently instantiating the property of having been a baby. But if I am wrong, that is a linguistic matter. Either way, we are animals and Lockean persons (thinkers).

If we make the Lockean move and claim that I am not an animal because I

would survive various transplants that my body would not, we either use “I” to refer to something with certain mental properties (beliefs, desires, goals), or to refer to a temporally extended thinker. If it is the former, then the physical conditions instantiating those mental properties presumably include, among other things, a particular bunch of biological tissues. I am partly those tissues. If it is the temporally extended thinker, that thinker once again has biological tissues as a temporal part, even if it has other physical conditions as temporal parts, and the temporally broader entity is partly those tissues. Either way, we are thinkers and animals.

To see this point in action, consider whether standard transplant intuitions would withstand the physical details of such procedures. Thus imagine transplanting my brain. It could well turn out that, after overcoming the challenges of wiring everything up correctly to compatible tissue, a period of intense therapy would be required to remap and recalibrate schemas and procedures in the brain to work with the almost certainly different host of bodily inputs. It is not hard to imagine that the initial state of this mind would be utter confusion. People who have just one sense modality restored often find the resulting cacophony of stimuli overwhelming. Even if this brain had all the plasticity of a baby’s brain, it might take years to acclimate to the new body (as it does for babies, who don’t remember these developments). Meanwhile, memory traces, if ever reconstructed, might seem like recollected dreams or *déjà vu*, hardly straightforward replicas of the experience of “this same person.” A new person, strictly speaking, would be born in the process. And the difficulty of transplanting our minds suggests that if we do value psychological continuity so highly (even if all we’re really doing is caring about a different experiencer whose pain and suffering we have particular influence over), we should value the extended bodies and communities and context on which psychological continuity depends. I mention communities and context, because the best hope of psychological continuity through transplant or transport would be one’s community’s support through rehabilitation and reaffirming or rebuilding one’s sense of past.

Notice, too, that a body, my body at that time—the physical conditions instantiating my mental life at that time—*does* survive. Even brains in vats

have to be hooked up to a world to be persons. And my new body is the “same” as my old one in the same way that a thing at one moment is ever the same as a thing at another moment. They are similar enough to be projectible. *Ex hypothesi*, the progenitor and inheritor of a mind share mental properties. And they are in causal continuity. After all, there are mechanisms involved in copying, uploading, or otherwise transplanting the mind. There’s no metaphysical difference between the continuity of a mental life during well-known metabolic turnover and the continuity of a mental life during, as I shall now put it, technological turnover. And the possibility of being not just transplanted but duplicated wouldn’t change anything, except that, prior to being copied, I would have two people’s futures to worry about, and after being copied, two people would praise or blame me for certain features of their lot in life. Analogous biological processes would have the same consequence. Suppose we reproduced by asexual budding and inherited, along with our parent’s DNA, a copy of some of their memories, desires and beliefs. These offspring would be *very* similar to their parents, without strict identity being at issue, whether the parent continued to live or not. Again, the claim cannot be that, when a mind is transplanted or copied or otherwise changed in some respect, the inheritor(s) are *the same*, in the strict sense, as the progenitor(s). They are just that similar.

The upshot here is that, as a special case of the mind–body problem, the apparent choice between physical or mental continuity as the grounds for our self and other-regarding practices is premised on construing an overly narrow physical basis for minds. So-called “transplant” thought experiments are supposed to show that psychological and not biological continuity is what intuitively matters to our self and other-regarding practices. What transplant thought experiments actually show—in showing that psychological continuity is what matters for self and other-regarding practices—is not that psychological continuity can be divorced from physical continuity generally, but that the physical continuity on which psychological continuity depends is broader than the biological animal.

11.2 Dynamic Bodies

I have suggested that we, our selves, are embodied minds, although as a physicalist, I take it that the “embodied” bit is redundant. The extent of that embodiment has been the topic of the thesis to this point. The properties we attribute to our minds, mental states and cognitive processes, are instantiated more broadly than the brain. And if we are embodied minds, what goes for our minds goes for our selves. I said in Section 10.1 that a system to which we attribute portability is not strictly sufficient for that property. Our brains don’t exactly “have”—all on their own—the property of being portable.

Nor do our biological bodies “have”—all on their own—the capacity to move. Technologies are quite clearly among those conditions instantiating the portability we attribute to the putatively core cognitive system, agent, or organism. We can’t go just anywhere. Where we can go is determined by flesh and blood, to be sure, but also by “the environment.” But it’s not the case that a human organism’s capacity to move around in a context, inclusive as it is of that context, nonetheless enables us to clearly distinguish the two, the way we might distinguish a goldfish and its bowl of water—the medium and the extent of its swimming capacity. Or if it does, in the human case, the distinction does not carve out a naked ape-like organism. So not only does the organism not have, all on its own, the capacity to move, technologies clearly travel along *within* a core to which that disposition might ordinarily be attributed. What moves is technological. But it does not even have a consistent skin-like boundary. Human movement does not carve out, say, clothed apes. In the human case, it is not a fixed body, clothed or otherwise, that moves against a fixed backdrop that partly constitutes its mobility. Rather, “agent” and “environment” are dynamic and interwoven. What moves drags along what we might have called bits of the environment with it. Backpackers, for example, seem like rather portable humans, nonetheless dragging along with them clothes, contact lenses, cameras, guidebooks, passports and credit cards. Conversely, travellers make well-travelled cities what they are as much as cities ground the possibility of travel. Human mobility within an elaborate network of people and technological structures is perhaps more analogous to a fetus’s capacity to move within the womb. The fetus is attached to the womb by an

umbilical cord that shifts with it, and as it moves, the fetus changes the shape of the amniotic sack and the womb. And as the fetus goes where its mother goes, where the Earth goes, we go; its future is ours. In the case of space travel, what we call the body of the astronaut is tethered by a technological umbilical cord to the mothership, itself tethered by radio waves to the Earth. At the scale of the commute or the trip to the grocery store, the situation is the same. My sitting at a desk and applying my mind to certain questions—like the question of what physical conditions are presently attempting to answer that question—depends, as a temporally extended process, on there being a desk in a heated room and brain-fuelling food in the cold fridge. Our stomachs reach out through our forks, and our distribution and agronomic systems, to the soils and seas our lives grow out of. These causal interactions are constitutive of our mental lives as temporally extended things.

And there is good phenomenological evidence for the dynamicity of any putative core as far as it might be phenomenally-given. Merleau-Ponty (1996) explored this point in some depth, saying,

A [person] may, without any calculation, keep a safe distance between the feather in her hat and things which might break it off. She feels where the feather is just as we feel where our hand is. If I am in the habit of driving a car, I enter a narrow opening and see that I can “get through” without comparing the width of the opening with that of the wings, just as I go through a doorway without checking the width of the doorway against that of my body. The hat and the car have ceased to be objects with a size and volume which is established by comparison with other objects. . . . The blind [person’s] stick has ceased to be an object for him, and is no longer perceived for itself; its point has become an area of sensitivity, extending the scope and active radius of touch, and providing a parallel to sight. In the exploration of things, the length of the stick does not enter expressly as a middle term. . . . The position of things is immediately given through the extent of the reach which carries him into it, which comprises besides the arm’s own reach the stick’s range of action. . . . To get used to a hat, a car or a stick is to be transplanted into them, or conversely, to incorporate them into the bulk of our own body. Habit expresses our power of dilating our being in the world, or changing our existence by appropriating fresh instruments. . . . But this power of habit is no

different from the general one which we exercise over our body: if I am ordered to touch my ear or my knee, I move my hand to my ear or my knee by the shortest route, without having to think of the initial position of my hand, or that of my ear, or the path between them. (p. 143)

Whatever other conclusions Merleau-Ponty derived from these observations together with his other views, I think they show that our minds are dynamically embodied. Suppose the phenomenal character of an experience² can differentiate—among all of what that experience depends upon for its duration—the conscious *contents* of that experience from its unconscious *vehicles*. I say “conscious,” and “unconscious,” to make the distinction explicitly mutually-exclusive, whereas the vehicle/content distinction doesn’t clearly do this. On an account in which the museum is partly constitutive of Otto’s belief about it, the museum may be thought of as being among the vehicles of the belief if vehicles are taken as a special case of constituents and defined as anything on which a meaningful thing depends for its duration. So I will always say “unconscious” vehicles when excluding the conscious contents on which an experience nonetheless also depends for its duration such that it has the property of being about anything at all. The unconscious vehicles of an experience are distinguishable in the sense that the character of experience counterfactually depends on them, and yet we aren’t aware of them while having just *that* experience. Now, if we want to use the character of experience to mark this distinction, then technologies are frequently among the unconscious vehicles of our mental lives.

This should come as no great surprise if we accept that our biological tissues are often among the unconscious vehicles of our mental lives, because, as Merleau-Ponty suggests, there is no difference between our powers of habit with respect to sticks as compared to knees. But I will say a bit more in defense

² I am using “experience” in a very general sense to signal that which has the property of being conscious. This is something of a departure from its use in traditional epistemological debates between empiricism and rationalism about whether knowledge is attained primarily or only via “experience” or “reason,” as mutually exclusive faculties of mind. I would rather use “perception” for the empiricist’s source of knowledge, so that thinkers can be also be said to experience dreams and “offline” reasoning.

of that claim. Merleau-Ponty seems a little indecisive in the above passage on what our “bodies” consist in, at first describing instruments as incorporated “into the bulk of our own body,” and later speaking of our bodies in the ordinary sense. But if we want to use the character of experience to mark, or help to mark, the distinction between our “bodies” as unconscious vehicles and the “world” we experience, then, since technologies are frequently among the unconscious vehicles of our mental lives, it follows that our bodies, as far as they might be phenomenally-given, actually comprise habituated technologies, and to a surprising extent.

