

Metaphysical Emergence

JESSICA WILSON

Department of Philosophy
University of Toronto
jessica.m.wilson@utoronto.ca

July 25, 2018

Contents

1 Key issues and questions	1
1.1 Metaphysical emergence: dependence with autonomy	1
1.1.1 The <i>prima facie</i> motivations	3
1.1.2 Two key questions	10
1.1.3 Outline of the book	19
1.2 Preliminaries	28
1.2.1 The target cases	28
1.2.2 The individuation of levels	33
1.2.3 The operative notion of the physical	41
1.2.4 A metaphysically neutral understanding of powers	46
1.2.5 Some brief remarks on methodology	50
2 Two schemas for emergence	53
2.1 The problem of higher-level causation	54
2.1.1 Kim’s overdetermination argument	55
2.1.2 The two ‘emergentist’ strategies for responding to the problem	60
2.2 Strong emergentism and the <i>New Power Condition</i>	61
2.2.1 The <i>New Power Condition</i>	64
2.2.2 The schema for Strong emergence	67
2.3 Non-reductive physicalism and the <i>Proper Subset of Powers Condition</i>	68
2.3.1 The <i>Proper Subset of Powers Condition</i>	70
2.3.2 The schema for Weak emergence	85
2.4 Merricks’s overdetermination argument	89
2.4.1 The status of configurations	94
2.5 The schemas for Strong and Weak emergence as core and crucial to metaphysical emergence	96

3 The viability of Weak emergence	99
3.1 Objection: anti-realism about higher-level features	101
3.1.1 Pragmatic/abstractionist anti-realism	101
3.1.2 Explanatory gap eliminativism	108
3.2 Objection: non-satisfaction of the <i>Proper Subset of Powers Condition</i>	110
3.2.1 Failure to motivate satisfaction of the <i>Proper Subset of Powers Condition</i>	110
3.2.2 Token multiple realizability as a barrier to satisfaction of the condition	115
3.3 Objection: compatibility with reductionism	117
3.3.1 Reduction to a conjunct of a lower-level conjunction?	118
3.3.2 Reduction to a disjunction of lower-level disjuncts?	122
3.3.3 Reduction to a metaphysical consequence of lower-level laws?	126
3.4 Objection: compatibility with physical unacceptability	129
3.4.1 Quiddities	131
3.4.2 Phenomenal aspects	134
3.4.3 Historical aspects and ‘backwards-facing’ powers	140
3.4.4 Conjuncts and conjunctions	142
3.4.5 Physically unacceptable individuation	143
3.4.6 Fundamentally mental powers	147
3.5 Objection: non-necessity	149
3.5.1 Token identity	150
3.5.2 Constitution	152
3.5.3 Grounding	155
3.6 Concluding remarks	159
4 The viability of Strong emergence	161
4.1 Incompatibility with physics, scientific practice, naturalism	163
4.2 Collapse	166
4.2.1 Collapse via power possession	167
4.2.2 Collapse via lower-level dispositions	169
4.2.3 Three responses to the collapse objection	172
4.3 Objection: non-necessity	182
4.3.1 Epiphenomenalism	183
4.3.2 Supervenience	184
4.3.3 Epistemic criteria	193

4.4	Concluding remarks	196
5	Complex systems	199
5.1	Are complex systems Strongly emergent?	201
5.1.1	Nonlinearity and unpredictability in the British emergentist tradition	202
5.1.2	The fall of nonlinearity and unpredictability as guides to fundamental novelty	206
5.1.3	Are all nonlinear phenomena at best Weakly emergent? . .	209
5.1.4	Nonlinearity's descendant	212
5.2	Are complex systems Weakly emergent?	214
5.2.1	Bedau's appeal to incompressibility	215
5.2.2	Mitchell's appeal to self-organization	219
5.2.3	Batterman's appeal to asymptotic singularities	222
5.2.4	Eliminations in degrees of freedom	229
5.3	Concluding remarks	247
6	Ordinary objects	249
6.1	Are ordinary objects Weakly emergent?	251
6.1.1	Classical objects	251
6.1.2	Sortal features and functional realization	258
6.1.3	Metaphysically indeterminate boundaries	265
6.2	Are ordinary objects Strongly emergent?	279
6.2.1	Two routes to the Strong emergence of artifacts . . .	279
6.3	Concluding remarks	282
7	Consciousness	287
7.1	Is consciousness Strongly emergent?	291
7.1.1	The knowledge arguments	291
7.1.2	The conceivability argument	303
7.2	Is consciousness Weakly emergent?	322
7.2.1	Determinable perceptions	322
7.2.2	The objections from mental multiple realizability and mental super=determinates	326
7.2.3	Responses to the objections	330
7.3	Concluding remarks	337

8 Free will	339
8.1 The generalized problem of mental quausation	341
8.2 Compatibilism and Weak emergence	346
8.2.1 Weak emergence and the non-reductive physicalist's proper subset strategy	347
8.2.2 The compatibilist's proper subset strategy	349
8.2.3 Deepening the parallel: a powers-based interpretation of the compatibilist's proper subset condition	355
8.3 Libertarianism and Strong emergence	358
8.3.1 Event-causal accounts	362
8.3.2 Agent-causal accounts	366
8.3.3 Noncausal accounts	368
8.4 Is free will either Weakly or Strongly emergent?	372
8.4.1 Is there compatibilist (Weakly emergent) free will?	372
8.4.2 Is there libertarian (Strongly emergent) free will?	373
8.5 Concluding remarks	381

Chapter 1

Key issues and questions

1.1 Metaphysical emergence: dependence with autonomy

Consider some mid-range macroscopic special science entities of the sort we can or do ordinarily experience: hurricanes (treated by meteorology), trees (treated by botany), birds (treated by zoology), and humans (treated by psychology). Such macro-entities, the scientists tell us, are dependent on complex configurations of smaller, typically less complex, and ultimately fundamental physical entities—‘micro-configurations’, for short—in that, at any given time, a macro-entity is materially constituted by some micro-configuration at that time, and a macro-entity’s features at a time (or over a temporal interval) are at least partly determined by features of its underlying micro-configuration at that time (over that interval). Notwithstanding this synchronic material dependence, however, these special science entities also seem, from both theoretical and experiential points of view (to be discussed in more detail shortly), to possess a certain degree of ontological and causal autonomy—that is, they appear to be *distinct* from and *distinctively efficacious* as compared to the micro-configurations upon which they depend. The same is true for special science entities at smaller and larger scales of existence (which are still broadly ‘macroscopic’ by way of comparison with fundamental physical

entities), including molecules (treated by chemistry), gasses (treated by thermodynamics), organelles (treated by cellular biology), ecosystems (treated by ecology), galaxies (treated by astronomy), and so on: on the one hand, these macro-entities are synchronically materially dependent on lower-level, and ultimately fundamental physical, configurations; yet on the other hand, there are cases to be made that these entities are also ontologically and causally autonomous from—distinct from and distinctively efficacious as compared to—the micro-configurations upon which they so depend.

It is the coupling of *synchronic material dependence* with *ontological and causal autonomy* which is most basically definitive of the notion of emergence, at least as suggested by the central cases of special science entities *vis-á-vis* the lower-level, and ultimately fundamental physical, configurations which are their constant companions.¹ This general notion of emergence is also motivated by attention to cases of artifacts, including buildings, sculptures, furniture, and so on; for these too are reasonably taken to be synchronically materially dependent on, yet distinct from and distinctively efficacious as compared to, underlying micro-configurations. And insofar as the dependence, distinctness, and distinctive efficacy seemingly at issue in all these cases appear to be characteristic of the entities themselves, as opposed to what we can know or represent about such entities, the notion of emergence initially motivated by these cases is that of distinctively *metaphysical* emergence.

As we'll see, there are questions about how to interpret the appearances of metaphysical emergence, and relatedly, about whether these appearances can be taken at realistic face value. These and related questions will occupy much of this book. First, though, it's worth saying a bit more about the *prima facie* motivations for there being metaphysical emergence, to substantiate that there is good reason to attend to the content, viability, and applicability of this interesting notion.

¹Note that the synchronic dependence at issue here is broad, in being compatible with holding over a temporal interval. The operative broad notion of synchronic (material) dependence is intended to contrast with diachronic relations (e.g., causation, as usually understood, where the relata occur or obtain at different times).

1.1.1 The *prima facie* motivations

Synchronic material dependence

Why think that macro-entities, including both special science entities and artifacts, are synchronically materially dependent, in the sense described above, on lower-level, and ultimately fundamental physical micro-configurations? The motivations here are broadly empirical, and reflect common scientific consensus on two points, of the sort that any physics or special science textbook is likely to confirm. First is that the only matter or substance is physical matter or substance, such that any and all macro-entities (including objects, systems, and so on) are, at each time of their existence, materially constituted by lower-level and ultimately fundamental physical configurations—that is, by lower-level and ultimately fundamental physical entities, understood as standing in certain spatiotemporal and other relations. Second is that the features (properties, states, and so on) had by any given macro-entity at a given time (over a temporal interval) are at least in part a function of the properties of the micro-configuration(s) which materially constitute the macro-entity at that time (over that temporal interval).

The introductory section of practically any scientific textbook is likely to register one or both of these baseline scientific assumptions, as in these examples, drawn at random from my bookshelf:

Everything is made of atoms. That is the key hypothesis. The most important hypothesis in biology, for example, is that [...] *there is nothing that living things do that cannot be understood from the point of view that they are made of atoms acting according to the laws of physics.* [...] This was not known from the beginning: it took some experimenting and theorizing to suggest this hypothesis, but now it is accepted [...]. If a piece of steel or a piece of salt, consisting of atoms next to the other, can have such interesting properties [...] is it possible that that “thing” walking back and forth in front of you, talking to you, is a great glob of these atoms in a very complex arrangement, such that the sheer complexity of it staggers the imagination as to what it can do? When we say we are a pile of atoms, we do not mean we are *merely* a pile of atoms [...]. Feynman 1963, Vol I, 1-9)

Quarks and leptons are the fundamental objects of which all matter is composed; they interact via the exchange of gauge bosons. [Kane 1993](#), 1)

In the present state of scientific knowledge, quantum mechanics plays a fundamental role in the description and understanding of natural phenomena. [...] when we are concerned only with macroscopic physical objects [...] it is necessary, in principle, to begin by studying the behaviour of their various constituent atoms, ions, electrons, in order to arrive at a complete scientific description. There are many phenomena which reveal, on a macroscopic scale, the quantum behaviour of nature. It is in this sense that it can be said that quantum mechanics is the basis of our present understanding of all natural phenomena, including those traditionally treated in chemistry, biology, etc. ([Cohen-Tannoudji et al. 1977](#), 9).

Every day, whether we know it or not, we witness changes in matter that are a result of the properties of the atoms and molecules composing that matter—ice melts, iron rusts, gasoline burns, fruit ripens, water evaporates. [...] If we want to understand the substances around us, we must understand the physical and chemical properties of the atoms and molecules that compose them—this is the central goal of chemistry. ([Tro et al. 2017](#), 2)

That there is *prima facie* motivation for taking macro-entities to be synchronically materially dependent on micro-configurations is also reflected in Kim's ([1993a](#)) description of the standard background picture that is the starting point of philosophical investigations into reduction and emergence, as presupposing ...

a hierarchically stratified structure of ‘levels’ or ‘orders’ of entities and their characteristic properties. It is generally thought that there is a bottom level, one consisting of whatever microphysics is going to tell us are the most basic physical particles out of which all matter is composed [...] As we ascend to higher levels, we find [entities] that are made up of entities belonging to the lower levels, and, moreover, the entities at any given level are thought to be characterized by a set of properties distinctive of that level. (190)

Kim's sketch leaves a lot open, as do the usual introductory gestures in science

textbooks, but the core idea that special science entities and artifacts are synchronically materially dependent on configurations of ultimately fundamental physical entities is about as uncontroversial as it gets in either science or philosophy. That said, as we will see there is controversy about exactly how to understand the operative sense(s) of dependence at issue here; moreover, we will also later discuss views, such as substance dualism and panpsychism, which reject the appearances of material dependence, at least on some understandings of such dependence. But at this point we are simply canvassing *prima facie* reasons to think that there is synchronic material dependence in the cases at issue; and there isn't any real question about there being an enormous amount of empirical support for such a view.

Ontological and causal autonomy

There is more to say about the *prima facie* motivations for the autonomy of synchronically materially dependent entities of either scientific or artifactual varieties. Three such motivations have to do with how special science entities are classified or characterized, as follows:

- *Distinctive special-scientific taxonomy*: Special science entities are classified as falling under distinctive types. As above, molecules, gasses, cells, hurricanes, trees, birds, humans, ecosystems, and galaxies are classified as falling, respectively, under distinctive types in chemistry, thermodynamics, biology, meteorology, botany, zoology, psychology, ecology, and astronomy. On the face of it, these special science types are different from those under which lower-level, and ultimately fundamental physical, entities fall, even when the latter occur in complex combinations or configurations. Classification practices thus provide *prima facie* support for the ontological autonomy of special science entities.
- *Distinctive special-science features*: Special science entities are characterized as having distinctive features, constitutive of the distinctive types under which they fall. A tree, for example, has roots, a trunk, branches, stems,

leaves; it obtains nutrients from air, sun, soil, and water through leaves and roots; it reproduces via seeds and may bear fruit; it is deciduous or evergreen; it is hardy in certain climate zones, and so on. On the face of it, such features are not appropriately attributed to even complex configurations of atoms or other lower-level entities; and the same is true for the characteristic features of other special science entities. The mental features of human persons are especially distinctive, as involving, among other salient characteristics, qualitative aspects of sensory experience; representational aspects of belief, desire, intention, and other mental states; appreciation of aesthetic, moral, and other broadly normative values; and seemingly free agency. By Leibniz's Law, of course, entities with different features (at the same time, at least) are distinct; hence that special science entities are characterized as having features different from the micro-configurations upon which they synchronically depend again supports special science entities' being ontologically autonomous.