Why think there is no difference between our powers of habit with respect to technological as opposed to biological tissues? First, there is no obvious phenomenally-appreciable difference between what it is like to learn to use hands, as opposed to learning to use the things in our hands. Consider ten-finger typing. Once proficient, ten-finger-typists don’t have to work backward from the intention to type a string of words to the first word in the string, to the first letter that composes it, and then to the location of that letter on the keyboard. And since developing motor skills particular to keyboard use, typists don’t have to work backward from the intention to press a key to checking where their fingers are in space as they inch a finger toward a target while monitoring it via multiple sense modalities. And since developing a basic body schema, most folks don’t have to work backward from seeing a finger to making sure it belongs to them. All that goes on neurologically when a ten-finger-typist records their thoughts are transparent to the conscious volition for which the goal, in this case, the expression, is its object. And much of what goes on physiologically—finger flexion and extension—is translucent to this volition. I say “translucent,” because although unnoticed *ex hypothesi* in the experience in question, it is not unnoticeable—it is available to be noticed as required. But, and this is the point, a properly integrated keyboard is as transparent or translucent to the experience as finger flexions are.

Pain does not supply a reliable phenomenal distinction between what it is like to have a genetically-given body and what it is like to have a technically-extended one. Not having nerve endings in a bat is like having a leprosy foot wherein bacteria have caused peripheral sensory nerve damage, but not

yet motor nerve damage, and having such a foot just shows a changeable interface for different functions (walking versus tissue damage detection), the way the spaces in a revolving door are at times continuous with the interior of a building and at other times continuous with the exterior. For the function of tissue damage detection, this leprous foot might as well belong to anyone, and the damage is anyone's to see, but for the purpose of walking, the foot extends the agent as well as it can for as long as it can. Poor performance is relative to a Sorites scale in the center of which we can call it a toss up whether something belongs within a creature's agency or not. In that case, a prosthetic leg or a pacemaker might make for a more whole or lively life than being stuck with a leprous foot or cardiac arrhythmia. Even if, again, some difference between how brains interact with their biotic body and how they interact with habituated technological extensions turned out to involve different brain matter, neither brain-based explanation is likely to appeal to the termini of nerves in a way that could reliably specify a strictly biotic boundary between agency and world. Nerves don't extend to the very surface of our skin, but the epidermal layer that covers them is not typically excluded from our sense of our body. Hair and whiskers are non-metabolizing biomaterial structures produced by animals to (in part) extend touch sensitivity. Whiskers still belong to cats, not cat habitats. Shells belong to snails, barnacles, and turtles all the way. Photoreceptive neurons in our retinas are far from the bleeding edge of vision, light having first been modulated by a variety of non-nervous structures, which we consider to be part of the organism. Whether the discarded mollusk's shell belongs within the bounds of the agency of the hermit crab that takes it up and inhabits it, in the absence of access to phenomenal experience of the crab, would be hinted at by empirically-discoverable functions. Can the crab predict or learn to predict what spaces it can fit through? If it can, and humans certainly can, what further motivation do we have to say that a hermit crab or a spectacle-wearing organism walks around, not with an extended body, but with a piece of the environment perched on it? It would be a benevolent environment that nicely refracts the rest of itself into my eyes, furnishing a custom acuity for the particular focal range of my cornea and lens array. Rather, the glasses are a hard-won part of this seeing organism. Of course, we might draw a line between

the biotic and the technic in terms of modes of production and inheritance. But modes of production and inheritance can't be what determines what is part of my body and what is not, since there's a very real sense in which 'I' didn't make my own eyes. The collection of organs that became part of 'me' included eyes before it had much in the way of agency. If being made by the agent is a criterion for parthood within agency, then spectacles are more a part of the spectacle-wearing optician than their own eyes, which might better be described as their mother's. Rather, the tissue-likeness of spectacles and other prosthetics consists in their part in the organism's seeing function.

Ten-finger typists experience themselves as willing sentences onto the screen. We don't usually say they fire neurons and twitch muscle fibers; to say so is not only "pointlessly complex" (Clark and Chalmers, 1998, p. 13), it's mistaken or at least involves mismatched grain-levels of description. Whole agents don't fire neurons at all; other neurons do. Similarly, when we say the cricket or baseball batter hits the ball, we can and should omit "with the bat." The implicitness goes to the point that there is no batter without a bat. Or, we might have specified any number of other particulars that usually go without saying, if we felt the need to emphasize them, such as, "having judged the ball's speed," or "with oxygenated blood carrying adenosine triphosphate circulated to striated muscles." Again, just like the swing, these processes are subsumed under the whole process as parts of it; they are not something the batter or the organism does, they *are* the batting that individuates the organism as a batter. Although we could describe the same event as, "The batter swung the bat and the bat hit the ball," we would only do so to suspend our attention in mid-air, so to speak, to emphasize what hangs momentarily in the balance between the whole batter and the ball. We wouldn't mean that agency is transferred to the bat as if hitting the ball becomes a goal for the bat upon being swung. Nor do batters experience their agency as so transferred. A properly integrated keyboard or bat is as translucent to this willing, this intention, as are a typist's or batter's fingertips, tendons, forearm muscles, and elbows (although not invisible like their motor cortex). Practised batters can keep their eyes on the ball and practised typists can keep their eyes on the screen. This thesis as it now stands nearing completion comprises nearly a million characters. I

estimate that, once note-taking, editing-while-writing, multiple re-edits, and deletions of large portions of draft material are taken into consideration, I have made *at least* five, and perhaps as many as ten times that many keystrokes. Just imagine—as far as doing so is even possible—if “I” had had to think about each of those millions of keystrokes. But I didn’t, and there is no inner thing that did which can itself properly be called “thinking” or “I.” Instead, I think about whole propositions, and write them down, and the I that does these things has fingers and a computer among its parts.

The phenomenally-appreciable shifting locus of our attention and intentions is a measure of the degree of integration of technologies. The less we notice our extensions, the better they work for us; the more under our control, the less noticeable technologies are. Over-attending to the steps of an already-internalized procedure can famously lead to “choking under pressure” (see Beilock and Carr, 2001). Subsequently, integration is a measure of the extent of our agency versus sources of activity that are not our own. The ownership is constantly contested in a tug of war that happens across a variety of scales in time and space, private and public. We can lose touch with our extensions. The momentary disorientation when a keyboard key sticks unexpectedly, or the surprise when a bat splinters, or a “no internet connection” warning appears, is like the surprise when our foot slips, or our knees buckle unexpectedly. Our bodies betray us, and we disown them until they cooperate again. Notice the differing attribution of agency: normally, *I* do the running, except when my knee acts up. Normally, *I* do the surfing, until the network acts up. We might never notice our bodies if they didn’t get sore, hungry or tired because we wouldn’t need to, and we might never notice our familiar tools if they didn’t break (if sometimes only relative to the emergence of “better” tools). Interestingly, when we under-attend to procedural execution, we can become too rigid, unable to easily adapt to changes, and unable to later recall elements of a performance in order to improve (see Langer and Imber, 1979). In order to maintain flexibility within a realm, a well-entrenched bad habit in the form of one’s swing might have to be worked over if one wanted to achieve a better swing. In this working over, allocating our limited attention and intention is a balancing act. We can lose touch with, or might need to loosen our

grip on our extensions, the way we temporarily do when changing keyboard layouts, the speed of a mouse cursor, the thickness of automotive brake pads, the height of shoe heels, or the stiffness of paintbrush bristles. We get brake work done and then have to recalibrate our footwork, using vestibular, visual, and proprioceptive feedback. Whereas our intentions normally go straight to where the rubber meets the road, they temporarily retreat to attend to the recalibration where the pads meet the disk (this might feel like our foot on the pedal). Likewise, having an injury, undergoing a growth spurt, becoming fitter or fatter, all require that we modify our schemas and procedures. Judging whether you can jump a gap when your fitness is downgraded is like judging whether you can shoot a gap through the roundabout in your new mini van. But don't inspect muscle fibers or cylinders to help you make the call. Just practise with forgiving gaps until your judgement is invisibly calibrated as your extensions become translucent.

Any phenomenally-given body we have expands and contracts. Whatever it is about our brains that makes what psychologists call procedural memory possible, neuroscience will have a story about. But introspection can't tell us whether there are one or two stories to be told—one about strictly biotic body schemas as related to strictly bodily procedural learning and corresponding skill execution (say, clapping and walking) and another story about technic schemas as related to technical procedural learning and corresponding skill execution (say, batting a ball and riding a bike). Whether there is one flexible body schema, or two (or more), or one kind of procedural memory or two (or more), the point is that the interface of our agency with the world is flexible, both ways.