- *Distinctive special science laws:* Special science entities are taken to be governed by special science laws (or regularities) describing states, properties and behaviours of, and associated causal interactions involving, such entities—laws that, on the face of it, are different from those governing even complex configurations of lower-level, ultimately physical entities. Of course, there is a deep question here (which will be a primary focus in what follows) about how to understand this higher-level efficacy, and whether such efficacy is compatible with the assumed efficacy of micro-configurations and constituent micro-entities.² At present, we can say this much: that there are seemingly distinctive, seemingly causal special science

²It is sometimes claimed, typically by appeal to the discussion in Russell 1912, that physics has as dispensed with the notion of ‘cause’, but Russell’s complaints are directed only against a certain implausible conception of causation, understood as involving universal generalizations. Nor does the fact that physical equations are expressed in broadly mathematical rather than explicitly causal terms imply that lower-level physical goings-on are not efficacious. We will return to the issue of lower-level physical efficacy down the line. For now, observe that even if it could be made out that physical goings-on were not properly seen as causal, the *prima facie* causal autonomy of special science entities would remain in place.

laws provides *prima facie* support for special science entities’ being causally autonomous as compared to their underlying micro-configurations.

Two other scientific reasons for thinking that special science entities are autonomous reflect, more specifically, that special science types, features, and laws often abstract from details relevant to characterizing lower-level types, features, and laws:

- *Universal properties and behaviour*: Many special science entities, including thermodynamic complex systems such as liquids and gasses, exhibit features that are functionally independent of various features of their underlying micro-configurations. For example, complex systems near critical points exhibit “universal” properties and behavior across widely diverse underlying configurations of molecules. Such cases provide support for thinking that some special science entities are causally, hence ontologically, autonomous, in that their behaviours are sensitive to a comparatively abstract level of causal grain.
- *Elimination of micro-physical ‘degrees of freedom’*: Degrees of freedom are independent parameters needed to specify an entity’s law-governed states and behaviours. It turns out to be characteristic of many special science entities that certain of their states and behaviours are specifiable by reference to strictly fewer degrees of freedom than are needed to specify the law-governed states and behaviours of the micro-configurations upon which they depend. For example, rigid bodies depend upon quantum-mechanical configurations, but the behaviour of rigid bodies does not generally rely upon certain quantum mechanical degrees of freedom, such as spin direction and magnitude. That specification of the law-governed states and behaviours of rigid bodies requires strictly fewer degrees of freedom than those of lower-level quantum-mechanical configurations again suggests that the former are causally and ontologically autonomous from the latter.

Four other motivations for autonomy apply to special science goings-on and artifacts of the sort available to ordinary experience, and reflect certain perceptual,

individuative, and semantic considerations, in ways that pick up on and confirm the scientific characterizations of such entities as having features which abstract away from certain micro-level details:

- *Perceptual unity*: Though the macro-entities of our acquaintance are, the scientists tell us, materially constituted by massively complex and constantly changing micro-configurations (ultimately involving whatever fundamental physical goings-on there might be), macro-entities do not perceptually appear to us as massively complex, constantly changing, configurations of micro-phenomena. A tree, for example, does not look like a complicated structure of cells or tissues, much less like a buzzing array of sub-atomic particles or other physical fundamenta; rather, a tree looks like a comparatively stable and unified entity, both at and over time; and the same is true for other familiar macro-entities. Moreover, in the case of human persons there is the evidence of introspection (a kind of internal perception) of ourselves as fairly unified and persisting “selves”.
- *Compositional flexibility*: The identity of many macro-entities appears to transcend that of their underlying micro-configurations, in the sense that the existence of a given special science entity (a cell, a hurricane, a tree, a bird, a human) does not depend on the existence of any *specific* micro-configuration or configurations, and relatedly, in that many macro-entities, both in scientific practice and in our ordinary practices of individuation, are understood as being capable of surviving even large changes in their underlying configurations.
- *Proper names and definite descriptions*: Our practices of giving names or definite descriptions to certain special science entities—the Atlantic Ocean, Hurricane Katrina, Benj Hellie—suggests that we often treat macro-entities as individuals, distinct from the ever-changing micro-configurations upon which they materially depend.
- *Truth and meaning*: Many of the sentences we take to be true appear to be about macroscopic goings-on, and relatedly, to contain subject terms

which appear to denote macro-entities (as either token individuals or types) and to contain predicate terms which appear to be used to attribute macro-features to these macro-entities—as with, e.g., ‘That table is well-made’, ‘Tigers typically have tails’, ‘Her novels are spellbinding’, and countless other examples.

One final motivation for ontological and causal autonomy is worth mentioning, as especially dear to our hearts and minds:

- *Seemingly free will*: The status of our actions as genuinely free is one that remains up for debate. However, a starting point in this debate is that it at least introspectively seems that creatures like ourselves are capable of making free choices to produce certain effects (or to intend to produce certain effects), where this efficacy appears to be quite different from that associated with the (deterministically or indeterministically) lawfully governed micro-configurations upon which we and our mental states depend.

Summing up: on the face of it, a number of considerations, drawn from science, perception, our practices of individuation, language, and introspective experience, provide *prima facie* support for thinking that many broadly natural entities are *synchronously materially dependent* on micro-configurations of less complex, ultimately fundamental physical, entities, yet are also *ontologically and causally autonomous* as compared to these underlying micro-configurations—that is, are metaphysically emergent.³ All this constitutes good reason to think that, prima

³Again, the notion of synchronic dependence at issue here is broad, in applying either at a time or over a temporal interval—i.e., is one allowing for temporally extended base and dependent phenomena. As for autonomy: throughout I distinguish ontological autonomy (distinctness) from causal autonomy (distinctive causal efficacy), and I assume that both are required of an account of metaphysical emergence aiming to vindicate special science entities as entering into distinctive (typically causal) laws or (if such there be) capable of making free choices; this assumption also reflects that causal as well as ontological autonomy is constitutive of the distinctively emergentist responses to the problem of higher-level causation, as I will later discuss. Of course, causal autonomy entails ontological autonomy, by Leibniz’s law. Ontological autonomy is compatible with an absence of causal autonomy, however, as with epiphenomenalist accounts of higher-level entities; correspondingly, though epiphenomenalist accounts of higher-level goings-on are occasionally presented as accounts of “emergence” (see Chalmers 2006a), they are not so in the sense at issue here.

facie, there is metaphysical emergence. And indeed, it is common to take the appearances of emergence to be the starting point of investigations into broadly natural reality, even by those who arrive at a different endpoint. We saw one example of this in Kim's description of the levelled hierarchy; for another, here Heil (2003b) prefaces his anti-levels argumentation (which we will consider in due course) by saying:

Many people these days express a belief that our world comprises *levels of reality*. In philosophy this idea is encountered in metaphysics, philosophy of science, and most especially in philosophy of mind. Talk of levels, of course, is by no means restricted to philosophers. Biologists, psychologists, anthropologists, historians, journalists [...] routinely appeal to 'higher-level' and 'lower-level' phenomena in discussions of a variety of topics. Reality, it is widely presumed, is hierarchical. Although items occupying higher levels are thought to be in some fashion dependent on lower level items (you could not remove the lower levels without thereby removing the higher levels), what exists at a higher level cannot in general be *reduced to* what exists at a lower level. Higher-level phenomena are in this regard taken to be *autonomous* with respect to phenomena at lower levels. (206)

This picture is also the starting point for the investigation here. For short, I will sometimes speak of the question of whether the *prima facie* support for metaphysical emergence can be sustained as the question of whether the 'appearances' of metaphysical emergence are genuine, where the notion of 'appearance' is one tracking certain seemings as opposed to anything strictly perceptual.

1.1.2 Two key questions

Given that the *prima facie* motivations for metaphysical emergence span such a wide range of phenomena, the interest in exploring and illuminating this notion is clear. Toward this end, two questions (and related sub-questions) are key.

The first pertains to the nature and varieties of metaphysical emergence. Here we ask: just what is metaphysical emergence, more precisely? How is it, exactly, that higher-level entities (and features) can synchronically depend on complex

configurations of lower-level entities (and features), while retaining some degree of ontological and causal autonomy? And is there more than one way in which this can be—is there more than one form of metaphysical emergence?

The second pertains to whether there really is any metaphysical emergence, either possibly or actually. To start: are there in-principle problems with taking the appearances of dependence with autonomy to be genuine, such that emergence is better understood as an epistemic or representational phenomenon? Supposing one or more varieties of emergence is metaphysically viable, are there any actual cases of such emergence? If so, then we might be on track to vindicate and illuminate the existence both the special sciences and the entities they treat, as well as much of our ordinary experience. But if not, that would still be worth knowing. Either way, the results of these investigations would have wide-spread ramifications for our understanding of the world around us.

There is good reason, then, to pursue answers to the key questions of what, more specifically, metaphysical emergence is—here and throughout, of the sort motivated above—and whether there is any. Indeed, in past decades there has been an explosion of interest in such emergence, in both philosophical and scientific contexts. Unfortunately, all this attention has left the answers to the key questions less clear than ever.

What, more specifically, is metaphysical emergence?

One source of unclarity stems from the huge diversity of accounts offered as providing comparatively specific answers to the question, ‘What is emergence?’ There are, to be sure, several points of agreement, as follows:

- Accounts of emergence typically agree that core to the notion is the combination of synchronic⁴ material dependence and some form of autonomy.

⁴Some take emergence to be diachronic, but for present purposes, this seeming distinction can be glossed. Mill (1843/1973) suggests that certain (“heteropathic”) effects emerge from temporally prior causes, but also suggests that the features having powers to produce such effects synchronically emerge from lower-level entities. O’Connor and Wong (2005) take emergence to be diachronic, on grounds that emergent features are caused by lower-level features (sometimes

These core components are sometimes explicitly flagged (see [Bedau 1997](#)), and sometimes are implicitly encoded in specific accounts of the dependence and autonomy at issue, as when [Kim \(2006\)](#) says “two [...] necessary components of any concept of emergence that is true to its historical origins [...] are supervenience and irreducibility” (548).

- Accounts of emergence typically agree in taking emergence to contrast with Cartesian or other forms of substance dualism or pluralism, and more generally, with any view (such as “vitalism”) according to which material configurations cause or are otherwise associated with wholly distinct non-material substances. As [Stephan \(2002\)](#) puts it:

The first feature of contemporary theories of emergence, the thesis of physical monism, is a thesis about the nature of systems that have emergent properties (or structures). The thesis says that the bearers of emergent properties are made up of material parts only. It denies that there are any supernatural components responsible for a systems having emergent properties. Thus, all substance-dualistic positions are rejected [...]. (79)

Relatedly, [Mill \(1843/1973\)](#), the father of British Emergentism, supposes that emergent entities are entirely materially composed:

All organised [living] bodies are composed of parts similar to those composing inorganic nature, and which have even themselves existed in an inorganic state; but the phenomena of life

in combination with other emergent features), and causation is diachronic; but here again talk of diachronic causation appears to presuppose the synchronic emergence of features having the powers to produce the effects in question, and in any case (as I'll argue down the line) the essentials of a causal account of dependence are preserved whether or not the causation at issue is diachronic. [Rueger \(2001\)](#) takes emergence to be diachronic since involving temporally extended processes; but the emergence of such processes is compatible with these “synchronously” depending on a temporally extended base (compare spatiotemporally global supervenience). Similarly for Mitchell's ([2012](#)) conception of emergence as involving ‘diachronic processes’. [Humphreys \(1997\)](#) characterizes a form of irreducibly diachronic emergence, involving the exhaustive (non-mereological) “fusion” of lower-level entities into another lower-level entity; but such same-level emergence is besides the point of accommodating higher-level entities, and so will be set aside here.

which result from the juxtaposition of those parts in a certain manner bear no analogy to any of the effects which would be produced by the action of the component substances considered as mere physical agents. (243)

As per these remarks, it is common to build into accounts of emergence a commitment to substance monism, and moreover, materialism, according to which the only substance is material or physical substance, in line with the usual scientific consensus. Although in what follows I will also suppose that the dependence base entities are ultimately physical (see the Preliminaries section to follow), I do not explicitly take material or physical substance monism to be constitutive of the notion of metaphysical emergence, primarily because it seems reasonable and desirable to suppose that our understanding of such emergence be portable to other naturalistically acceptable conceptions of the base entities (the schemas for metaphysical emergence that I will later propose are so portable). Stephan's remark also reflects a related common assumption of accounts of emergence, to be next discussed in the text.