Rather, the phenomenological continuity, shall we say, between biological body use and habituated technology use would be explained by their having the same neurological basis. And in fact there probably isn't, say, a technological cortex adjacent to the visual and motor cortices. For one thing, there is no clear developmental case for a technological cortex. For example, if the human body schema was hard-wired in utero, the integration of keyboards and the like would be all the more likely to involve some further neural resources. But it is a matter of debate whether the human body schema is hard-wired. The main

reason to think it might not be is that, as Price (2006) summarizes, “Reports of congenital phantoms have led a number of experts to infer that neural representations of the body are hard-wired and underwritten by a genetically encoded blueprint” (p. 312). But Price questions whether phantom pain in reportedly mentally “complete,” but actually missing or malformed limbs of people with congenital amputations really demonstrates that some of this mapping is baked in. He suggests how an alternative model could account for even these strange cases:

It is proposed that body image develops in two stages: first, in utero, a primitive bilateral body image is constructed, driven by spontaneous limb movements. Second, after birth, visual and sensory information can lead to the development of phantoms by means of mirror neurons and the incorporation of prosthesis into body image. This hypothesis is capable of explaining all cases of aplasic phantoms without recourse to an independent genetically determined neural representation of self. (p. 320)

I think Price is interpreting the available research very well. But even if embryogenesis provided us with some modest genetically-determined neural representation of our biological bodies, that would fall short of demonstrating distinct neural bases for biotic-body-learning and technic-body-learning, because even a soft wired schema would have to be changeable. Evidence popularly summarized by Doige (2007) suggests that “spike-timing-dependent synaptic plasticity” (the tendency of connection strength between neurons to increase when their action potentials spike closely in time, a principle summarized in the slogan “neurons that fire together, wire together”), is a mechanism for associative learning and even cortical remapping, where neural resources are reorganized to serve new functions when input conditions or output requirements change. So, even if grasping nothing and grasping things were in some way different neurological processes, both involve neuroplasticity. Doige also summarizes related research that seems to bear out phenomenal experience wherein attention to the instrument being integrated is greatest during learning. As familiarity with a task grows, connections are strengthened, and yet later simplified to an efficient form. Attention during learning, and later trimming of neurological resources to the bare essentials suggests that proceduralization

during a learning phase is scaffolded by resources that are later reappropriated and put to other use, as if the familiar does not command attention so that attention can oversee other familiarization.

Further, soft wired body schemas would have to be able to be built upon procedurally, because, although babies' brains might, for example, have plenty of soft wiring genetically dedicated to their hands, they don't grow fingers in utero that are any more under their command to touch thumb to forefinger than able to pinch a crayon. So what is generally known as procedural learning has to be able to build on any putatively given body schema. If preliminary wriggling and twitching served to check and calibrate the schema, still, generally, learning to use hands in so many cases seems to be the same as learning to use the things hands use. People don't acquire the necessary motor control for writing, typing, batting, or riding a bike before trying their hands at these activities. So not even soft wired body schemas would show that procedural learning (of routines which are by definition not hard-wired) involves distinct resources in the case of biotic and technically-extended procedures.

Finally, extreme altriciality (protracted development requiring protracted postnatal investment from conspecifics) in the human case is consistent with the "same-neurology" hypothesis for biotic and technic integration I'm advocating. If biotic body-use and technology-use were separate neural bases for our abilities, we would have some more explaining to do, because they could have been separately optimized by selective processes in the course of human evolution. In the absence of any advantages as far as merely biotic bodily skills that might be conferred by altriciality—protracted in humans far beyond that of our great ape cousins—selection would plausibly have pressured human neonates to keep up with their cousins' earlier perceptual and motor development if it could do so without precluding their ability to learn new technical tricks later on. Rather, it seems like altriciality is a package-deal, such that the flexibility of the resources that underpin both biotic body-use and technology-use is a setback for walking and a leap forward toward one of the traits walking on the moon requires. The ability to learn is a short-term metabolic cost to the group, and its greatest asset over the longer term. As the schemas and procedures of our ancestors were decreasingly fixed by their genetics, they became increasingly

flexible and capable of being culturally modified. Flexible brains like ours require protracted sociocultural sheltering, but for this tradeoff of longer boot up and sociocultural induction, our minds are poised to exploit novel spaces and subsequently extend the sheltering effect that buys us more time to learn.

So our bodies as they would be phenomenally-given are flexible and often technically extended. That makes sense, given the probably shared neurological basis of the flexibility and a plausible cognitive evolutionary story as to why our biology is technological: our biological tissues are adapted to technology use because technologies are adaptive. At best, in defense of our traditionally rigid view, our biotic bodies are slightly less changeable than our technically-extended bodies in a few contingent ways. We are embodied minds, and that embodiment is dynamic.

11.3 Shared Tissues and Humane Futures

Now, the analogy between prosthetics and tissue in the case of what look like personal devices, however counter-intuitive, might seem even less clear in the case of shared technologies, like corporate buildings and public roads. But actually, what I kept back for rhetorical flourish here is that not only does the view entail the possibility of human organisms' having shared tissues, but it allows that these may be among the most habituated of all; they are the ones upon which the character of our experiences and thus our particular mental lives and selves depend unconsciously most of the time. In fact, the analogy with tissues here is *deeper*, in the sense that our dependence on them, not just for our particular lives, but for anything like our lives at all, is deeper. We individually own certain living investments, our hearts, legs, best shoes, and the tools of our crafts. Yet so many of our personal devices—cheap and broken down, disused or under-used belongings that clutter our lives—are what can be thought of as metabolic wastes. (In contrast to the metabolic wastes of our biotic tissues, of course, much technological waste is toxic to other organisms.) In continuity with the other wastes we produce, and as opposed to the tissues we grow, so many of our personal belongings are things that either do not qualify as tissue defined as homeostatic metabolic capital (living

investments), or are only “tissues” insofar as vegetative or cancerous tissues are. Many shared tissues, on the other hand, are clearly living investments. There aren’t as many shared structures as there are personal tissues that we could dispense with and still lead anything close to the particular lives we lead, or have the experiences we have. My community’s food systems, institutions, transportation and communication networks are part of me because they make up the livelihood of the “we” in which both my agential and mental being have life and breath. To have a human life and mind is to be woven into a radically joint venture which is fundamentally technological, so that, to the cognitive process resulting in the thought that “I’ll have to leave for my meeting right after breakfast,” the presence of the milk in my fridge, the car in my driveway, roads and petrol stations, and a chair in a room with lights in the university building, are not contents, but unconscious vehicles. And few of these are “mine.” In fact, many of us have house and car payments to make. The ownership—in the sense of ultimate power or control—is shared.

But that is not to say these shared tissues are the best tissues. I thus want to close by noting a particular consequence of this broadened way of looking at ourselves. As I said in Section 10.4, despite the surprisingly wide extent of the total realizers of our mental lives, there are interesting subsets of the total realizers of our mental lives. For example, I think it is very interesting that technologies, especially the ancient and mundane, can be among those conditions of our mental lives that we don’t notice at one time, while designing them at another. As we saw just above, our phenomenally-given bodies are dynamic such that technologies often fall under the subset of the total realizers of our mental lives that we might call their unconscious vehicles, and yet the same parts can at other times fall under another interesting subset of the total realizers of our mental lives. They fall under those conditions within our influence. Add to this the widespread idea that moral responsibility presupposes agency and that the extent of our agency is the extent of our moral responsibility. These extents plausibly track the epistemic horizons of our influence—those outcomes we can be said to have designed, or ought to have designed in the sense that we might have known what they would be. For example, we use the concept of ecological footprint to evaluate the impact of our ways of life on the

planet. These footprints do not belong to a merely metaphorical “foot” that is incidental to us and beyond our control. Rather, since technology is more like tissue than like parts of the environment, our ecological footprints reflect our very size as organisms. And we *can* know ourselves and see ourselves in this way, because, as I have argued, that is the way we are. We can see ourselves as technologically extended, for example by recognizing the sloppiness of our ordinary boundary drawing.

Now, the particular facts underlying the generalization that we are partly what we make are that you may be partly what I make, and that I may be partly what you make. Things of our making will certainly be among the realizers of our offspring’s mental lives. This means we are shaping each other’s minds in ways that we are often unaware of. We are often unaware in two senses.

First, we so quickly habituate technologies—and many of these from a very early age—that we not only become unconscious of them, but we forget life before them. While we perceive the world through technologies—they extend our discriminatory apparati—they constrain our imaginative possibilities. When someone says, “I can’t live without my smartphone,” they reveal a mental state whose realizability includes a technological apparatus. We are different creatures with different minds when we change our technological bodies. Indeed, *that* person could not live without a smartphone, grocery stores, petrol stations, and a great many other tissues or tissue-like apparati. Our particular mental lives would fail to transplant to bodies without these appendages, in which our minds would be made anew. Unsurprisingly, Westerners born since the 1990’s find it difficult to imagine life without the internet. I only faintly remember. And we can all imagine what it is like to be an illiterate medieval peasant about as well as we can imagine what it is like to be a bat. That seems irrelevant in the sense that it is not desirable to recreate medieval life. But the general point is relevant insofar as, reflecting on our own experiences, we may want to, as far as it is possible, preserve what is worthwhile even as we seek to reduce conditions of suffering. At least, we ought not to conflate novelty with progress. In conflating these we set ourselves up to forget, even if by nostalgically misconstruing, the goods we already had.

Second, as designers of technology, we may be unaware that we are building

each other's minds. And *this* is why it matters to get the causal/constitutive distinction right for the temporal frame of some given explanatory target. When we make things, we are plausibly not responsible for *all* of the effects of doing so a million years hence. But we are plausibly partly responsible for the states of affairs of which our creations are *parts* soon afterward. The results of my productive activity are fairly proximate effects of that activity. But the mental states of which those same results become a part are not further causal "effects." They are as proximate in time as the results of my productive activity. And our mental lives are plausibly more predictable if we could grasp this point about the breadth of the conditions upon which our minds depend. As such, human experiences are among the things we design or ought to have designed in the sense that we might have known what they would be, not as distant effects but as immediately constituted by the results of our productive activities.

The upshot is first of all that if we can learn to see ourselves in this broadened way, we might take our technical carapaces more personally. Second, it may be that our mental health declines as our technological extensions grow, as popularly claimed in media and books like *The Shallows* (Carr, 2010), and increasingly corroborated by research (e.g. Takeuchi et al., 2018; Shakya and Christakis, 2017) in the relatively new but suddenly ubiquitous arenas of gaming and social media. More generally, our embodied mental health—our cognitive abilities, imaginations, happiness, and our anxieties—is partly constituted by the functioning of our tissues, including technology. Since our minds are dear to us, being partly what we make is a reason to take our jobs as designers of our selves more seriously. And since our similar genetic cores give rise to flexible phenotypes that are non-viable as organisms until extended technologically, the rest of us is up to us to make. Human organisms are polymorphic, and the question we get to ask, and ought to ask is, *what kinds of creatures do we want to be?* Do we want to be *Homo technologicus automobilis sapiens*, sitting in traffic on the way to work to pay for the cars that all too often tear themselves to shreds, with our brains on board? We might want to be healthier. In light of Western lifestyle diseases—autoimmune disorders, obesity and diabetes, heart disease, some cancers, anxiety and depression—we know our health is partly

constituted by the modes in which our stomachs reach out through our forks, and our distribution and agronomic systems, to the soils and seas our lives grow out of. And we might want to be smaller. If we notice that the technological tissues with which we make our lives are grown and require maintenance, that they occupy precious biocapacity, we might ask whether the quantitatively burgeoning Western life is qualitatively richer, in more respects than heart disease and anxiety, than the multitude of lives one Western life supplants or limits. The beauty of the fact that we can share huge parts of ourselves and our minds means that we can share resources. We can cultivate more life—more mental life—out of less stuff, if our personal unshared tissues are minimized. It might be that less stuff of certain kinds could be more flourishing all around.

Of course, if extended self-awareness was to lead some of us to pursue simpler lives, that wouldn't be enough. This self-awareness needs a systemic, political inflection, because of the systemic nature of present ecological destruction and because the technological appendages that make our lives are often corporate. Present ecological ailments are not primarily epistemic failures of an otherwise rational corporate coordination of human activity. While ecological damage is often foreseen, consequences merely accrue, first because of our misshapen senses of responsibility (rooted in sufficiency errors), but also, and I think relatedly, because there is no robust coordination between epistemic elements of society that take a wider and longer view of nutrient cycles, and executive elements of society that could correct the market forces shaping the way we make our lives. We already know these forces undervalue "natural capital" and externalize the costs of undermining and overburdening "ecosystem services." Like the dissonance between our glorification of freedom and avoidance of responsibility, the relationship between economics and ecology is oddly troubled. *Eco*-nomics is housekeeping. *Eco*-logy is the study of homes. Shouldn't we keep house according to our study of home? Yet too often, we are using an incomplete construal of our niche to think about markets. Our shared lives, considered on ecological terms, have their full-fledged niche within the world only in combination with one another *and* with technology. We need an economics, say mega-economics, whose calculus aligns with the full extent of our corporate tissues and shared total way of life—our full-fledged niche. We ought to redefine

wealth as healthy tissue, and reconsider its source in the ultimate cooperation of diverse and modest organisms. As long as we are trying to build healthy societies on a flawed model of what it is that we share as humans, my view is that we'll struggle to build institutions capable of coordinating our complete extended corporate existence.

Wrapping Up

I've argued in Section 11.1 that there is a close relation between our minds and our selves as bearers of praise and blameworthiness. Most plausibly, we *are* minds, which are extensive physical conditions that include our bodies and technological carapaces. In every case of psychological continuity there is also circumstantial continuity constituting a person's being of a kind with their past and future selves. I argued in Section 11.2 that phenomenological evidence of our personal boundaries suggest that our bodies are dynamically extensible. The upshot is that our particular mental lives would fail to be duplicated if our biological duplicates were to be transplanted into different technical carapaces and wider technological circumstances. Thus, in Section 11.3, I suggested that it is plausible and worthwhile to distinguish another morally salient subset of the complete constituents or total realizers of our minds and selves from the rest. Technologies are among those special constituents of our mental lives which we design, which has moral implications for our sizes as extended—and overlapped—beings.

To sum up, everything is physical. When we take apart big or complicated things—whether events, processes, systems, objects or states—we find smaller or simpler things. But unlike some physicalists, I don't think it makes much sense to deny that complicated things *are* sets of simpler things, their relations, and interactions. The latter are all physical and located in space and time, and so are complicated things, like us. But the natures of these complicated things, or what it is that we want to explain about them, show that, very often, those things are either not quite *what* we think they are, or they are not wholly *where* we think they are. Sometimes things aren't what we thought. There is no lumniferous aether or phlogiston. But often, complicated things

elude identification with their sets of simpler things, relations and interactions, not because they don't belong in predictively successful models of reality, but because they aren't wholly *where* we tend to look. Our minds are like that, and so are we. These may seem mysterious because minds are not *just* brains. They are actually brains embodied and embedded in a world. And our particular minds are brains technologically embodied and embedded in a technological world. That means we are partly what we make. And the fact that this comes as a surprise means our self-engineering could be more self-aware, and should be more self-reflective.

References

- Adams, F., and Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, 14(1), 43–64. <https://doi.org/10.1080/09515080120033571>
- Adams, F., and Aizawa, K. (2008). *The Bounds of Cognition*. John Wiley & Sons.
- Adams, F., and Aizawa, K. (2010). Defending the bounds of cognition. In R. Menary (Ed.), *The Extended Mind* (pp. 67–80). MIT Press.
- Aizawa, K., and Adams, F. (2005). Defending Non-Derived Content. *Philosophical Psychology*, 18(6), 661–669. <https://doi.org/10.1080/09515080500355186>
- Aizawa, K. (2009). Neuroscience and Multiple Realization: A Reply to Bechtel and Mundale. *Synthese*, 167(3), 493–510.
- Aizawa, K., and Gillett, C. (2009). The (Multiple) Realization of Psychological and other Properties in the Sciences. *Mind & Language*, 24(2), 181–208. <https://doi.org/10.1111/j.1468-0017.2008.01359.x>
- Alexander, S. (1927). *Space, Time, and Deity*. MacMillan and Co., Limited.
- Allen, B. (2004). *Knowledge and Civilization*. Boulder: Westview Press.
- Armstrong, D. M. (1989). *Universals: An Opinionated Introduction* (1 edition). Boulder: Westview Press.
- Baker, L. R. (1997). Why Constitution is Not Identity. *The Journal of Philosophy*, 94(12), 599–621.
- Baker, L. R. (1999). Unity without Identity: A New Look at Material

- Constitution. *Midwest Studies In Philosophy*, 23(1), 144–165. <https://doi.org/10.1111/1475-4975.00008>
- Baker, L. R. (2000). *Persons and Bodies: A Constitution View*. Cambridge University Press.
- Baumgartner, M., and Casini, L. (2016). An Abductive Theory of Constitution. *Philosophy of Science*, 84(2), 214–233. <https://doi.org/10.1086/690716>
- Bechtel, W. (2001). The Compatibility of Complex Systems and Reduction: A Case Analysis of Memory Research. *Minds and Machines*, 11(4), 483–502. <https://doi.org/10.1023/A:1011803931581>
- Beilock, S. L., and Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General*, 130(4), 701–725. <https://doi.org/10.1037/0096-3445.130.4.701>
- Bennett, K. (2003). Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It. *Noûs*, 37(3), 471–497. <https://doi.org/10.1111/1468-0068.00447>
- Bennett, K. (2004). Spatio-Temporal Coincidence and the Grounding Problem. *Philosophical Studies*, 118(3), 339–371.
- Bennett, K. (2008). Exclusion again. In J. Hohwy and J. Kallestrup (Eds.), *Being Reduced: New Essays on Reduction, Explanation, and Causation*. Oxford University Press.
- Bennett, K. (2011). Construction area (no hard hat required). *Philosophical Studies*, 154(1), 79–104.
- Bennett, K. (2017). *Making Things Up*. Oxford University Press.
- Bickle, J. (2016). Multiple Realizability. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016). Metaphysics Research Lab, Stanford University.
- Bigelow, J. (1996). Presentism and Properties. *Philosophical Perspectives*, 10, 35–52. <https://doi.org/10.2307/2216235>
- Biro, J. (2017). Constitution and Identity. *Erkenntnis*, 83(6), 1127–1138.

<https://doi.org/10.1007/s10670-017-9932-0>

- Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O., and Piddock, L. J. V. (2014). Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*, 13(1), 42–51. <https://doi.org/10.1038/nrmicro3380>
- Block, N. (1997). Anti-Reductionism Slaps Back. *Noûs*, 31, 107–132. <https://doi.org/10.1111/0029-4624.31.s11.5>
- Block, N., and Fodor, J. A. (1972). What Psychological States are Not. *The Philosophical Review*, 81(2), 159–181. <https://doi.org/10.2307/2183991>
- Boyd, R. (1999). Kinds, complexity and multiple realization. *Philosophical Studies*, 95(1–2), 67–98.
- Burge, T. (1986). Individualism and psychology. *The Philosophical Review*, 95(1), 3–45.
- Burke, M. B. (1992). Copper Statues and Pieces of Copper: A Challenge to the Standard Account. *Analysis*, 52(1), 12–17. <https://doi.org/10.2307/3328875>
- Bursten, J. R. (2014). Microstructure without Essentialism: A New Perspective on Chemical Classification. *Philosophy of Science*, 81(4), 633–653.
- Carnap, R. (1967). *The Logical Structure of the World*. University of California Press.
- Carr, N. G. (2010). *The Shallows: What the Internet is Doing to Our Brains*. W.W. Norton.
- Chisholm, R. M. (1973). Parts as Essential to Their Wholes. *The Review of Metaphysics*, 581–603.
- Clark, A. (2005). Intrinsic Content, Active Memory and the Extended Mind. *Analysis*, 65(1), 1–11.
- Clark, A. (2010a). Coupling, constitution and the cognitive kind: A reply to Adams and Aizawa. In R. Menary (Ed.), *The Extended Mind* (pp. 81–99). MIT Press.
- Clark, A. (2010b). Memento’s revenge: The extended mind extended. In R.