- Accounts of emergence typically agree that the emergence of entities (including objects, systems, events, processes, and other particulars) can be investigated by attention to the emergence of features (including properties, relations, behaviours, and states) of the entities at issue. On this understanding, any emergence there might be involves an emergent feature. As **Bedau** (2002) puts it:

[A]n entity with an emergent property is an emergent entity and an emergent phenomenon involves an emergent entity possessing an emergent property—and they all can be traced back to the notion of an emergent property. (6)

As I'll discuss further in the 'Preliminaries' section, I too will focus primarily on the emergence of features as offering a comparatively straightforward way of investigating the emergence of entities.

- A final typical point of agreement is that emergence has certain **correlational connotations**, sometimes expressed by saying that emergent features ‘supervene’ on base features with at least ‘nomological’—i.e., natural law-based—necessity (as I’ll sometimes put it: emergent features ‘minimally nomologically supervene’ on base features).⁵ Here the idea is that in every world (actual or hypothetical) with the same or relevantly similar laws of nature, the occurrence of an emergent feature *S* requires the occurrence of some or other lower-level base feature *P*, and in every such world, the occurrence of any such *P* will be accompanied by the occurrence of such an *S*. For example, Broad (1925) maintains that emergent features of a compound are “completely determined” by features of its parts when appropriately configured, in that “whenever you have a whole composed of these [...] elements in certain proportions and relations you have something with the [compound’s] characteristic properties” (64).

As we’ll see, some accounts of emergence take the correlations to hold with just nomological necessity, whereas others take them to moreover hold with metaphysical necessity (that is, in every possible world—not just worlds with relevantly similar laws of nature). Either way, these accounts agree that emergent features minimally nomologically supervene on base features.

Beyond these core points of agreement, however, accounts of emergence diverge into a bewildering variety, primarily reflecting that the core notions of dependence and autonomy have multiple, often incompatible interpretations. The extent of this diversity has led some to claim that references to emergence “seem to have no settled meaning” (Byrne 1994, 206), that accounts of emergence are “not obviously reconcilable with one another” (O’Connor 1994, 91), that “those discussing emergence, even face to face, more often than not talk past each other” (Kim 2006, 548), that “‘emergent’ and all its semantic kin have come to stand for a hopeless jumble of different ideas” (Ladyman and Ross 2007, 193), and that “[w]ithin philosophy and the sciences the term ‘emergence’ is used in such a be-

⁵See Kim 1990 and McLaughlin and Bennett 2018 for discussion.

wildering variety of ways that it seems the word itself is the only thing shared across these various usages” ([Silberstein 2009](#), 254).

One might wonder if the diversity here primarily reflects that some accounts of emergence take this to be a merely epistemic or representational phenomenon. On these other approaches (which we will also have opportunity to discuss in what follows), the seeming autonomy of emergent phenomena is cashed not in metaphysical terms, but rather in terms of such phenomena being, e.g., unpredictable or underivable from lower-level theories, and where the failures of predictability or derivability aren’t taken to have any clear metaphysical consequences for whether there are distinct and distinctively efficacious higher-level entities. But even restricting ourselves to accounts of emergence that are intended to have such metaphysical consequences, there remains a bewildering variety of options.⁶

Candidate conceptions of the synchronic dependence at issue in metaphysical emergence include:

- mereological ('part-whole') determination⁷
- causation or nomological connection⁸
- functional realization⁹
- constitutive mechanism¹⁰
- the determinable-determinate relation¹¹
- inheritance of causal powers¹²

⁶Both the lists and citations to follow are representative rather than exhaustive; there are many hundreds of papers and books on these notions and their variations, as entering into accounts of either physically unacceptable emergence (sometimes called ‘strong’ or ‘robust’ emergence) or physically acceptable emergence (a.k.a. ‘realization’ or ‘weak emergence’).

⁷See [Stephan 2002](#), [Gillett 2002a](#).

⁸See [Searle 1992](#), [O’Connor and Wong 2005](#).

⁹See [Putnam 1967](#), [Boyd 1980](#), [Poland 1994](#), [Antony and Levine 1997](#), [Melnyk 2003](#).

¹⁰See [Craver 2001](#), [Haug 2010](#).

¹¹See [MacDonald and MacDonald 1986](#), [Yablo 1992](#), [Ehring 1996](#), [Wilson 2009](#).

¹²See [Kim 1992a](#), [Wilson 1999](#), [Shoemaker 2000/2001](#).

- primitive ‘Grounding’¹³

Candidate conceptions of the ontological and/or causal autonomy at issue are even more various. Explicitly metaphysical (that is, explicitly non-epistemic and non-representational) conceptions include:

- nomological but not metaphysical supervenience¹⁴
- non-fundamental novelty (of properties, powers, laws, entities)¹⁵
- fundamental novelty (of properties, powers, forces/interactions, laws, entities)¹⁶
- non-additivity/non-linearity¹⁷
- “downward” causal efficacy¹⁸
- multiple realizability or compositional plasticity¹⁹
- symmetry breaking²⁰
- elimination in degrees of freedom²¹
- the holding of a proper subset relation between token powers²²

And “epistemic criteria” accounts of ontological and/or causal autonomy include:

¹³See Leuenberger in progress.

¹⁴See van Cleve 1990, Chalmers 1999, Noordhof 2010.

¹⁵See Anderson 1972, Humphreys 1996, Wimsatt 1996, Crane 2001, Pereboom 2002, Megill 2013.

¹⁶See the British Emergentists (e.g., Mill 1843/1973, Alexander 1920, Broad 1925), Kim 1992a, Cunningham 2001, O’Connor 2002, Wilson 2002a, Barnes 2012.

¹⁷See the British Emergentists, Newman 1996, Bedau 1997, Silberstein and McGeever 1999.

¹⁸See Sperry 1986, Klee 1984, Thompson and Varela 2001, Searle 1992, Schroder 1998, Stephan 2002.

¹⁹See Putnam 1967, Fodor 1974, Boyd 1980, Klee 1984, Wimsatt 1996, Aizawa and Gillett 2009.

²⁰See Morrison 2012.

²¹See Wilson 2010b.

²²See Wilson 1999, Shoemaker 2000/2001, Clapp 2001.

- in-principle failure of deducibility, predictability, or explicability²³
- predictability, but only by simulation²⁴
- lack of conceptual or representational entailment²⁵
- theoretical/mathematical singularities²⁶

Given this plethora of options, it's no surprise that many discussions of specifically metaphysical emergence aim primarily to taxonomize its varieties.²⁷

Now, though in general a thousand flowers may fruitfully bloom, this much diversity is unhelpful as regards answering the first key question, concerning the nature and varieties of specifically metaphysical emergence. It would be one thing if different accounts of metaphysical emergence targeted different phenomena. But different accounts often target the same phenomena, while disagreeing about whether these are metaphysically emergent; and when accounts do agree about a given case, there is often no clear basis for the agreement.

In particular, and importantly for the relevance of metaphysical emergence to contemporary debate, different accounts also often disagree over whether such emergence is compatible with *physicalism*, the view that all broadly scientific goings-on are, to speak schematically, “nothing over and above”, “realized in”, completely metaphysically dependent on”, or “grounded in” physical goings-on. For example, [Kim \(1999\)](#) takes physical realization (an especially intimate form of dependence) to be incompatible with metaphysical emergence, while [Gillett \(2002a\)](#) takes physical realization to be a form of metaphysical emergence. More generally, the extent of variability in both content and application here might well lead one to suppose that accounts of specifically metaphysical emergence, like accounts of emergence generally, have nothing systematic in common.

²³See [Broad 1925](#), [Hempel and Oppenheim 1948](#), [Klee 1984](#), [LePore and Loewer 1989](#).

²⁴See [Newman 1996](#), [Bedau 1997](#).

²⁵See [Chalmers 1996](#), [Van Gulick 2001](#).

²⁶See [Batterman 2002](#).

²⁷See [Klee 1984](#), [Van Gulick 2001](#), [Stephan 2002](#), [Gillett 2002b](#), [O'Connor and Wong 2015](#).

Is there (really) any metaphysical emergence?

The answer to the second key question, of whether there is any emergence of a substantively metaphysical variety, has also remained unclear, reflecting still-live concerns about whether the appearances of such emergence (or of an associated hierarchy of levels of natural reality) are genuine. Among these concerns are

- that the very notion of metaphysical emergence, combining dependence with ontological and causal autonomy, is incoherent;
- that the notion, while coherent, is either trivially fulfilled or trivially never fulfilled;
- that metaphysical emergence is naturalistically unacceptable;
- that considerations of parsimony push against taking the appearances of distinctness ontologically seriously;
- that metaphysically emergent entities or features, were they to exist, would give rise to problematic causal overdetermination of lower-level effects.

Here the diversity of accounts of emergence again muddies the waters; for while some accounts have resources to respond to some of these concerns, the absence of any systematic treatment of the notion of metaphysical emergence renders it unclear whether the notion can survive all the various attacks. And to the extent that the in-principle viability of metaphysical emergence remains unclear, the further project of determining whether there actually is any such emergence cannot even get off the ground.

My aim: providing clear, compelling answers to the key questions

The point and purpose of this book is to provide clear, compelling, and systematic answers to the two key questions of what, more precisely, metaphysical emergence is, and whether there really is any such emergence.

In response to the first key question, I will argue that for the sort of target cases motivating the notion of metaphysical emergence, there are two and only two schemas for metaphysical emergence, one of which is compatible with physicalism (on the assumption that the base-level goings-on are physical), and the other of which is not. And I will show that a representative range of existing accounts of metaphysical emergence plausibly aim to instantiate one or other schema, such that much of the apparent diversity of these accounts is superficial.

In response to the second key question, I will first argue that each of these two forms of metaphysical emergence is viable. More specifically, I will argue that each form of emergence is coherent, substantive, naturalistically acceptable, such as to avoid both causal exclusion and causal overdetermination, and more generally such as to vindicate and illuminate the *prima facie* scientific and other motivations for thinking that there is metaphysical emergence—again, understood as involving synchronically materially dependent yet distinct and distinctively efficacious higher-level entities and features. I will go on to consider, for a variety of interesting actual phenomena, whether these phenomena can be seen as metaphysically emergent in one or the other of these two ways; I will argue that one form of metaphysical emergence (the sort compatible with physicalism) is plausibly quite common, and the other (incompatible with physicalism) remains, for some cases, a live and in-principle empirically verifiable possibility, and for one special case is plausibly actually instantiated.

1.1.3 Outline of the book

The plan for carrying out this project is as follows.²⁸

In the remainder of Chapter 1 (§1.2: ‘Preliminaries’), I discuss certain basic presuppositions or prerequisites of the project. Here I lay out certain target cases of synchronic material dependence to which the concept of emergence might be

²⁸This manuscript is a draft, and though it is fairly complete, it is not entirely complete; among the recent discussions that I have yet to appropriately engage with are those in Yates 2016 and Paoletti 2017. More generally, I welcome contact in re work that it would be appropriate to cite or engage with here (in particular, by authors of such work).

naturally applied, and motivate the associated presupposition that the emergence of entities (objects, systems, processes, and so on) can be explored by focusing on the emergence of their features (properties, relations, behaviours, states, and so on) of these entities; here the notion of synchronic material dependence is extended to features of entities in the natural way.²⁹ I address the question of how to individuate ‘levels’, and present two approaches each of which is potentially useful in not ruling reductionist positions out of court; I flag the working assumption that the physical entities are compositionally basic, and present the operative conception of the ‘physical’; and I highlight the metaphysically neutral notion of ‘power’ that plays a role in the forthcoming schemas for emergence.

In Chapter 2 (‘The two schemas for metaphysical emergence’), I present what is seen by many as the most pressing challenge to taking the appearances of emergent structure as genuine—namely, the problem of higher-level causation, made salient by Jaegwon Kim in his 1989, 1993a, 1998, and elsewhere. The general concern here is that, given that higher-level entities and features synchronically materially depend on lower-level physical entities and features, such that the former minimally nomologically supervene on the latter, any purported effects of the former are reasonably taken to be already produced by higher-level entities or features are problematically causally overdetermined—that is, implausibly caused twice over. I argue, following discussions in Wilson 1999, 2001, 2011, and elsewhere, that there are two and only two strategies of response to this problem that make sense of higher-level entities and features’ being metaphysically emergent—that is, as being synchronically materially dependent on yet also ontologically and causally autonomous from lower-level physically acceptable base entities and features. One of these strategies provides a schematic basis for ‘Weak’ (physically acceptable) emergence; the other provides a schematic basis for ‘Strong’ (physically unacceptable) emergence. And for each of these strategies and associated schemas, I show that a representative range of seemingly diverse accounts of emergence are plausibly seen as aiming to satisfy the conditions in one or the

²⁹To wit: what it is for a special science feature S to synchronically materially depend on a lower-level physical feature P is for the entity bearing S to synchronically materially depend on the entity bearing P .

other schema, and thus are more unified than they appear. I go on to discuss the overdetermination argument put forth by Merricks (2003) against the existence of certain composed higher-level objects; I highlight certain differences between Merricks's and Kim's arguments, and argue that, *mutatis mutandis*, the Weak and Strong emergentist strategies also block Merricks's eliminativist conclusion.