- Menary (Ed.), *The Extended Mind* (pp. 43–66). MIT Press.
- Clark, A. (2010c). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press.
- Clark, A., and Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7–19.
- Cotnoir, A. J., and Baxter, D. L. M. (2014). *Composition as Identity*. Oxford University Press, USA.
- Craver, C. F. (2007). Constitutive Explanatory Relevance. *Journal of Philosophical Research*, 32, 3–20. <https://doi.org/10.5840/jpr20073241>
- Craver, C. F. (2014). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Clarendon Press.
- Decock, L., and Douven, I. (2011). Similarity After Goodman. *Review of Philosophy and Psychology*, 2(1), 61–75. <https://doi.org/10.1007/s13164-010-0035-y>
- Descartes, R. (1996). Selections from Meditations on First Philosophy. In P. A. Morton (Ed.), *A Historical Introduction to the Philosophy of Mind: Readings with Commentary*. Broadview Press.
- Deutsch, H. (2017). Relative Identity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017). Metaphysics Research Lab, Stanford University.
- Doige, N. (2007). *The Brain That Changes Itself: Stories of Personal Triumph from the Frontiers of Brain Science*. Penguin Group US.
- Elder, C. (2003). Alexander’s Dictum and the Reality of Familiar Objects. *Topoi*, 22, 163–171. <https://doi.org/10.1023/A:1024930306624>
- Endicott, R. (2016a). Developing the explanatory dimensions of part–whole realization. *Philosophical Studies*, 173(12), 3347–3368. <https://doi.org/10.1007/s11098-016-0674-7>
- Endicott, R. (2016b). Functionalism, superduperfunctionalism, and physicalism: Lessons from supervenience. *Synthese*, 193(7), 2205–2235. <https://doi.org/>

10.1007/s11229-015-0839-5

- Fodor, J. A. (1968). *Psychological Explanation: An Introduction to the Philosophy of Psychology*. Random House.
- Fodor, J. A. (1974). Special Sciences (Or: The Disunity of Science as a Working Hypothesis). *Synthese*, 28(2), 97–115.
- Fodor, J. A. (1981). The Mind-Body Problem. *Scientific American*, 244, 114–25.
- Fodor, J. A. (1997). Special Sciences: Still Autonomous after All these Years. *Noûs*, 31, 149–163. <https://doi.org/10.1111/0029-4624.31.s11.7>
- Frigg, R., and Nguyen, J. (2016). Scientific Representation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University.
- Gallagher, S. (2013). The socially extended mind. *Cognitive Systems Research*, 25, 4–12.
- Gallois, A. (2016). Identity Over Time. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University.
- Gärdenfors, P. (2004). *Conceptual Spaces: The Geometry of Thought*. MIT Press.
- Gärdenfors, P. (2014). *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. MIT Press.
- Geach, P. T. (1980). *Reference and Generality: An Examination of Some Medieval and Modern Theories*. Cornell University Press.
- Gillett, C. (2002a). The dimensions of realization: A critique of the Standard view. *Analysis*, 62(276), 316–323. <https://doi.org/10.1111/1467-8284.00377>
- Gillett, C. (2002b). The metaphysics of realization, multiple realizability, and the special sciences. *The Journal of Philosophy*, 100(11), 591–603.
- Gillett, C. (2007). Understanding the New Reductionism: The Metaphysics of

- Science and Compositional Reduction. *The Journal of Philosophy*, 104(4), 193–216.
- Gillett, C. (2013). Constitution, and Multiple Constitution, in the Sciences: Using the Neuron to Construct a Starting Framework. *Minds and Machines*, 23(3), 309–337. <https://doi.org/10.1007/s11023-013-9311-9>
- Goodman, N. (1972). *Problems and Projects*. Indianapolis: Bobbs-Merrill.
- Goodman, N. (2012). *The Structure of Appearance*. Springer Science & Business Media.
- Grandy, R. E. (2016). Sortals. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University.
- Heller, M., Heller, P. M., and Mark, H. (1990). *The Ontology of Physical Objects: Four-Dimensional Hunks of Matter*. Cambridge University Press.
- Homer. (1919). *The Odyssey* (Vols. 1–2; A. T. Murray, Trans.). W. Heinemann.
- Horgan, T. (1993). From Supervenience to Superdupervenience: Meeting the Demands of a Material World. *Mind*, 102(408), 555–586.
- Hubble, P. (2009). *A Certainty of Death: Appreciating Human Animalhood* (Institute for Christian Studies). Retrieved from <http://hdl.handle.net/10756/288472>
- Hurley, S. L. (2010). Varieties of Externalism. In R. Menary (Ed.), *The Extended Mind* (pp. 101–153). MIT Press.
- Johnston, M. (1992). Constitution Is Not Identity. *Mind*, 101(401), 89–105.
- Kim, J. (1984). Concepts of Supervenience. *Philosophy and Phenomenological Research*, 45(2), 153–176.
- Kim, J. (1987). “Strong” and “Global” Supervenience Revisited. *Philosophy and Phenomenological Research*, 48(2), 315–326.
- Kim, J. (1992a). “Downward Causation” in Emergentism and Nonreductive Physicalism. In A. Beckermann and H. Flohr (Eds.), *Emergence or Reduction?: Essays on the Prospects of Nonreductive Physicalism*. Walter de

Gruyter.

- Kim, J. (1992b). Multiple Realization and the Metaphysics of Reduction. *Philosophy and Phenomenological Research*, 52(1), 1–26.
- Kim, J. (1997). Does the Problem of Mental Causation Generalize? *Proceedings of the Aristotelian Society*, 97, 281–297.
- Kim, J. (1999). Making sense of emergence. *Philosophical Studies*, 95(1), 3–36.
- Kim, J. (2000). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation* (Reprint edition). Cambridge, Mass.: A Bradford Book.
- Kim, J. (2003). Blocking Causal Drainage and Other Maintenance Chores with Mental Causation¹. *Philosophy and Phenomenological Research*, 67(1), 151–176. <https://doi.org/10.1111/j.1933-1592.2003.tb00030.x>
- Kim, J. (2012). The very idea of token physicalism. In S. Gozzano and C. S. Hill (Eds.), *New perspectives on type identity: The mental and the physical* (pp. 167–193). Cambridge University Press.
- Kingsbury, J. (2004). Biologising the Mind. *Biology and Philosophy*, 19(3), 473–482. <https://doi.org/10.1023/B:BIPH.0000036148.32654.90>
- Kingsbury, J. (2006). A Proper Understanding of Millikan. *Acta Analytica*, 21(3), 23–40. <https://doi.org/10.1007/s12136-006-1008-8>
- Kingsbury, J. (2008). Learning and selection. *Biology & Philosophy*, 23(4), 493–507. <https://doi.org/10.1007/s10539-008-9113-2>
- Kirchhoff, M. D. (2015). Extended Cognition & the Causal-Constitutive Fallacy: In Search for a Diachronic and Dynamical Conception of Constitution. *Philosophy and Phenomenological Research*, 90(2), 320–360. <https://doi.org/10.1111/phpr.12039>
- Koslicki, K. (2015). The coarse-grainedness of grounding. In *Oxford Studies in Metaphysics, Volume 9* (Vol. 9). Oxford University Press.
- Langer, E. J., and Imber, L. G. (1979). When practice makes imperfect: Debilitating effects of overlearning. *Journal of Personality and Social Psychology*,

- 37(11), 2014–2024. <https://doi.org/10.1037/0022-3514.37.11.2014>
- Leuridan, B. (2012). Three Problems for the Mutual Manipulability Account of Constitutive Relevance in Mechanisms. *The British Journal for the Philosophy of Science*, 63(2), 399–427. <https://doi.org/10.1093/bjps/axr036>
- Lewis, D. (1973). Causation. *The Journal of Philosophy*, 70(17), 556–567.
- Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, 61(4), 343–377.
- Lewis, D. (1991). *Parts of Classes*. B. Blackwell.
- Lewis, D. (2000). Causation as Influence. *The Journal of Philosophy*, 97(4), 182–197.
- Locke, J. (1841). *An Essay Concerning Human Understanding*.
- Mackenzie, C. (2014). Embodied agents, narrative selves. *Philosophical Explorations*, 17(2), 154–171. <https://doi.org/10.1080/13869795.2014.886363>
- Maclaurin, J., and Sterelny, K. (2008). *What Is Biodiversity?* University of Chicago Press.
- Menary, R. (2010). *The Extended Mind*. MIT Press.
- Merleau-Ponty, M. (1996). *Phenomenology of Perception*. Motilal Banarsidass Publishers.
- Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. MIT Press.
- Millikan, R. G. (2017). *Beyond Concepts: Unicepts, Language, and Natural Information*. Oxford University Press.
- Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation*. London: Routledge & Kegan Paul.
- Noë, A. (2010). *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness* (1 edition). Hill and Wang.
- Noonan, H., and Curtis, B. (2017). Identity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017). Metaphysics Research

Lab, Stanford University.

Noonan, H. W. (1993). Constitution is Identity. *Mind*, 102(405), 133–146.

O'Connor, T., and Wong, H. Y. (2005). The Metaphysics of Emergence. *Noûs*, 39(4), 658–678. <https://doi.org/10.1111/j.0029-4624.2005.00543.x>

Olson, E. T. (1999). *The Human Animal: Personal Identity without Psychology*. Oxford University Press.

Olson, E. T. (2007). *What Are We?: A Study in Personal Ontology*. Oxford University Press.

Parfit, D. (1986). *Reasons and Persons*. Oxford University Press.

Parfit, D. (2012). We Are Not Human Beings. *Philosophy*, 87(01), 5–28. <https://doi.org/10.1017/S0031819111000520>

Paul, L. A. (1998). Keeping Track of the Time: Emending the Counterfactual Analysis of Causation. *Analysis*, 58(3), 191–198. <https://doi.org/10.1111/1467-8284.00121>

Pereboom, D. (2002). Robust nonreductive materialism. *The Journal of Philosophy*, 99(10), 499–531.

Plato. (1914). *Plato in twelve volumes: With an English translation* (H. N. Fowler, W. R. M. Lamb, R. G. Bury, and P. Shorey, Trans.). W. Heinemann.

Polger, T. W. (2007). Realization and the metaphysics of mind. *Australasian Journal of Philosophy*, 85(2), 233–259. <https://doi.org/10.1080/00048400701343085>

Polger, T. W. (2009). Evaluating the evidence for multiple realization. *Synthese*, 167(3), 457–472.

Polger, T. W., and Shapiro, L. A. (2008). Understanding the Dimensions of Realization. *The Journal of Philosophy*, 105(4), 213–222.

Polger, T. W., and Shapiro, L. A. (2016). *The Multiple Realization Book*. Oxford University Press.

Price, E. H. (2006). A critical review of congenital phantom limb cases and

- a developmental theory for the basis of body image. *Consciousness and Cognition*, 15(2), 310–322. <https://doi.org/10.1016/j.concog.2005.07.003>
- Prior, E. W., Pargetter, R., and Jackson, F. (1982). Three Theses about Dispositions. *American Philosophical Quarterly*, 19(3), 251–257.
- Putnam, H. (1967). Psychological Predicates. In W. H. Capitan and D. D. Merrill (Eds.), *Art, Mind, and Religion*. University of Pittsburgh Press.
- Putnam, H. (1975a). Minds and Machines. In *Mind, Language and Reality: Vol. Philosophical Papers: Volume 2*. Cambridge University Press.
- Putnam, H. (1975b). The Meaning of “Meaning.” In K. Gunderson (Ed.), *Language, Mind, and Knowledge*. University of Minnesota Press.
- Quine, W. V. (1969). Natural Kinds. In N. Rescher (Ed.), *Essays in Honor of Carl G. Hempel: A Tribute on the Occasion of his Sixty-Fifth Birthday* (pp. 5–23). https://doi.org/10.1007/978-94-017-1466-2_2
- Ramsey, W. (2016). Eliminative Materialism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University.
- Ross, D., and Ladyman, J. (2010). The Alleged Coupling-Constitution Fallacy and the Mature Sciences. In R. Menary (Ed.), *The Extended Mind* (pp. 155–166). MIT Press.
- Rupert, R. D. (2010). Representation in Extended Cognitive Systems: Does the Scaffolding of Language Extend the Mind? In R. Menary (Ed.), *The Extended Mind* (pp. 325–354). MIT Press.
- Schaffer, J. (2010). Monism: The Priority of the Whole. *The Philosophical Review*, 119(1), 31–76. <https://doi.org/10.1215/00318108-2009-025>
- Schaffer, J. (2012). Grounding, transitivity, and contrastivity. In F. Correia and B. Schnieder (Eds.), *Metaphysical Grounding: Understanding the Structure of Reality* (pp. 122–138). Cambridge University Press.
- Schechtman, M. (2014). *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. Oxford University Press.

- Shakya, H. B., and Christakis, N. A. (2017). Association of Facebook Use With Compromised Well-Being: A Longitudinal Study. *American Journal of Epidemiology*, 185(3), 203–211. <https://doi.org/10.1093/aje/kww189>
- Shapiro, L. A. (2000). Multiple Realizations. *The Journal of Philosophy*, 97(12), 635–654. <https://doi.org/10.2307/2678460>
- Shapiro, L. A. (2008). How to Test for Multiple Realization. *Philosophy of Science*, 75(5), 514–525.
- Shapiro, L. A., and Polger, T. W. (2012). Identity, variability, and multiple realization in the special sciences. In S. Gozzano and C. S. Hill (Eds.), *New Perspectives on Type Identity: The Mental and the Physical* (pp. 264–88). Cambridge University Press.
- Shoemaker, D. (2016). Personal Identity and Ethics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University.
- Shoemaker, S. (2003). Realization, Micro-Realization, and Coincidence. *Philosophy and Phenomenological Research*, 67(1), 1–23.
- Smid, J. (2015). The ontological parsimony of mereology. *Philosophical Studies*, 172(12), 3253–3271. <https://doi.org/10.1007/s11098-015-0468-3>
- Sober, E. (1999). The Multiple Realizability Argument against Reductionism. *Philosophy of Science*, 66(4), 542–564.
- Sterelny, K. (2004). Externalism, epistemic artefacts and the extended mind. *The Externalist Challenge*, 239–254.
- Takeuchi, H., Taki, Y., Asano, K., Asano, M., Sassa, Y., Yokota, S., ... Kawashima, R. (2018). Impact of frequency of internet use on development of brain structures and verbal intelligence: Longitudinal analyses. *Human Brain Mapping*, 39(11), 4471–4479. <https://doi.org/10.1002/hbm.24286>
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327–352. <https://doi.org/10.1037/0033-295X.84.4.327>
- Unger, P. (1979). There Are No Ordinary Things. *Synthese*, 41(2), 117–154.

- van Inwagen, P. (1995). *Material Beings*. Cornell University Press.
- Varzi, A. (2016). Mereology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University.
- Wachowski, L. (Director), and Wachowski, L. (Director). (1999). *The Matrix*. Retrieved from <http://www.imdb.com/title/tt0133093/>
- Walter, S., and Kyselo, M. (2009). Review of *Review of The Bounds of Cognition*, by Fred Adams and K. Aizawa. *Erkenntnis* (1975-), 71(2), 277–281.
- Wasserman, R. (2004). The Constitution Question. *Noûs*, 38(4), 693–710. <https://doi.org/10.1111/j.0029-4624.2004.00489.x>
- Wasserman, R. (2017). Material Constitution. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017). Metaphysics Research Lab, Stanford University.
- Weatherson, B. (2016). The Problem of the Many. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University.
- Weir, P. (Director). (1998). *The Truman Show*. Retrieved from <http://www.imdb.com/title/tt0120382/>
- Weisberg, M. (2012). Getting Serious about Similarity. *Philosophy of Science*, 79(5), 785–794. <https://doi.org/10.1086/667845>
- Wilson, J. (2009). Determination, Realization and Mental Causation. *Philosophical Studies*, 145(1), 149–169.
- Wilson, J. (2011). Non-reductive Realization and the Powers-based Subset Strategy. *The Monist*, 94(1), 121–154.
- Wilson, J. (2014). No Work for a Theory of Grounding. *Inquiry*, 57(5–6), 535–579. <https://doi.org/10.1080/0020174X.2014.907542>
- Wilson, J. (2015). Metaphysical Emergence: Weak and Strong. *Metaphysics in Contemporary Physics*, 251–306.

- Wilson, R. A. (2001). Two Views of Realization. *Philosophical Studies*, 104(1), 1–31.
- Wilson, R. A. (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences*. Cambridge University Press.
- Wilson, R. A. (2007). A Puzzle About Material Constitution and How to Solve It: Enriching Constitution Views in Metaphysics. *Philosopher's Imprint*, 7(5).
- Wilson, R. A. (2009). The Transitivity of Material Constitution. *Noûs*, 43(2), 363–377.
- Wilson, R. A., and Craver, C. F. (2007). Realization: Metaphysical and Scientific Perspectives. In P. Thagard (Ed.), *Philosophy of Psychology and Cognitive Science* (pp. 81–104). Elsevier.
- Witmer, G. (2003). Multiple Realizability and Psychological Laws: Evaluating Kim's Challenge. In S. Walter and H.D. Heckmann (Eds.), *Physicalism and Mental Causation: The Metaphysics of Mind and Action* (pp. 59–84). Imprint Academic.
- Yi, B. (2018). Nominalism and Comparative Similarity. *Erkenntnis*, 83(4), 793–803. <https://doi.org/10.1007/s10670-017-9914-2>
- Ylikoski, P. (2013). Causal and Constitutive Explanation Compared. *Erkenntnis*, 78(2), 277–297. <https://doi.org/10.1007/s10670-013-9513-9>
- Zimmerman, D. W. (1995). Theories of Masses and Problems of Constitution. *The Philosophical Review*, 104(1), 53–110. <https://doi.org/10.2307/2186012>

- Adams, F., and Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, 14, 43–64.
- Adams, F., and Aizawa, K. (2008). *The Bounds of Cognition*. John Wiley & Sons.
- Adams, F., and Aizawa, K. (2010). Defending the bounds of cognition. In R. Menary (Ed.), *The Extended Mind* (pp. 67–80). MIT Press.
- Aizawa, K. (2009). Neuroscience and Multiple Realization: A Reply to Bechtel and Mundale. *Synthese*, 167, 493–510.
- Aizawa, K., and Adams, F. (2005). Defending Non-Derived Content. *Philosophical Psychology*, 18, 661–669.
- Aizawa, K., and Gillett, C. (2009). The (Multiple) Realization of Psychological and other Properties in the Sciences. *Mind & Language*, 24, 181–208.
- Alexander, S. (1927). *Space, Time, and Deity*. MacMillan and Co., Limited.
- Allen, B. (2004). *Knowledge and Civilization*. Boulder: Westview Press.
- Armstrong, D. M. (1989). *Universals: An Opinionated Introduction* (1 edition). Boulder: Westview Press.
- Baker, L. R. (1997). Why Constitution is Not Identity. *The Journal of Philosophy*, 94, 599–621.
- Baker, L. R. (1999). Unity without Identity: A New Look at Material Constitution. *Midwest Studies in Philosophy*, 23, 144–165.
- Baker, L. R. (2000). *Persons and Bodies: A Constitution View*. Cambridge University Press.
- Baumgartner, M., and Casini, L. (2016). An Abductive Theory of Constitution. *Philosophy of Science*, 84, 214–233.
- Bechtel, W. (2001). The Compatibility of Complex Systems and Reduction: A Case Analysis of Memory Research. *Minds and Machines*, 11, 483–502.
- Beilock, S. L., and Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology*:

- General*, 130, 701–725.
- Bennett, K. (2003). Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It. *Noûs*, 37, 471–497.
- Bennett, K. (2004). Spatio-Temporal Coincidence and the Grounding Problem. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 118, 339–371.
- Bennett, K. (2008). Exclusion again. In J. Hohwy and J. Kallestrup (Eds.), *Being Reduced: New Essays on Reduction, Explanation, and Causation*. New York: Oxford University Press.
- Bennett, K. (2011). Construction area (no hard hat required). *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 154, 79–104.
- Bennett, K. (2017). *Making Things Up*. Oxford University Press.
- Bickle, J. (2016). Multiple Realizability. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016). Metaphysics Research Lab, Stanford University.
- Bigelow, J. (1996). Presentism and Properties. *Philosophical Perspectives*, 10, 35–52.
- Biro, J. (2017). Constitution and Identity. *Erkenntnis*, 83, 1127–1138.
- Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O., and Piddock, L. J. V. (2014). Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*, 13, 42–51.
- Block, N. (1997). Anti-Reductionism Slaps Back. *Noûs*, 31, 107–132.
- Block, N., and Fodor, J. A. (1972). What Psychological States are Not. *The Philosophical Review*, 81, 159–181.
- Boyd, R. (1999). Kinds, complexity and multiple realization. *Philosophical Studies*, 95, 67–98.
- Burge, T. (1986). Individualism and psychology. *The Philosophical Review*, 95,

3–45.