I conclude that we have *prima facie* reason to think that satisfaction of the conditions in the schemas for *Weak* and *Strong* emergence is, as I put it, “core and crucial” to metaphysical emergence of both physically acceptable and physically unacceptable varieties, respectively. I prefer this terminology to the usual though to my mind overly coarse-grained terms of necessary and sufficient conditions, since any schematic account needs to be sensibly filled in. But modulo this caveat, the results of this chapter can also be seen as providing *prima facie* reason to think that the conditions in the schemas are both necessary and sufficient for metaphysical emergence of both physically acceptable and physically unacceptable varieties—a bold claim, but one that, as I argue in ensuing chapters, is surprisingly robust.

Since the schemas play a large and structuring role in what follows, it is worth prefiguring their content and the associated strategies for avoiding problematic overdetermination. A Strongly emergent feature has, on a given occasion, at least one token power not had by the base system feature upon which it synchronically materially depends on that occasion; overdetermination is avoided by denying that the base goings-on produce the effect (or, more weakly, produce the effect in the same way as the higher-level goings-on). Strong emergence is of the anti-mechanistic or anti-physicalist variety most commonly associated with British Emergentism, according to which, at certain levels of compositional complexity, fundamentally novel features and associated powers or laws come to exist (be instantiated, obtain). By way of contrast, a Weakly emergent feature has, on a given occasion, a proper subset of the token powers had by the base system feature upon which it synchronically materially depends on that occasion; problematic overdetermination is avoided via the token-identity of each power of the higher-level feature with a power of its dependence base, while the distinctive efficacy of the

higher-level feature is preserved (for reasons to be discussed in detail down the line) as a result of its having a distinctive power profile. Weak emergence is the sort associated with the many varieties of non-reductive physicalism, according to which some higher-level features are, while completely metaphysically dependent on certain complex configurations of lower-level, ultimately physical goings-on, nonetheless both ontologically irreducible to and distinctively efficacious as compared to the latter.³⁰

Two other points regarding the two schemas for emergence are worth registering, given their importance for what follows. First, for purposes of appreciating

³⁰The ‘proper subset of powers’ strategy for accommodating physically acceptable emergence (a.k.a. ‘non-reductive realization’) is sometimes inaccurately called ‘Shoemaker’s strategy’ or ‘Shoemaker’s account’ of realization, following Shoemaker 2000/2001—inaccurately, since my 1999 paper (first written for a Spring 1998 seminar with Richard Boyd on naturalism during my third year of graduate school at Cornell, and submitted to a *Philosophical Quarterly* competition that year) was the first published paper presenting and defending the proper subset strategy. There I motivated the strategy as required to block the Strong emergence of higher-level features from lower-level physical features; I moreover argued that apparently diverse accounts of non-reductive physicalism were more similar than they appeared, in having in common that the preferred realization relations each arguably satisfied the subset condition on powers. More generally, my paper directed attention to powers as suitably metaphysical means, going beyond appeals either to supervenience or to explanation, of distinguishing reductive from non-reductive versions of physicalism, and non-physicalist accounts from any form of physicalism. The powers-based approach I endorse has certain advantages over Shoemaker’s—importantly, as I’ll rehearse down the line, it is not required to implement the strategy that one accept Shoemaker’s (1980) view of properties as essentially and exhaustively characterized by their powers.

The pedigree of the proper subset strategy traces to John Heil’s 1996 NEH summer seminar in the metaphysics of mind, which took place at Cornell following my first year of graduate school, and which Heil graciously allowed me to attend. During the seminar, Michael Watkins struck upon the idea of “solving” the problem of mental causation—avoiding the threat that real and efficacious mental features would systematically overdetermine effects of their physical realizers—by taking the powers of a mental feature to be a proper subset of those of its physical realizer(s). The original idea for the proper subset strategy is thus Watkins’s; however, he did not go on to much develop the approach, whereas both I and Shoemaker (chair of my dissertation) did so, in parallel. Unfortunately, though I cited Shoemaker’s then work-in-progress, he did not and has never cited any of my work on this topic, which has, perhaps predictably, led to its being commonly assumed that he is the sole originator of the view. It didn’t help that the title of my 1999 paper (‘How Superduper does a Physicalist Supervenience Need to Be?’) was less than informative about the key results therein. Be all this as it may, I hope that those informed about this citation and priority issue will do what they can to ensure that my contribution to the original and subsequent development of the proper subset approach to realization is appropriately tracked.

the generality of the schemas, it is crucial to register that the notion of ‘power’ here is metaphysically highly neutral, reflecting commitment just to the plausible thesis that the causes an entity may potentially bring about are associated (perhaps only contingently) with how the entity is—that is, with its features. No controversial theses pertaining to the nature of powers, properties, causation, or laws are presupposed. As I later discuss, even a categoricalist contingentist Humean—that is, someone who rejects the notion of irreducible dispositions or powers, and who thinks that what causes what is ultimately a matter of contingent regularities—could accept powers in the weak sense at issue in the schemas.³¹

Second, others have observed that accounts of emergence may be broadly sorted into ‘weak’ and ‘strong’ varieties, that are and are not compatible with physicalism, respectively.³² My treatment goes beyond these (typically gestural) treatments in explicitly cashing the distinction between Weak and Strong emergence in metaphysical rather than epistemological or semantic terms, in more specifically identifying the differing schematic metaphysical bases for these two types of emergence, and in explicitly locating the schemas in a representative spectrum of existing accounts of emergence. My treatment also goes beyond previous taxonomically disunified descriptions of the varieties of emergence, in that the schemas for Weak and Strong emergence exhaust the available ways in which higher-level, broadly scientific entities might metaphysically emerge from lower-level such entities.

In Chapter 3 (‘The viability of Weak emergence’) I consider and respond to a number of objections to the schema for Weak emergence and the associated proper subset approach to realization presented and developed in Wilson 1999,

³¹What about the presupposition in the schemas that there are features (i.e., properties), which certain nominalists reject? Nominalists have their preferred strategies for converting talk of properties into talk of objects, which they are invited to apply here.

³²See Smart 1981, Bedau 1997, Chalmers 2006a, and Clayton 2006. These accounts typically take ‘weak’ emergence to be (merely) epistemological, and ‘strong’ emergence to be metaphysical; but this distinction doesn’t provide a useful basis for understanding the difference between reductive and non-reductive physicalism (even granting, what could be denied, that non-reductive physicalists are committed to epistemic gaps of some sort), or the difference between non-reductive physicalism (understood, as per usual, as a metaphysical thesis) and physically unacceptable emergentism.

2009, 2010*b*, and 2011, Shoemaker 2000/2001 and 2007, and Clapp 2001, among other venues. These objections include that satisfaction of the conditions in Weak emergence ...

- is compatible with anti-realism about higher-level features (as per Ney 2010; here also Heil 2003*b* is relevant)
- is compatible with reductionism (as per Kim 2010, Morris 2011*a*, and Dosanjh 2014);
- does not itself guarantee satisfaction of other conditions purportedly required for physically acceptable emergence/non-reductive realization (as per Melnyk 2006, McLaughlin 2007, and Gibb 2013);
- is compatible with epiphenomenalism (as per Walter 2010);
- is unmotivated, in that standard strategies for showing that the proper subset condition is satisfied (e.g.. appeals to multiple realizability) do not succeed (as per Morris 2013);
- fails to explain or accommodate qualitative distinctness of realizer and realized features (as per Gillett 2010);
- is not necessary for physically acceptable emergence (as per Davidson 1970, Macdonald and Macdonald 1995, Ehring 1996, Robb 1997, Pereboom 2002 and 2011, and Noordhof 2010).

These diverse challenges can, I argue, be answered. As we'll see, each challenge admits of one or more responses that are generally available to any sensible instantiation of the schema for Weak emergence. Upon occasion, additional (and to my mind, optimal) responses draw on features of my preferred accounts of Weak emergence—namely, one appealing to the determinable-determinate relation (as per my 2009 and 2009, and as originally proposed in MacDonald and MacDonald 1986 and Yablo 1992), according to which Weakly emergent features are determinables of lower-level realizers, and another appealing to an account of

Weak emergence as involving an elimination in degrees of freedom (as per my 2010*b*), according to which a Weakly emergent entity is associated with strictly fewer degrees of freedom (independent parameters needed to specify an entity's law-governed states and behaviours) than are associated with the system of its composing entities.

In Chapter 4 ('The viability of Strong emergence'), I consider and respond to a number of objections to Strong emergence. These objections include

- that Strong emergence is naturalistically unacceptable (as per Ladyman and Ross 2007 and Campbell and Bickhard 2011) or 'scientifically irrelevant' (as per Bedau 2002)
- that the having of a fundamentally novel power is compatible with physicalism (as per the interpretation of Alexander 1920 in O'Connor and Wong 2015)
- that depending on the operative account of what it is to have a power, the novel power condition on Strong emergence will never be satisfied (as per Kim 1999 and as discussed in Wilson 2002*a*); and relatedly, that there is no way to prevent dispositions to give rise to Strongly emergent features from being part of the fundamental physical base (as per Howell 2009 and Taylor 2015).

Here again, I argue that the diverse challenges can be answered. And here again, for each challenge one or more strategies of response are available to any sensible instantiation of the conditions in the schema for Strong emergence. Upon occasion, however, additional (and to my mind, optimal) responses appeal to specific features of my preferred 'fundamental interaction-relative' account of Strong emergence, according to which a Strongly emergent entity (feature) has at least one power that is grounded, at least in part, in a novel (non-physical) fundamental interaction.

Having established the in-principle viability of both Weak and Strong conceptions of metaphysical emergence, I go on to put this result to work, in considering

whether complex systems, ordinary (inanimate) objects, consciousness (characteristic of persons), and free will (characteristic of agents), are plausibly seen as either Weakly or Strongly emergent.

In Chapter 5 ('The emergence of complex systems'), I first discuss how the historical assumption that non-linearity was a marker of fundamental novelty of the sort at issue in Strong emergence was undermined by the discovery of complex nonlinear systems which were clearly physically acceptable. I then argue (drawing on work in Wilson 2010b and Wilson 2013b) that contrary to what is often claimed (e.g., in Bedau 1997 and Batterman 1998), features of complex systems such as non-linearity, unpredictability, algorithmic incompressibility, and universalizability do not themselves provide a decisive basis for taking complex systems to be either Weakly or Strongly metaphysically emergent, for such features are—at least for all the authors of these accounts say—compatible with complex systems and their features being merely epistemologically emergent. I go on to argue that complex systems in the thermodynamic limit, and more generally, complex systems near critical points, have degrees of freedom that are eliminated relative to the systems of their composing lower-level entities, and so are appropriately seen as Weakly emergent, by lights of a degrees of freedom-based account. Here also I address a recent challenge to this account of the emergence of complex systems, due to Lamb (2015), and related to the account of emergence proposed in Morrison 2012, according to which complex systems involve not *fewer*, but *more*, degrees of freedom, associated with 'order parameters' that emerge near critical points.

In Chapter 6 ('The emergence of ordinary objects'), I consider whether ordinary (inanimate) objects, of either natural or artifactual varieties, are plausibly seen as either Strongly or Weakly metaphysically emergent. I argue that insofar as such objects, *qua* inanimate, are governed by Newtonian mechanics, we have reason to think these are at least Weakly emergent, by lights of a degrees of freedom-based account. While here again the Strong emergence of ordinary objects remains a live empirical (if not commonly endorsed) possibility, the best such case involves that of artifacts, the Strong emergence of which ultimately depends

on whether the states of consciousness that ultimately determine what powers are possessed by artifacts are themselves Strongly emergent, as will be explored in Chapter 7. I close Chapter 6 by observing that the results herein have consequences not just for the status of ordinary objects as existing and at least Weakly emergent, but also for the proper assessment of Amie Thomasson's metaontological view, as discussed in her (2010) and elsewhere, that investigations into the status of ordinary objects should proceed differently from investigations into the status of special science entities.

In Chapter 7 ('The Emergence of consciousness') I turn to considering whether consciousness of the sort that we and other creatures enjoy is plausibly seen as either Weakly or Strongly emergent. Existing arguments for the Strong emergence of consciousness rely, one way or another, upon the supposition that consciousness, or certain of its characteristic features, lies beyond the explanatory reach of any lower-level physical goings-on. Though, as will have been established by this point, the presence of even an insuperable explanatory gap is not a sufficient indicator of Strong emergence, the proponents of explanatory gap arguments take such gaps to be metaphysically significant, in reflecting not just broadly mathematical barriers to explanation such as non-linearity, but rather that certain features of consciousness—notably, subjective or qualitative aspects of conscious experience—depart so greatly from lower-level physical features that this divergence provides reasonable grounds for thinking that no physicalist account of consciousness of either reductive or non-reductive (i.e., Weakly emergent) varieties could possibly be correct. I consider the two most promising forms of explanatory gap argument: knowledge arguments of the sort advanced in Nagel 1974 and Jackson 1982 and 1986), and the conceivability argument advanced in Chalmers (1996), 1999, and elsewhere. As I will argue, each of these forms of explanatory gap argument is uncompelling, for reasons that have not been previously much explored. The upshot will be that while it remains a live empirical possibility that consciousness is Strongly emergent, at present we have no compelling philosophical or empirical motivation for taking this to actually be so. I go on to argue that, on the supposition that consciousness is not Strongly emergent, attention to the ir-

reducibly determinable nature of qualitative conscious states provides good reason to see certain such states as realized in determinable-based fashion by lower-level physical states—that is, to see such states as Weakly emergent.

In Chapter 8 ('The emergence of free will'), I consider whether free will of the sort that we appear to have and to exercise is plausibly seen as either Weakly or Strongly emergent. I start by drawing on Bernstein and Wilson 2016 to present a framework for connecting existing positions on free will—most importantly, compatibilism and libertarianism—with existing positions in the mental (more generally: higher-level) causation debates. In our paper, Bernstein and I argued that compatibilists and non-reductive physicalists were each implementing a similar 'proper subset' strategy for responding to their respective problematics of free will and of mental causation, respectively; here I extend this result to establish good reason to think that the compatibilist strategy entails satisfaction of the conditions on Weak emergence. I then argue that a range of Libertarian positions are appropriately seen as entailing satisfaction of the conditions on Strong emergence. I then argue that free will of the compatibilist/Weakly emergent variety is plausibly seen as widespread, then go on to present a novel argument for the conclusion that at least some instances of seemingly free choice are properly taken to be of the libertarian/Strongly emergent variety.

1.2 Preliminaries

Henceforth, unless otherwise specified, by 'emergence' I mean 'metaphysical emergence', and by 'reduction' or 'reducibility' I mean 'ontological or metaphysical reduction/reducibility'.

1.2.1 The target cases

As above, both science and ordinary experience motivate taking many seemingly higher-level entities to be metaphysically emergent from configurations of lower-level entities. In what follows, we will have occasion to consider certain represen-

tative sorts of case of entities which appear to instantiate the coupling of dependence with autonomy which is characteristic of metaphysical emergence. In the course of what follows, I will discuss a number of specific cases where emergence might be thought to be at issue, including (though not restricted to) the following:

- ~ ... a “classical” rigid body from a quantum-mechanical configuration
- ~ ... a spherical conductor from a field array of charged particles
- ~ ... a thermodynamic complex system (fluid or gas) from a dynamically interacting configuration of molecules
- ~ ... a cell from a highly organized and dynamically interacting configuration of cell structures
- ~ ... a tree from a highly organized and dynamically interacting configuration of plant cells and tissues
- ~ ... a conscious mind from a highly organized and dynamically interacting configuration of neural processes
- ~ ... a table from a highly organized and dynamically interacting configuration of molecules

Sometimes the emergence of the *type* of entity will be the focus, where a type of entity is something like a (natural) kind, that can have many instances or tokens; other times the emergence of one or other *token* of an entity of a certain type will be at issue. When the difference between type and token is important for the discussion, this will be made clear.

More important is that, though in the target cases our concern is, in the first instance, with the emergence of (types or tokens of) entities, accounts of emergence nearly uniformly focus on the emergence of *features*—that is, tokens or

types of properties (including relations and behaviours) or states—from lower-level features of lower-level configurations.³³ The focus on features reflects that, as previously, it is commonly assumed that, as [Bedau \(2002\)](#) put it, all emergent phenomena “can be traced back to the notion of an emergent property” (6), in the sense that if some entity (object, system, etc.) is emergent, this is because it has some emergent feature (e.g., the feature of having a charged surface, in the case of a spherical conductor, or the feature of being in a qualitative experiential state, in the case of humans or their minds), which can be the direct target of investigation. Relatedly, a focus on features is useful in that talk of entities can be naturally translated into talk of features—namely, the property or feature of being an entity of the type in question. Correspondingly, the bulk of discussion in what follows will be on the proper characterization and assessment of accounts of the metaphysical emergence of higher-level features from lower-level features, and where the operative understanding of synchronic material dependence is as follows: what it is for a special science feature S to synchronically materially depend on a lower-level physical feature P is for the entity bearing S to synchronically materially depend on the entity bearing P .³⁴

So, for example, to explore the status as emergent of certain of the previous cases, it serves to focus on the status as emergent of:

- ~ . . . the property/state (of a rigid body) of *being governed by the laws of classical mechanics* from the lower-level property/state (of a quantum-mechanical configuration) of *being composed of sub-atomic particles having certain positions, spins and momenta*
- ~ . . . the property/state (of a spherical conductor) of *having a charged surface*

³³What it comes to for a feature of a lower-level configuration to itself be lower-level will be discussed in detail shortly. Here and elsewhere, those with nominalist or other scruples are invited to translate talk of features (properties/states) into their preferred terms.

³⁴A third reason to focus on the emergence of features is that, and notwithstanding that the initial motivations for metaphysical emergence are naturally expressed as involving higher-level special scientific and artifactual entities, one might want to remain neutral on or make room for one feature of an entity to emerge from another feature of that same entity. We will revisit the question of whether emergence always brings emergent entities in its wake, and if so, whether this is in tension with the assumption of physical monism, in Ch. 4.

from the lower-level property/state (of a field array of charged particles) of *being composed of particles having certain positions and charges*

- ~ ... the property/state (of a liquid or gas near a critical point) of *undergoing a change of phase* from the lower-level property/state (of a configuration of molecules) of being composed of molecules having certain positions, momenta, and energies
- ~ ... the property/state (of a cell) of being capable of reproduction, from the lower-level property/state (of the complex configuration of cell components) of being such as to undergo certain chemical interactions and exchanges
- ~ ... the property/state (of a plant) of being phototropic, from the lower-level property/state (of the plant's cellular walls) of being such as to undergo certain cellular wall weakenings and cellular expansions

As regards human persons and their minds, it will be useful to consider a variety of distinctive features of mentality, including:

- ~ ... the mental property/state (of a person) of *believing that $2+2=4$* , on a certain lower-level neurophysiological property/state (of certain neurons standing in certain neuronal relations).
- ~ ... the mental property/state (of a person) of *seeing something red*, on a certain lower-level neurophysiological property/state (of certain neurons standing in certain neuronal relations).
- ~ ... the mental property/state (of a person) of *choosing to go for a walk*, on a certain lower-level neurophysiological property/state (of certain neurons standing in certain neuronal relations).

For systematicity, and reflecting the translation strategy noted above, it will also sometimes be convenient to translate talk of an entity of a given kind into talk of the property of being an entity of that kind.

Emergence, as applying to the above sorts of cases, is naturally treated as a one-one relation between a lower-level feature and a higher-level feature, consonant with contrasting claims of ontological property reduction, which also involve a one-one relation—namely, identity—between lower-level and (only apparently) higher-level features. Relatedly, it is natural to suppose that certain complex relational micro-configurations (e.g., configurations of atoms standing in certain spatiotemporal and atomic relations) exist, as having the lower-level features in question.³⁵

All this presupposes that we are in position to specify which configurations of lower-level entities, and associated features of such configurations, are also plausibly deemed lower-level (that is, are clearly not emergent in any interesting sense). More generally, our investigation requires that we have some plausible

³⁵An alternative approach (inspired by Gillett in his 2001 and elsewhere, though his discussion more specifically concerns a form of realization, potentially related just to Weak emergence) treats emergence as a many-one relation between many non-relational lower-level features (say, multiple instances of atomic charge) and a higher-level feature. As I see it, the ‘one-one’ and ‘many-one’ approaches to emergence target the same phenomena in ways that are different, but not substantively so: the many-one approach considers the nature of the dependence of a seemingly higher-level feature on multiple, comparatively non-relational lower-level features, understood as combining in various unproblematic ways, whereas the one-one approach considers the nature of the dependence of a seemingly higher-level feature on a lower-level feature (of a relational lower-level entity), understood as already combined in various unproblematic ways. See also related discussion in [Bennett 2017](#), 10.

That said, the one-one approach, which presupposes that relational lower-level entities and features exist, is dialectically preferable, for three reasons. First, this approach is typically assumed in the accounts of emergence that we will be discussing. Second, as mentioned in the text, it more naturally accommodates the usual understanding of ontological reduction (contrasting with any form of metaphysical emergence) as involving identity between the features or entities in question. To be sure, some (e.g., proponents of the view that composition is identity) suppose that there can be many-one identity claims; but such a view does not wear its intelligibility on its sleeve (taking one feature to be identical to many appears to violate Leibniz’s Law—though see [Cotnoir 2013](#) for an interesting defense of composition as identity against this and other objections). Third, on the most natural reading of the target cases, the seeming emergence is between a higher-level feature and a feature of some, ultimately fundamental physical, structural entity (plurality or configuration). Also worth noting that Gillett’s main objection to the standard approaches is that they often assume that the realized and realizer features are had by the same entity, which by lights of the many-one approach is incorrect. I agree that this assumption can and in many cases should be dropped—which it can, even if the one-one approach is taken; but see discussion of [Gillett 2010](#) in Ch. 3 for further discussion.

means of individuating levels—that is, of saying which entities and features may exist at a given level. I turn now to considering two strategies for individuating levels in a dialectically sensible way.

1.2.2 The individuation of levels

As previously discussed, the existence of the special sciences provides *prima facie* support for natural reality’s being hierarchically structured, with entities at higher ‘levels’ existentially depending on entities at lower ‘levels’, and distinct levels being associated with different characteristic entities, features, and laws. More generally, it is natural to think of emergence in the target cases, whether natural or artifactual, as going hand-in-hand with the suggestion that emergent entities and features are ‘higher-level’ *vis-á-vis* the ‘lower-level’ goings-on upon which they depend. Still, as Wimsatt (1994, 225) notes, “the notion of a compositional level of organization is left unanalyzed by virtually all extant analyses of inter-level reduction and emergence”, notwithstanding that “levels and other modes of organization cannot be taken for granted, but demand characterization and analysis”. How should we understand talk of levels in what follows, and most importantly, which entities and features should be taken to exist at a given level? The question is a delicate one in various dialectical respects.

To start, at the present stage of the dialectic—antecedent to having considered reductionist or otherwise deflationist (e.g., eliminativist) reasons for thinking that the appearance of multiple levels is incorrect—talk of levels should be taken to reflect the appearances, so as not to rule ‘one-level’ positions out of court.

Also important for dialectical purposes is that levels (or the one level, if reductionism or some form of deflationism turns out to be correct) be individuated so as to include any combinations or configurations of entities and features to which the reductionist or deflationist may reasonably appeal. Suppose, for purposes of illustration, that the fundamental physical entities and features are atoms and pairwise bonding relations between atoms. Then, beyond these characteristic entities and relations, we should allow as existing, at the atomic level, not just small numbers of atoms standing in atomic relations, but also (among other aggregate combi-

nations) large numbers of atoms standing in highly complex atomic (including spatiotemporal) relations, of the sort that might, if reductionism were correct, be identical with a rock, a plant, or a person, at least at any given time. Similar resources are needed to make sense of deflationist views, such as Heil's (to be discussed further in Ch. 3), which reject both reductionism and the independent posit of higher-level features.³⁶ If we are suitably generous to the reductionist or deflationist, substantive debate can proceed over whether some apparently higher-level entity or feature really is higher-level, or is rather identical to some lower-level entity or feature (or perhaps doesn't exist at all, even as so reduced).³⁷

On the other hand, at this stage of the dialectic, neither we do not want to be so generous to the reductionist (or other deflationist) as to rule the possibility of either Weak or Strong emergence out of court.

With these constraints in mind, the question to be answered is: as a dialectical starting point, which combinations of entities and associated features should be taken to exist at a given level L of broadly scientific reality, beyond the entities and

³⁶As Heil (2003a) puts it:

I am inclined to think that ‘this is a statue’ can be, and often is, literally true. What makes it true is a complex, dynamic, arrangement of particles. A statue’s boundaries are, at the particle level, fuzzy. The collection of particles that we might regard as making up the statue at a given time can gain and lose member particles over time. We cannot hope to paraphrase or replace talk of statues with talk of such collections. Even so, it seems clear that, with few exceptions, objects like statues that make up our everyday surroundings owe their existence to arrangements of more fundamental constituents. We deploy predicates like ‘is a statue’ to mark off salient features of the world. These features are grounded in properties and arrangements of the fundamental constituents. (217)

³⁷Being suitably generous means that we cannot rest with certain ways of individuating levels. For example, Wimsatt's understanding of “compositional levels of organization . . . as constituted by families of entities usually of comparable size and dynamical properties, which characteristically interact primary with one another” (226) rules reductionism out of court. Indeed, unless the compositionally basic level can contain complex aggregates of (perhaps vastly) different sizes, one will not be able to make sense of Wimsatt's reductionist-friendly claim that “Because any complex material objects can be described at a number of different levels of organization, identity relations must hold between descriptions of the same object at different levels” (227–8). That said, as we'll see shortly, Wimsatt does aim to make sense of such identity claims by reference to a number of modes of ‘aggregation’.

features typically taken, by lights of the associated science S , to be characteristic of L ? In what follows, I'll discuss two different approaches to answering this question, neither of which is perfect, but either of which suffices to get discussion off the ground.