- Burke, M. B. (1992). Copper Statues and Pieces of Copper: A Challenge to the Standard Account. *Analysis*, 52, 12–17.
- Bursten, J. R. (2014). Microstructure without Essentialism: A New Perspective on Chemical Classification. *Philosophy of Science*, 81, 633–653.
- Carnap, R. (1967). *The Logical Structure of the World*. University of California Press.
- Carr, N. G. (2010). *The Shallows: What the Internet is Doing to Our Brains*. W.W. Norton.
- Chisholm, R. M. (1973). Parts as Essential to Their Wholes. *The Review of Metaphysics*, 581–603.
- Clark, A. (2005). Intrinsic Content, Active Memory and the Extended Mind. *Analysis*, 65, 1–11.
- Clark, A. (2010a). Coupling, constitution and the cognitive kind: A reply to Adams and Aizawa. In R. Menary (Ed.), *The Extended Mind* (pp. 81–99). MIT Press.
- Clark, A. (2010b). Memento’s revenge: The extended mind extended. In R. Menary (Ed.), *The Extended Mind* (pp. 43–66). MIT Press.
- Clark, A. (2010c). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press.
- Clark, A., and Chalmers, D. (1998). The Extended Mind. *Analysis*, 58, 7–19.
- Cotnoir, A. J., and Baxter, D. L. M. (2014). *Composition as Identity*. Oxford University Press, USA.
- Craver, C. F. (2007). Constitutive Explanatory Relevance. *Journal of Philosophical Research*, 32, 3–20.
- Craver, C. F. (2014). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Clarendon Press.
- Decock, L., and Douven, I. (2011). Similarity After Goodman. *Review of*

Philosophy and Psychology, 2, 61–75.

- Descartes, R. (1996). Selections from Meditations on First Philosophy. In P. A. Morton (Ed.), *A Historical Introduction to the Philosophy of Mind: Readings with Commentary*. Broadview Press.
- Deutsch, H. (2017). Relative Identity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017). Metaphysics Research Lab, Stanford University.
- Doige, N. (2007). *The Brain That Changes Itself: Stories of Personal Triumph from the Frontiers of Brain Science*. Penguin Group US.
- Elder, C. (2003). Alexander’s Dictum and the Reality of Familiar Objects. *Topoi*, 22, 163–171.
- Endicott, R. (2016a). Developing the explanatory dimensions of part–whole realization. *Philosophical Studies*, 173, 3347–3368.
- Endicott, R. (2016b). Functionalism, superduperfunctionalism, and physicalism: Lessons from supervenience. *Synthese*, 193, 2205–2235.
- Fodor, J. A. (1968). *Psychological Explanation: An Introduction to the Philosophy of Psychology*. Random House.
- Fodor, J. A. (1974). Special Sciences (Or: The Disunity of Science as a Working Hypothesis). *Synthese*, 28, 97–115.
- Fodor, J. A. (1981). The Mind-Body Problem. *Scientific American*, 244, 114–125.
- Fodor, J. A. (1997). Special Sciences: Still Autonomous after All these Years. *Noûs*, 31, 149–163.
- Frigg, R., and Nguyen, J. (2016). Scientific Representation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University.
- Gallagher, S. (2013). The socially extended mind. *Cognitive Systems Research*, 25, 4–12.

- Gallois, A. (2016). Identity Over Time. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University.
- Gärdenfors, P. (2004). *Conceptual Spaces: The Geometry of Thought*. MIT Press.
- Gärdenfors, P. (2014). *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. MIT Press.
- Geach, P. T. (1980). *Reference and Generality: An Examination of Some Medieval and Modern Theories*. Cornell University Press.
- Gillett, C. (2002a). The dimensions of realization: A critique of the Standard view. *Analysis*, 62, 316–323.
- Gillett, C. (2002b). The metaphysics of realization, multiple realizability, and the special sciences. *The Journal of Philosophy*, 100, 591–603.
- Gillett, C. (2007). Understanding the New Reductionism: The Metaphysics of Science and Compositional Reduction. *The Journal of Philosophy*, 104, 193–216.
- Gillett, C. (2013). Constitution, and Multiple Constitution, in the Sciences: Using the Neuron to Construct a Starting Framework. *Minds and Machines*, 23, 309–337.
- Goodman, N. (1972). *Problems and Projects*. Indianapolis: Bobbs-Merrill.
- Goodman, N. (2012). *The Structure of Appearance*. Springer Science & Business Media.
- Grandy, R. E. (2016). Sortals. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University.
- Heller, M., Heller, P. M., and Mark, H. (1990). *The Ontology of Physical Objects: Four-Dimensional Hunks of Matter*. Cambridge University Press.
- Homer. (1919). *The Odyssey* (Vols. 1–2; A. T. Murray, Trans.). W. Heinemann.
- Horgan, T. (1993). From Supervenience to Superdupervenience: Meeting the

- Demands of a Material World. *Mind*, 102, 555–586.
- Hubble, P. (2009). *A Certainty of Death: Appreciating Human Animalhood* (Institute for Christian Studies). Institute for Christian Studies.
- Hurley, S. L. (2010). Varieties of Externalism. In R. Menary (Ed.), *The Extended Mind* (pp. 101–153). MIT Press.
- Johnston, M. (1992). Constitution Is Not Identity. *Mind*, 101, 89–105.
- Kim, J. (1984). Concepts of Supervenience. *Philosophy and Phenomenological Research*, 45, 153–176.
- Kim, J. (1987). "Strong" and "Global" Supervenience Revisited. *Philosophy and Phenomenological Research*, 48, 315–326.
- Kim, J. (1992a). "Downward Causation" in Emergentism and Nonreductive Physicalism. In A. Beckermann and H. Flohr (Eds.), *Emergence or Reduction?: Essays on the Prospects of Nonreductive Physicalism*. Walter de Gruyter.
- Kim, J. (1992b). Multiple Realization and the Metaphysics of Reduction. *Philosophy and Phenomenological Research*, 52, 1–26.
- Kim, J. (1997). Does the Problem of Mental Causation Generalize? *Proceedings of the Aristotelian Society*, 97, 281–297.
- Kim, J. (1999). Making sense of emergence. *Philosophical Studies*, 95, 3–36.
- Kim, J. (2000). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation* (Reprint edition). Cambridge, Mass.: A Bradford Book.
- Kim, J. (2003). Blocking Causal Drainage and Other Maintenance Chores with Mental Causation¹. *Philosophy and Phenomenological Research*, 67, 151–176.
- Kim, J. (2006). Emergence: Core ideas and issues. *Synthese*, 151, 547–559.
- Kim, J. (2012). The very idea of token physicalism. In S. Gozzano and C. S. Hill (Eds.), *New perspectives on type identity: The mental and the physical*

- (pp. 167–193). Cambridge University Press.
- Kingsbury, J. (2004). Biologising the Mind. *Biology and Philosophy*, 19, 473–482.
- Kingsbury, J. (2006). A Proper Understanding of Millikan. *Acta Analytica*, 21, 23–40.
- Kingsbury, J. (2008). Learning and selection. *Biology & Philosophy*, 23, 493–507.
- Kirchhoff, M. D. (2015). Extended Cognition & the Causal-Constitutive Fallacy: In Search for a Diachronic and Dynamical Conception of Constitution. *Philosophy and Phenomenological Research*, 90, 320–360.
- Korman, D. Z. (2016). Ordinary Objects. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016). Metaphysics Research Lab, Stanford University.
- Koslicki, K. (2015). The coarse-grainedness of grounding. In *Oxford Studies in Metaphysics, Volume 9* (Vol. 9). Oxford University Press.
- Langer, E. J., and Imber, L. G. (1979). When practice makes imperfect: Debilitating effects of overlearning. *Journal of Personality and Social Psychology*, 37, 2014–2024.
- Leuridan, B. (2012). Three Problems for the Mutual Manipulability Account of Constitutive Relevance in Mechanisms. *The British Journal for the Philosophy of Science*, 63, 399–427.
- Lewis, D. (1973). Causation. *The Journal of Philosophy*, 70, 556–567.
- Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, 61, 343–377.
- Lewis, D. (1991). *Parts of Classes*. B. Blackwell.
- Lewis, D. (2000). Causation as Influence. *The Journal of Philosophy*, 97, 182–197.
- Locke, J. (1841). *An Essay Concerning Human Understanding*. Retrieved from

<http://books.google.com?id=cjYIAAAAQAAJ>

- Mackenzie, C. (2014). Embodied agents, narrative selves. *Philosophical Explorations*, 17, 154–171.
- Maclaurin, J., and Sterelny, K. (2008). *What Is Biodiversity?* University of Chicago Press.
- McLaughlin, B., and Bennett, K. (2014). Supervenience. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014). Metaphysics Research Lab, Stanford University.
- Menary, R. (2010). *The Extended Mind*. MIT Press.
- Merleau-Ponty, M. (1996). *Phenomenology of Perception*. Motilal Banarsidass Publishers.
- Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. MIT Press.
- Millikan, R. G. (2017). *Beyond Concepts: Unicepts, Language, and Natural Information*. Oxford University Press.
- Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation*. London: Routledge & Kegan Paul.
- Noë, A. (2010). *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness* (1 edition). New York: Hill and Wang.
- Noonan, H., and Curtis, B. (2017). Identity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017). Metaphysics Research Lab, Stanford University.
- Noonan, H. W. (1993). Constitution is Identity. *Mind*, 102, 133–146.
- O'Connor, T., and Wong, H. Y. (2005). The Metaphysics of Emergence. *Noûs*, 39, 658–678.
- Olson, E. T. (1999). *The Human Animal: Personal Identity without Psychology*. Oxford University Press.