The ‘ontologically lightweight’ combinations approach

One common approach to the individuation of levels proceeds by allowing that various ‘ontologically lightweight’ combinations of the characteristic entities and features treated by a given science S and placed at a level L are also appropriately placed at L . For example, Hellman and Thompson (1975) start by characterizing the compositionally basic level in terms of the entities and features taken to be characteristic of fundamental physics, as including any

... satisfying any predicate in a list of basic positive physical predicates of [fundamental physical theory]. Such a list might include, e.g., ‘is a neutrino’, ‘is an electromagnetic field’, ‘is a four-dimensional manifold’, ‘and are related by a force obeying the equations (Einstein’s, say) listed’, etc. (554)

(See also Melnyk 1997 and others.) Hellman and Thompson then expand beyond the entities and features picked out by the basic physical predicates to include at the basic physical level (to simplify somewhat) any and all mereological sums (i.e., ‘fusions’ of parts into wholes) of spatiotemporally located instances of basic physical entities and features.³⁸

A number of other modes of combination are usually allowed, applying to entities, features, or both, which are supposed by all parties not to result in any interesting form of emergence. These are typically operations which are uncontroversially aggregative using resources of the science at issue (e.g., iterations of

³⁸Classical mereology is a theory of parts and wholes on which composition (‘fusion’) is ontologically lightweight, in that any objects *qua* parts automatically form a whole. There is controversy over whether material composition (as when, e.g., some atoms compose a molecule, and so on) can or should be understood in terms of mereological composition; see Simons 1987, McDaniel 2001, Paul 2002, Koslicki 2008, Bennett 2015, and Wilson *in progressa* for discussion.

pair-wise relations, as in the atomic example above), or are broadly logical, as involving certain Boolean, classical mereological, or set-theoretic combinations of entities or features. Some mathematical modes of combination are also considered ontologically lightweight in this context; in particular, it is common to suppose that features corresponding to linear combinations of L -level features should also be placed at L . Hence at a level L associated with a given science S , the L -level entities and features (including properties, states, and relations) would typically be taken to include:

- any characteristic entities or features treated by S ;
- any relational entity (configuration or plurality) consisting in any number of L -level entities standing in L -level relations;
- any relational feature of the form “being composed of some L -level entities with L -level features standing in L -level relations”³⁹
- any entity (or feature) consisting in a set or plurality of L -level entities (or features), understood as (merely) jointly existing

³⁹ Features of this form are sometimes called “micro-based” or “micro-structural” features; see Kim (1998) and Shoemaker (2007), following Armstrong (1978). Kim (1998) offers on the reductionist’s behalf the feature of being a water molecule as a case in point: “it is the property of having two hydrogen atoms and one oxygen in such and such bonding relationship” (84). More generally, Kim (1998) characterizes a micro-based property as follows:

P is a micro-based property just in case P is the property of being completely decomposable into nonoverlapping proper parts, a_1, a_2, \dots, a_n , such that $P_1(a_1), P_2(a_2), \dots, P_n(a_n)$, and $R(a_1, a_2, \dots, a_n)$. (84)

Shoemaker’s related characterization is as follows:

[Micro-structural] properties [...] can be specified entirely in terms of the micro-manifest powers of the constituent micro-entities together with how these micro-entities are related i.e., in terms of what could be known about them prior to their entering into emergence engendering combinations. Such a property will be the property of being composed of particles with such and such micro-manifest causal powers and related in such and such a way. [...] If emergentism is false, manifest causal powers are the only ones the micro-entities have, and physical micro-structural properties are the only ones macro-objects have, and the other properties of macro-objects are realized in their physical micro-structural properties. (2007, 55)

- any entity (or feature) consisting in a disjunction of any L -level entities (or features)
- any entity (or feature) consisting in a conjunction of L -level entities (or features)
- any entity (or feature) consisting in a mereological fusion of L -level entities (or features)
- any feature consisting in a linear (scalar or vector) combination of L -level features⁴⁰

Note that the specification here allows for iterative closure under the operations; for example, an entity consisting in a disjunction of conjunctions of L -level relational entities would also be L -level. Though in principle there might be further ontologically lightweight operations, it is common to suppose, on this approach, that the closure of L -level entities under these operations is more or less exhaustive of the (individual and aggregative) entities at L . Hence it is that debate over the status of a given seemingly higher-level entity or feature frequently proceeds by considering whether the entity or feature can be reduced to one or other of these ontologically lightweight combinations of characteristic lower-level entities or features. For example, a non-reductive physicalist might argue that the multiple realizability of a given mental feature type indicates that it is not reducible to any type of lower-level physical or physically acceptable feature, while a reductive physicalist might respond by suggesting that the multiple realizability of the mental type can be accommodated, compatible with reduction, by taking the mental type to be identical to a disjunction of physical types.

⁴⁰Wimsatt (1994) more generally suggests that “the conditions required for a system property to be an aggregate of the properties of the parts of the system—conditions on the ‘composition function’ relating system and parts’ properties” are “associativity, commutativity, inter-substitutivity, linearity, and invariance under decomposition and reaggregation” (237). For simplicity in what follows, I’ll understand talk of linearity as subsuming these additional features; so far as I can tell, nothing turns on whether the features on Wimsatt’s list besides linearity strictly speaking are added to the mix.

Though common and sufficient for purposes of getting debate off the ground, the ‘ontologically lightweight’ approach faces the concern that it is overly restrictive, in failing to allow that non-linear features of L -level aggregates might be L -level. The supposition that failure of additivity is sufficient for emergence goes back to Mill and other British Emergentists, and reflects a natural understanding of emergent entities and characteristic features as being “more than the sum of their parts”, by way of contrast with paradigm cases of non-emergence, involving features (such as mass and shape) and causal capacities (e.g., to exert certain forces) which were reasonably seen as scalar or vector resultants. Relatedly, given the force-based, broadly Newtonian science at the time, Mill and other British Emergentists reasonably assumed that a failure of additivity (in the production of effects, in particular) was indicative of a new fundamental force being on the scene, associated with a new level of emergent entities and features. While reasonable, the supposition that non-linearity is sufficient for emergence of the Strong variety endorsed by British emergentists is arguably incorrect: 20th-century investigations into complex systems have revealed that non-linearity of features (including behaviours) of physically acceptable aggregates (e.g., gasses and other thermodynamic systems) is pervasive. It has also been claimed (correctly, in my view; see Wilson 2013b) that non-linearity is insufficient even for Weak emergence; as Wimsatt has claimed, “System properties needn’t be purely additive or aggregative functions [...] consistent with reductionism or mechanism” (237).⁴¹

I will later revisit the question of whether and how non-linearity bears on either Strong or Weak emergence—first, in the Chapter 4 discussion of Strong emergence, and second, in the Chapter 5 discussion of complex systems. For other purposes, however—most crucially, for purposes of considering how the problem of higher-level causation motivates the two schemas for emergence—nothing in particular hinges on the answer to this question. As such, no harm comes from taking levels to be individuated along lines of the ontologically lightweight approach, as again is commonly done.

⁴¹A similar concern is associated with the suggestion that micro-structural features exist at levels different from their components, for on a natural understanding of what makes a property (merely) micro-structural, the suggestion is a variant of the linearity suggestion.

The law-consequence approach

An alternative approach to the individuation of levels expands upon the appeal to scientific laws, understood as applying to entities at a distinctive level of natural reality. Laws so understood are, in the first instance, metaphysical; they are, or encode, the “rules” governing the entities and features at issue, though, consonant with a suitably fallibilist realism about theories, discussion of laws typically focuses on claims made in or by scientific theories as appropriate stand-ins for claims about the associated laws.

On the law-consequence approach, the suggestion is that the laws governing entities characteristic of a given level can also do the work of expanding the domain of entities and features at that level in such a way that the reductionist is reasonably accommodated. The laws of fundamental physics, for example, are capable of taking as input or initial conditions various complex aggregations of characteristic physical entities and features; hence the laws/theories themselves have resources to expand beyond the explicit commitments of the laws/theory treating some L -level entities and features, to admit at that level any entities and features whose (potential) existence is deemed a metaphysical consequence—not to be confused with either mere necessitation or representational entailment—of the L -level laws. This is the sort of account that, I believe, Lewis (1983a, 34) has in mind when he says that the physical theory at issue in physicalism is “something not too different from present-day physics, though presumably somewhat improved”.

A law-consequence approach to the individuation of levels has certain advantages over the ‘ontologically lightweight’ approach. For example, unlike the latter approach a law-consequence approach need not antecedently specify whether non-linear entities or features are or are not to be placed at a given level L ; whether or not this is so will follow from the laws governing the characteristic entities at that level. More generally, it can allow that entities and features which are causal consequences just of the L -level laws may also be placed at L . Another advantage of a law-consequence approach is that it needn’t allow that every ontologically lightweight combination of entities or features at L is available for potential reduc-

tionist or otherwise deflationary purposes—if, say, some complex aggregates of atoms could not exist, for some law-based reason. On a law-based approach, there is no concern about positing L -level aggregates that do not (even potentially) exist by lights of the L -level laws, since such aggregates would not be found among the consequences of these laws.

That said, one might be reasonably concerned that, while a law-consequence approach to levels clearly makes room for the possibility of Strong emergence (since fundamental higher-level powers or other features of reality will not be metaphysical consequences of lower-level laws), it does not clearly make in-principle room for the possibility of Weak emergence of the sort associated with non-reductive physicalism. For non-reductive physicalists grant that the higher-level entities and features that they take to be genuine, as well as the higher-level laws governing these entities and features, are *in some sense* metaphysical consequences of the fundamental physical laws, even if these higher-level goings-on are (as non-reductive physicalists suppose) different from any lower-level goings-on (and moreover, some non-reductive physicalists think, are at least sometimes epistemically beyond our ken).

As I will discuss in more detail in Ch. 3, however, there is a way of making sense of (the possibility of) Weak emergence on a law-consequence approach to the individuation of levels, based in the notion of a degree of freedom—that is, an independent parameter needed to characterize the law-governed properties and behaviour of a given entity or feature; hence a law-consequence approach, properly understood, does not rule the possibility of such emergence out of court. To prefigure: the entities and features which are metaphysical consequences of the laws at a level L which are appropriately placed at L will be those whose specification includes all the degrees of freedom needed for the laws at L to operate. For example, in order to operate, the laws at the fundamental physical level require quantum degrees of freedom, such as spin. If it were to be the case that the degrees of freedom needed to characterize some entity or feature which was a metaphysical consequence of the physical laws did not include quantum degrees of freedom, then any such entities or features would not be appropriately placed at

the fundamental physical level, since in the absence of the relevant quantum information the physical laws wouldn't be able to operate on such entities or features. The upshot is that a law-consequence approach to the individuation of levels does not entail that any and all metaphysical consequences of the laws at a given level L should be placed at that level; rather, only those whose degrees of freedom allow the laws at level L to operate.⁴²

A related idea is that the consequences of a given set of laws relevant to individuating levels are those that are in the state space of the laws. Though the notions of a degree of a freedom and of a state space are somewhat technical, the underlying idea in both cases is intuitive; namely, that laws require certain kinds of information in order to operate, and that among the entities and features that are consequences of level- L laws, only those that preserve the information needed for the level- L -laws to operate are appropriately placed at L .

1.2.3 The operative notion of the physical

The physical as compositionally basic

In theorizing about reduction and emergence it is typically supposed that there is a lowest level of compositionally basic or compositionally fundamental entities, which (as Bennett evocatively puts it in her 2017) ‘build’ the substance, so to speak, of all broadly natural entities. But what if there are no compositionally basic entities—what if the world is gunky, such that notwithstanding that there is reason to think that parts bring about wholes (as opposed to vice versa),⁴³ ev-

⁴²I earlier noted that a *prima facie* motivation there being metaphysical emergence adverted to certain broadly scientific reasons for taking special science entities to be more abstractly characterized than lower-level physical goings-on, including conceptions of the former as having fewer degrees of freedom than the latter. The present point is just that a law-consequence approach to the individuation of levels doesn't itself build in the truth of either reductionism or of emergentism. Whether there are in fact goings-on with eliminated degrees of freedom is a matter for empirical determination, to be further considered down the line.

⁴³On a Monist view of the sort endorsed in Schaffer 2010, the Cosmos is the only compositionally basic entity, and parts of the Cosmos are less fundamental decompositions of the whole. Discussions of reduction and emergence tend to presuppose a ‘bottom-up’ rather than a ‘top-down’ conception, at least so far as the compositional relations at issue in the natural sciences are con-

erything can be further decomposed (as discussed, e.g., in [Zimmerman 1995](#) and [Schaffer 2003](#))? Granting the possibility of gunky worlds, in any case there are at least two ways to retain a close cousin of the usual supposition (that there are compositionally basic entities) in gunky worlds, in a way that allows theorizing about emergence and other ‘inter-theoretic’ relations proceed. First is if there is a level of relatively compositionally basic or fundamental entities, which can be effectively treated as atomic, in the sense that the influence of any entities associated with parts (or parts of parts, etc.) of the relatively basic entities is either exhausted or inherited by the relatively basic entities (see [Montero 2006](#) for a sophisticated variation on this theme). Second is if the decomposition converges to a limit corresponding to an effectively compositionally basic level (see [Wilson 2016c](#)). In what follows, I will assume that there is a level of compositionally basic (or relatively basic) entities, but will keep track of any concerns that may arise from this supposition.