- Olson, E. T. (2007). *What Are We?: A Study in Personal Ontology*. Oxford University Press.
- Parfit, D. (1986). *Reasons and Persons*. OUP Oxford.
- Parfit, D. (2012). We Are Not Human Beings. *Philosophy*, 87, 5–28.
- Paul, L. A. (1998). Keeping Track of the Time: Emending the Counterfactual Analysis of Causation. *Analysis*, 58, 191–198.
- Pereboom, D. (2002). Robust nonreductive materialism. *The Journal of Philosophy*, 99, 499–531.
- Plato. (1914). *Plato in twelve volumes: With an English translation* (H. N. Fowler, W. R. M. Lamb, R. G. Bury, and P. Shorey, Trans.). London; Cambridge: W. Heinemann; Harvard University Press.
- Polger, T. W. (2007). Realization and the metaphysics of mind. *Australasian Journal of Philosophy*, 85, 233–259.
- Polger, T. W. (2009). Evaluating the evidence for multiple realization. *Synthese*, 167, 457–472.
- Polger, T. W., and Shapiro, L. A. (2008). Understanding the Dimensions of Realization. *The Journal of Philosophy*, 105, 213–222.
- Polger, T. W., and Shapiro, L. A. (2016). *The Multiple Realization Book*. Oxford University Press.
- Price, E. H. (2006). A critical review of congenital phantom limb cases and a developmental theory for the basis of body image. *Consciousness and Cognition*, 15, 310–322.
- Prior, E. W., Pargetter, R., and Jackson, F. (1982). Three Theses about Dispositions. *American Philosophical Quarterly*, 19, 251–257.
- Putnam, H. (1967). Psychological Predicates. In W. H. Capitan and D. D. Merrill (Eds.), *Art, Mind, and Religion*. University of Pittsburgh Press.
- Putnam, H. (1975a). Minds and Machines. In *Mind, Language and Reality: Vol. Philosophical Papers: Volume 2*. Cambridge University Press.

- Putnam, H. (1975b). The Meaning of "Meaning". In K. Gunderson (Ed.), *Language, Mind, and Knowledge*. University of Minnesota Press.
- Putnam, H. (1979). *Philosophical Papers: Volume 2, Mind, Language and Reality*. Cambridge University Press.
- Quine, W. V. (1969). Natural Kinds. In N. Rescher (Ed.), *Essays in Honor of Carl G. Hempel: A Tribute on the Occasion of his Sixty-Fifth Birthday* (pp. 5–23). Dordrecht: Springer Netherlands.
- Ramsey, W. (2003). *Eliminative Materialism*. Retrieved from <https://plato.stanford.edu/archives/win2016/entries/materialism-eliminative/#BriHis>
- Rescher, N. (1991). Translation of the Monadology of Gottfried Wilhelm von Leibniz. In G. W. Leibniz's *Monadology* (pp. 17–30). University of Pittsburgh Press.
- Ross, D., and Ladyman, J. (2010). The Alleged Coupling-Constitution Fallacy and the Mature Sciences. In R. Menary (Ed.), *The Extended Mind* (pp. 155–166). MIT Press.
- Rupert, R. D. (2010). Representation in Extended Cognitive Systems: Does the Scaffolding of Language Extend the Mind? In R. Menary (Ed.), *The Extended Mind* (pp. 325–354). MIT Press.
- Schaffer, J. (2010). Monism: The Priority of the Whole. *The Philosophical Review*, 119, 31–76.
- Schaffer, J. (2012). Grounding, transitivity, and contrastivity. In F. Correia and B. Schnieder (Eds.), *Metaphysical Grounding: Understanding the Structure of Reality* (pp. 122–138). Cambridge University Press.
- Schechtman, M. (2014). *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. OUP Oxford.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3, 417.
- Shakya, H. B., and Christakis, N. A. (2017). Association of Facebook Use With Compromised Well-Being: A Longitudinal Study. *American Journal*

- of Epidemiology*, 185, 203–211.
- Shapiro, L. A. (2000). Multiple Realizations. *The Journal of Philosophy*, 97, 635–654.
- Shapiro, L. A. (2008). How to Test for Multiple Realization. *Philosophy of Science*, 75, 514–525.
- Shapiro, L. A., and Polger, T. W. (2012). Identity, variability, and multiple realization in the special sciences. In S. Gozzano and C. S. Hill (Eds.), *New Perspectives on Type Identity: The Mental and the Physical* (pp. 264–288). Cambridge University Press.
- Shoemaker, D. (2016). Personal Identity and Ethics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University.
- Shoemaker, S. (2003). Realization, Micro-Realization, and Coincidence. *Philosophy and Phenomenological Research*, 67, 1–23.
- Smid, J. (2015). The ontological parsimony of mereology. *Philosophical Studies*, 172, 3253–3271.
- Sober, E. (1999). The Multiple Realizability Argument against Reductionism. *Philosophy of Science*, 66, 542–564.
- Sterelny, K. (2004). Externalism, epistemic artefacts and the extended mind. *The Externalist Challenge*, 239–254.
- Takeuchi, H., Taki, Y., Asano, K., Asano, M., Sassa, Y., Yokota, S., ... Kawashima, R. (2018). Impact of frequency of internet use on development of brain structures and verbal intelligence: Longitudinal analyses. *Human Brain Mapping*, 39, 4471–4479.
- Thelen, E., and Smith, L. B. (1996). *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352.
- Unger, P. (1979). There Are No Ordinary Things. *Synthese*, 41, 117–154.

- van Inwagen, P. (1995). *Material Beings*. Cornell University Press.
- Varzi, A. (2016). Mereology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University.
- Wachowski, L., and Wachowski, L. (Directors). (1999). *The Matrix*. Retrieved from <http://www.imdb.com/title/tt0133093/>
- Wallace, M. (2011a). Composition as identity: Part 1. *Philosophy Compass*, 6, 804–816.
- Wallace, M. (2011b). Composition as identity: Part 2. *Philosophy Compass*, 6, 817–827.
- Walter, S., and Kyselo, M. (2009). Review of The Bounds of Cognition. *Erkenntnis* (1975-), 71, 277–281.
- Wasserman, R. (2004). The Constitution Question. *Noûs*, 38, 693–710.
- Wasserman, R. (2017). Material Constitution. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017). Metaphysics Research Lab, Stanford University.
- Weatherson, B. (2016). The Problem of the Many. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University.
- Weir, P. (Director). (1998). *The Truman Show*. Retrieved from <http://www.imdb.com/title/tt0120382/>
- Weisberg, M. (2012). Getting Serious about Similarity. *Philosophy of Science*, 79, 785–794.
- Wilson, J. (2009). Determination, Realization and Mental Causation. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 145, 149–169.
- Wilson, J. (2011). Non-reductive Realization and the Powers-based Subset Strategy. *The Monist*, 94, 121–154.
- Wilson, J. (2014). No Work for a Theory of Grounding. *Inquiry*, 57, 535–579.

- Wilson, J. (2015). Metaphysical Emergence: Weak and Strong. *Metaphysics in Contemporary Physics*, 251–306.
- Wilson, R. A. (2001). Two Views of Realization. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 104, 1–31.
- Wilson, R. A. (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences*. Cambridge University Press.
- Wilson, R. A. (2007). A Puzzle About Material Constitution and How to Solve It: Enriching Constitution Views in Metaphysics. *Philosopher's Imprint*, 7. Retrieved from <http://hdl.handle.net/2027/spo.3521354.0007.005>
- Wilson, R. A. (2009). The Transitivity of Material Constitution. *Noûs*, 43, 363–377.
- Wilson, R. A., and Craver, C. F. (2007). Realization: Metaphysical and Scientific Perspectives. In P. Thagard (Ed.), *Philosophy of Psychology and Cognitive Science* (pp. 81–104). Elsevier.
- Witmer, G. (2003). Multiple Realizability and Psychological Laws: Evaluating Kim's Challenge. In S. Walter and H.-D. Heckmann (Eds.), *Physicalism and Mental Causation: The Metaphysics of Mind and Action* (pp. 59–84). Imprint Academic.
- Yi, B. (2018). Nominalism and Comparative Similarity. *Erkenntnis*, 83, 793–803.
- Ylikoski, P. (2013). Causal and Constitutive Explanation Compared. *Erkenntnis*, 78, 277–297.
- Zimmerman, D. W. (1995). Theories of Masses and Problems of Constitution. *The Philosophical Review*, 104, 53–110.