Again, in these contexts it is also typically supposed by all parties to the debate—reductive and non-reductive physicalists, as well as Strong emergentists—that the compositionally basic entities are (only) physical. Hence, as previously mentioned, these parties to the debate contrast their views with substance dualist or pluralist accounts, on which there are basic substances besides physical or material substance. I too will assume that the compositionally basic (or relatively basic) entities are (only) physical, though as previously mentioned the schemas for Weak and Strong emergence can more generally be used to characterize two different sorts of emergence, irrespective of the status as physical of the lower-level entities at issue.

Finally, in these contexts it is typically supposed that the compositionally basic physical entities are the characteristic entities treated by fundamental physics. Such a physics-based characterization of the physical incorporates the transition

cerned; and I will operate with this presupposition in what follows. That said, for the most part little turns on this issue, since the Monist can grant that, notwithstanding that the Cosmos *qua* whole is compositionally prior to any of its proper parts (including, e.g., atoms and molecules), atoms are compositionally prior to molecules. Modulo the background presupposition of the Cosmos as the fundamental, then, the debate can proceed as usual.

from a priori to a posteriori characterizations of the compositionally basic entities. As [Crane and Mellor \(1990\)](#) tell the story, materialists (and their opponents) characterized the compositionally basic entities in terms taken to be definitive of matter—being impenetrable, being conserved, being such as (only) to deterministically interact, and so on; materialism was then understood as the thesis that everything was ‘nothing over and above’ the material. But contemporary physics has shown that the compositionally basic entities have few, if any, of these characteristics. Hence the materialist position has evolved into physicalism, where the specification of the compositionally basic entities is to be determined a posteriori by physical science (more precisely, physics) alone.

The physical/non-physical distinction

A physics-based approach to the physical, though common, faces the concern, articulated in ‘Hempel’s Dilemma’ (acknowledging [Hempel 1979](#)), that neither current nor future physics will do so far as characterizing the domain of the physical is concerned—at least if this domain is supposed to serve as a substantive dialectical basis for exploring whether or not some form of physicalism is correct. Here the domain of the physical is understood as including any entities and features at the physical level.

The first horn is straightforward: if the physics at issue is current physics, then since current physics is to some extent inaccurate and incomplete, then so will be the associated domain. There are three versions of the second horn, to the effect that if the physics at issue is rather future (or ideal) physics, then any version of physicalism based in this conception will be either indeterminate in content ([Hempel 1979](#); [Hellman 1985](#) 1985), trivially true ([Chomsky 1968](#), [Crook and Gillett 2001](#)), or compatible with entities that are intuitively physically unacceptable—most pressingly, in being fundamentally mental, in either having or bestowing mentality, at odds with the intended conception of the physical ([Papineau 1993](#), [Loewer 2001](#)).

Hempel’s dilemma can be avoided, however (see [Wilson 2006a](#)).⁴⁴ We start

⁴⁴What follows is my preferred strategy; see [Stoljar 2001](#), [Dowell 2006](#), and [Ney 2008](#) for dis-

by whittling down the difficulties associated with the future physics horn. First, it is definitive of fundamental (or relatively fundamental) physics that it treats entities at a fairly low level of compositional complexity (the characteristic physical entities), or lower-level aggregates thereof; this feature will also characterize future physics, thus blocking at least some of the concern about indeterminacy (I'll discuss a remaining concern shortly). Relatedly, to the extent that physics aims to be a ‘complete’ science, this is at best in the extended sense of its aiming to provide (as physicalists believe it does) a basis for all broadly scientific goings-on, in either reductive or non-reductive fashion; whether this can be done is an open empirical question, however, so again there is no danger that a future-physics-based characterization of the physical will end up rendering physicalism trivially true.

The third concern—that the characteristic entities posited by future physics might end up being fundamentally mental—is to my mind the most pressing. If, for example, future physics were to posit fields or particles themselves possessing mental properties (as opposed to composing or constituting complex entities possessing such properties), then physicalism would be thereby falsified; for on any plausible historically grounded understanding, physicalism is incompatible with pan- or proto-psychism, views according to which mentality exists at relatively low levels of constitutional complexity.

This difficulty may be avoided, however, by noting that Crane and Mellor’s genealogy omits a crucial fact: that physicalists have not handed over all authority to physics to determine, *a posteriori*, what is physical. Reflecting the historical roots of physicalism in materialism, as foundationally committed to understanding mentality as nothing over and above complex material goings on, one feature has remained definitive of the term ‘physical’ (as this term enters into formulating physicalism, at any rate); namely, that the compositionally basic physical entities (presumably, those characteristically treated by physics) are not fundamentally mental: the basic physical entities do not individually either possess or bestow mentality. Hence a physics-based account of the physical should not be under-

cussion of alternative conceptions of the physical and associated approaches to resolving Hempel’s dilemma.

stood as the view that any and all entities treated by physics—current, future, or ideal—are physical. It should rather be understood as incorporating a ‘No fundamental mentality’ constraint, according to which an entity is physical only if it is not fundamentally mental.

This much serves, in the main, to address the concerns with a future physics-based conception of the physical. The problem remains, however—to return to the **issue of indeterminacy**—that if the entities posited by future physics are of too different a character than those posited by present physics, the present content and applicability of the account will be compromised; this observation lies at the heart of efforts to make sense of a present-physics account of the physical.⁴⁵ Though, as I argue in my (2006a), present-physics accounts are not ultimately sustainable, the apt concern about indeterminacy can, I believe, be accommodated by taking the physical entities to be those that are approximately accurately treated by present or future (in the limit of inquiry, ideal) physics. My preferred formulation, which conforms to the intentions of most physics-based characterizations of the physical while avoiding the most pressing objections to these accounts, is as follows:

The physics-based NFM account: An entity is physical iff:

- (i) it is treated, approximately accurately, by present or future (in the limit of inquiry, ideal) physics; and
- (ii) it is not fundamentally mental⁴⁶

Positing the physicality of non-fundamentally-mental entities treated by better versions of physics prevents physics’ present failures from immediately falsifying physicalism, while providing continuous content to the account of the physical through the needed revisions.

This account conforms to the sort of physics-based account of the physical typically supposed to be at issue in debates over reduction and emergence, though participants do not always make the relevant respects of similarity (most importantly, as not involving fundamental mentality) explicit; see Lewis’s (1983a, 34)

⁴⁵See especially Melnyk 1997.

⁴⁶See Hellman and Thompson 1975, Papineau 1993, Ravenscroft 1997, Papineau 2001, and Loewer 2001 for variations on this theme.

characterization of the physical theory at issue in physicalism is “something not too different from present-day physics, though presumably somewhat improved”. In what follows this account of the physical will be in some sense operative, though as we’ll see not much will turn on the specific details of this account. The main take-home point is that there is at least one physics-based account of the physical suitable for our dialectical purposes.

1.2.4 A metaphysically neutral understanding of powers

As prefigured, the schematic accounts of Weak and Strong emergence that I will propose and defend each impose (different) conditions on the powers possessed by emergent and dependence base features.

Here, talk of ‘powers’ is simply shorthand for talk of what causal contributions possession of a given feature makes (or can make, relative to the same laws of nature) to an entity’s bringing about an effect, when in certain circumstances. That features are associated with actual or potential causal contributions (‘powers’) reflects the uncontroversial fact that what entities do (can do, relative to the same laws of nature) depends on how they are (what features they have). So, for example, a magnet attracts nearby pins in virtue of being magnetic, not massy; a magnet falls to the ground when dropped in virtue of being massy, not magnetic. Moreover, a feature may contribute to diverse effects, given diverse circumstances of its occurrence (which circumstances may be internal or external to the entity possessing the feature). Anyone accepting that what effects an entity causes (can cause, relative to the same laws of nature) is in part a function of what features it has—effectively, all participants to the present debate—is in position to accept ‘powers’, in this shorthand, metaphysically neutral, and nomologically motivated sense.

Besides commitment to the platitude that what entities can do (cause), relative to the same laws of nature, depends on how they are (what features they have), only one metaphysical condition is required in order to make sense of the powers-based conditions to follow; namely, that one’s account of (actual or potential) causal contributions (powers) has resources sufficient to ground the identity

(or non-identity) of a token causal contribution associated with a token higher-level feature, with a token causal contribution associated with a token lower-level feature. Here again, effectively all participants to the debate can make sense of such identity (non-identity) claims as applied to token (actual or potential) causal contributions (token ‘powers’).⁴⁷

Of course, beyond the neutral characterization of powers, understood as tracking the nomologically determined causal contributions associated with a given feature, philosophers disagree. It is of the first importance, in order to appreciate the generality of the upcoming schemas for emergence, to see that no commitment to any controversial theses about powers (or associated notions such as property or law) will be required payment in what follows. Three key points of non-commitment, to be further discussed and defended in Chapter 2, are worth highlighting.

 First, nothing in what follows requires accepting that it is essential to features that they have the powers they actually have. Maybe powers are essential to features; maybe they aren’t. As we will shortly see, it suffices to characterize the Strong emergentist and non-reductive physicalist strategies, and associated schemas for emergence, that powers are contingently had by the features at issue.

Second, nothing in what follows requires accepting that features are exhaustively individuated by powers. Maybe they are; maybe they aren’t: perhaps features are also or ultimately individuated by quiddities or other non-causal aspects of features. In any case, the presence or absence of quiddities, which primarily serve to locate actually instanced features in worlds with different laws of nature, plays no role in actually individuating broadly scientific features in either scientific law or practice; and similarly for our ordinary practices of individuation of macro-entities and features. As such, the presence or absence of non-causal aspects of the features at issue can play no interesting role in a metaphysical account aiming to vindicate the scientific appearances supporting higher-level emergence; and nor does it, in the schemas to come.

⁴⁷See Wilson 2011 and (2015c) for substantiation of this point, according to which even a contingentist categoricalist Humean has resources sufficient unto implementing the powers-based schemas.

Third, nothing in what follows requires accepting that powers are or are not reducible to categorical features, or that attributions of powers are or are not reducible to certain conditionals or counterfactuals, etc. Maybe powers, or talk of them, are reducible to other entities or terms; maybe they aren't. Again, scientific theorizing and practice, and our ordinary practices of individuation, are transparent to such further metaphysical details, and so too should be—and are—our associated conceptions of emergence.

Is even a neutral reliance on powers problematic?

Before continuing, I want to consider the concern that even a metaphysically neutral reliance on powers is problematic, since, some claim, physics rejects causation, and so rejects powers. To be sure, I am committed to rejecting this view, as are physicalists in general, who as we will see typically accept *Physical Causal Closure*, according to which every lower-level physical effect has a purely lower-level physical cause. But the view is well-rejected.

To start, many of those claiming that physics dispenses with the notion of cause often cite [Russell \(1912\)](#) in this regard. As [Reutlinger \(2017\)](#) describes Russell's view:

Russell famously held that fundamental physics teaches us that—contrary to the beliefs of philosophers (of his time)—causal relations are not part of the ontology of fundamental physics. Call this the orthodox Russellian claim. (2291)

But notwithstanding Russell's influential rhetoric—e.g., the oft-cited claim that causation is “a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm” (1), his discussion in fact targeted only a single, implausible, notion of ‘cause’—one according to which causal relations are exceptionless universal generalizations. Perhaps Russell is right that physics—and no other science, for that matter—appeals to such a notion of cause, but that is a far cry from establishing that physics does not appeal to the notion of cause at all.

More recently, [Ladyman and Ross \(2007\)](#) say, “There is no justification for the neo-scholastic projection of causation all the way down to fundamental physics and metaphysics” (280). But on the face of it physics as well as the other sciences depend upon the notion of cause. The notion of a fundamental interaction, for example, is on the face of it explicitly causal; and special science laws are (as Ladyman and Ross grant) typically couched in explicitly causal terms.

Even so, it is sometimes claimed that these appearances cannot be correct, since, for example, [the laws of physics are time-reversible](#), unlike causation (which is typically supposed to go in only one direction—from past to future) [or are expressed in terms of probabilities which do not entail a specific causal evolution](#), or are expressed using mathematical rather than causal relations (such as the identity symbol in ‘ $F = ma$ ’ or ‘ $PV = nRT$ ’). But none of these considerations support the skeptical conclusion. To start, as [Field \(2003\)](#) notes, “it is not obvious that the claim that the basic laws of physics are time-symmetric is correct; indeed, the notion of the time symmetry of a law itself is not as clear as it sounds” (page), and indeed, even if these laws are in some sense time-symmetric, “there is still room in principle for arguing that their character in the forward direction is importantly different from their character in the backwards direction” (page). For example, while quantum mechanics is expressed in terms of the Schrödinger equation and associated probabilities, the causal notion of a measurement is also part of the theoretical mix; as [Field \(2003\)](#) notes, “certain kinds of indeterministic laws might be such as to give rise to a fundamental distinction in temporal direction: for instance, those with well-defined probabilities only in the forward direction. (If one takes quantum mechanical laws to include “the collapse of the wave packet”, then the law governing the collapse would appear to be an example).

Of course, such “collapse” interpretations of quantum mechanics are controversial; but the more general point is that there’s no in-principle reason to suppose that physical theory doesn’t or can’t appeal to the notion of causation. It is also worth noting that laws which appear to be time-reversible or expressed in probabilistic or mathematical terms may be understood as reflecting certain convenient modes of expression which are, properly interpreted, compatible with the states of

affairs so described being causal. To take a simple example, the fact that the relation between force, mass, and acceleration is typically expressed using the symbol for identity is compatible with these quantities being causally connected—as they evidently are. And that physicists use mathematical symbols to represent lawful connections may be seen as reflecting representational convenience or predictive necessity as opposed to any intended rejection of causation.

More generally, there is as yet no reason to think that conceptions of emergence cannot rely on an appeal to (appropriately neutrally understood) powers.

1.2.5 Some brief remarks on methodology

In theorizing about a given phenomenon, I aim to accommodate as much of the relevant data as possible, in as illuminating and sensible a way as possible. This desideratum, in metaphysics as in science, acts as an important constraint. After all, if a theory about X fails to accommodate a sufficient range of the data about X , why think that the theory is about X , as opposed to about something else entirely?

The data here, as above, primarily concern a range of considerations—again, drawn from science, perception, our practices of individuation, language, and introspective experience—that individually and together provide *prima facie* support for there being entities and associated features combining the characteristic features of synchronic material dependence and ontological and causal autonomy—that is, combining dependence with distinctness and distinctive efficacy—which combination of features is taken to be broadly definitive of metaphysical emergence of the sort motivated by the target cases. Other things being equal, I think it would be better, and I will aim, to make realistic good sense of the *prima facie* motivations for emergence, as generally consonant with the scientific and other practices that seem to presuppose that natural reality has a layered structure. Moreover, insofar as humans like ourselves are among the dependent entities whose seeming ontological and causal autonomy is at stake, there is to my mind special interest in preserving the appearances of higher-level autonomy—especially given the bearing on human action, intention, morality, creativity, and other topics and activities which seem to crucially constitute our understanding of ourselves and others.

This basic methodological standpoint in turn informs what I take to be the dialectical burden so far as responding to those who argue that the appearances of metaphysical emergence cannot be taken as genuine. In providing these responses, as I think I am able to do, my aim is not (at least not in the first instance) to argue that a reductionist or eliminativist position is untenable. My primary aim is not to knock opponents of metaphysical emergence off their horse, but rather to assist those endorsing metaphysical emergence—the common sense view, it seems to me—in staying on their own.

Chapter 2

Two schemas for emergence

In this chapter, I start by presenting what is seen by many as the most pressing challenge to taking the appearances of emergent structure as genuine—namely, the problem of higher-level causation, made salient by Jaegwon Kim in his 1989, 1993a, 1998, and elsewhere, according to which irreducible higher-level features would problematically causally overdetermine the effects of the lower-level dependence base features upon which they synchronically materially depend. I argue, following discussions in Wilson 1999, 2001, 2011, and elsewhere, that there are two and only two strategies of response to this problem that make sense of higher-level entities and features’ being metaphysically emergent—that is, as being synchronically materially dependent on yet also ontologically and causally autonomous from lower-level physically acceptable base entities and features. One of these strategies provides a schematic basis for ‘Weak’ (physically acceptable) emergence; the other provides a schematic basis for ‘Strong’ (physically unacceptable) emergence. And for each of these strategies, I show that a representative range of seemingly diverse accounts of emergence are plausibly seen as aiming to satisfy the conditions in one or the other schema, and thus are more unified than they appear. I go on to discuss the overdetermination argument put forth by Merricks (2003) against the existence of certain composed higher-level objects; I highlight certain differences between Merricks’ and Kim’s arguments, and argue that, mutatis mutandis, the Weak and Strong emergentist strategies also block

Merricks's eliminativist conclusion. I conclude that we have *prima facie* reason to think that satisfaction of the conditions in the schemas for *Weak* and *Strong* emergence is, as I put it, "core and crucial" to metaphysical emergence of both physically acceptable and physically unacceptable varieties, respectively. I prefer this terminology to the usual though to my mind overly coarse-grained terms of necessary and sufficient conditions, since any schematic account needs to be filled in and moreover filled in in sensible fashion, if it is to be really adequate. But modulo this caveat, the results of this chapter can also be seen as providing *prima facie* reason to think that the conditions in the schemas are both necessary and sufficient for metaphysical emergence of both physically acceptable and physically unacceptable varieties—a bold claim, but one that, as we will see in ensuing chapters, is surprisingly robust.

2.1 The problem of higher-level causation

I start with three clarificatory remarks. First, following Kim and common practice, I assume that entities (again, including objects, systems, events, and so on) are efficacious in virtue of having efficacious features (properties, relations, states, behaviours, and so on); for example, the effects that a billiard ball causes (can cause) are a matter of what properties it has—its mass, shape, and so on. As such, in what follows talk of entities' causing effects is suppressed in favor of talk of their features' causing effects. The assumption that the efficacy of entities lies in their having efficacious features is conveniently consonant with the operative assumption (discussed in Ch. 1, §2.1) that the emergence of entities is ultimately a matter of the emergence of features. Second, given that causation is in the first instance a relation between spatiotemporally located goings-on, reference to 'features' in what follows is to be understood, unless otherwise qualified, as reference to spatiotemporally located tokens (e.g., property instances, particular states, particular events) of a given type (property, state type, event type).¹ Third, to fix

¹That said, I will sometimes gloss the type/token distinction—e.g., when discussing necessitation of one feature by another, below.

ideas I set up the problem as directed at special science entities and features, but nothing deep hangs on the reference to science(s) here; the problem arises more generally for both natural and artifactual higher-level entities and features.

2.1.1 Kim's overdetermination argument

Six premises lead to the problem of higher-level causation.² Four of these concern special science features, and are motivated by considerations similar to those giving rise to there seeming to be metaphysical emergence. These are:

1. *Dependence*. Special science features synchronically materially depend on lower-level physically acceptable features (henceforth, “base features”) in such a way that, at a minimum, the occurrence of a given special science feature at a time or over a temporal interval (at least nomologically) requires and is (at least nomologically) necessitated by a physically acceptable base feature at that time or over that interval. In other words, special science features minimally nomologically supervene on base features.³
2. *Reality*. Both special science features and their physically acceptable base features are real.
3. *Efficacy*. Special science features are causally efficacious.
4. *Distinctness*. Special science features are distinct from their base features.

The remaining two premises concern causation. The fifth is a standard physicalist commitment, sometimes called ‘the causal closure of the physical’:

²What follows reflects my preferred way of presenting the problem and slate of candidate resolutions, as set out in Wilson 2009, 2011, and elsewhere. Kim’s own presentations more specifically target motivating reductive over non-reductive versions of physicalism.

³Recall: in worlds with laws of nature relevantly similar to the actual laws of nature, any given token of a special science type requires, for its occurrence, a token of some physically acceptable type; and in such worlds, if any token of that physically acceptable type occurs, then a token of that special science type will occur.

5. *Physical Causal Closure.* Every lower-level physically acceptable effect has a purely lower-level physically acceptable cause.⁴

The sixth reflects the common supposition that there is no systematic causal overdetermination (henceforth, just: ‘overdetermination’) of effects by distinct individually sufficient causes, with the exception of causes which form part of a single diachronic causal chain (which cases are not relevant to cases of synchronic material dependence), and “double-rock-throw”-type cases, where a given effect (e.g., a window’s breaking) is, on a given occasion, the result of two distinct causes (e.g., two rock-throwings), each of which is individually sufficient for an effect of the type at issue. Note that in neither of these ‘exception’ cases, does one of the competing causes stand in a relation of synchronic material dependence to the other.

6. *Non-overdetermination.* With the exception of double-rock-throw cases, effects are not causally overdetermined by distinct individually sufficient synchronic causes.

On to the problem. There are two cases to consider, reflecting two sorts of effects that might be at issue. In the first case, a special science feature S is assumed to cause another special science feature S^* ; in the second case, S is assumed to cause a lower-level physical feature P^* . In Kim’s classic presentation, S is taken to be a mental feature (e.g., a token state of being thirsty); P is taken to be a lower-level physically acceptable feature upon which mental state S depends, on a given occasion; and mental state S is taken to cause either another mental state S^* (e.g., a desire to quench one’s thirst) or a lower-level physically acceptable state P^*

⁴In being formulated in terms of lower-level physical/physically acceptable causes and effects, this characterization of *Closure* is similar to those in [Baker 1993](#), 79 (“Every instantiation of a micro-physical property that has a cause at t has a complete micro-physical cause at t ” and [Sturgeon 1998](#), 124 (“Every quantum event has a fully disclosive, purely quantum history”). *Closure* is sometimes expressed in more generous terms (e.g., every physical/physically acceptable effect has a physical/physically acceptable cause), but such a formulation is overly broad, not least because overdetermination and exclusion concerns also arise for higher-level physically acceptable features. See [Montero 2003](#) and [Garcia 2014](#) for further discussion.

(e.g., a physical reaching for a glass of water). More generally, however, the considerations to follow raise a concern about how any real and distinct higher-level feature might be unproblematically efficacious.

First (case 1), suppose that S causes special science feature S^* on a given occasion (compatible with *Efficacy*). S^* is synchronically materially dependent on some base feature P^* (*Dependence*), such that P^* necessitates S^* , with at least nomological necessity. Moreover, P^* has a purely lower-level physically acceptable cause (*Physical Causal Closure*)—plausibly, and without loss of generality, P . If P causes P^* , and P^* (at least nomologically) necessitates S^* , then it is plausible that P causes S^* , by causing P^* . So, it appears, both P and S cause S^* , and given that P and S are both real and distinct (*Reality*, *Distinctness*), S^* is causally overdetermined; moreover (given *Dependence*), this overdetermination is not of the double-rock-throw variety (contra *Non-overdetermination*).

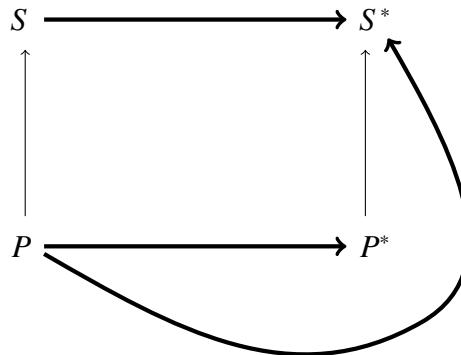


Figure 1: Case 1 of the problem of higher-level causation: S causes S^*

Second (case 2), suppose that S causes some base feature P^* on a given occasion (compatible with *Efficacy*). P^* has a purely lower-level physically acceptable cause (*Physical Causal Closure*)—plausibly, and without loss of generality, P . So, it appears, both P and S cause P^* , and given that P and S are both real and distinct (by *Reality* and *Distinctness*), P^* is causally overdetermined; moreover, (given *Dependence*) this overdetermination is not of the double-rock-throw variety (contra *Non-overdetermination*).

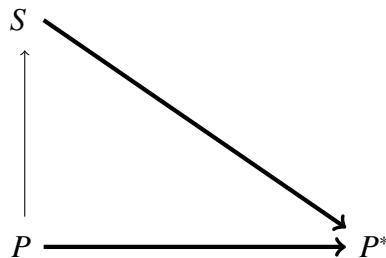


Figure 2: Case 2 of the problem of higher-level causation: S causes P^*

So goes Kim's argument that real, distinct and efficacious higher-level features induce problematic overdetermination.

Kim sees his argument as motivating rejection of the premise that special science features are distinct from their base features—that is, he sees it as motivating reductionism (more specifically: reductive physicalism). For present purposes, however (and following Wilson 2011 and elsewhere), it is useful to more generally note that rejection of each of the premises of the argument is associated with one or more fairly comprehensive positions in the metaphysics of science. Rejection of one or other of the first four premises gives rise to the following strategies of response, and associated positions:

1. *Substance dualism or Pan/proto-psychism.* Deny *Dependence*: avoid overdetermination by denying that S and S^* synchronically materially depend on physically acceptable P and P^* , respectively.⁵ If higher-level features do not so depend on lower-level features, there is no motivation for positing a lower-level physical property P as a dependence base for M , hence no motivation for positing a competing causal chain from P to M^* (case 1) or from P to P^* (case 2).
2. *Eliminativism.* Deny *Reality*: avoid overdetermination by denying that S and S^* are real.⁶ If higher-level features are not real, they do not compete with lower-level features for efficacy.

⁵See, e.g., Descartes 1641–7/1984 and Chalmers 1996.

⁶See, e.g., Churchland 1981, Churchland 1984, and Churchland 1986; see also Merricks 2003, to which I will shortly return.

3. *Epiphenomenalism.* Deny *Efficacy*: avoid overdetermination by denying that S is efficacious.⁷ If higher-level features are not efficacious, they do not compete with lower-level features for efficacy.

4. *Reductive physicalism.* Deny *Distinctness*: avoid overdetermination by denying that S is distinct from P .⁸ If higher-level and lower-level features are identical, they do not compete for efficacy.

Each of these strategies avoids overdetermination, but not in a way making sense of the seeming emergence of higher-level features. In the case at hand, for S to be emergent, it must synchronically materially depend on physically acceptable P while being both ontologically and causally autonomous from P —that is, while being distinct from and distinctively efficacious as compared to P . But the substance dualist and pan/proto-psychist strategies each deny that S materially synchronically materially depends on P : substance dualists deny this on grounds that S is instantiated in a non-physical substance; pan/proto-psychists deny this on grounds that the dependence base property is not physically acceptable, or at least is not physically acceptable in the usual sense according to which fundamental physical goings-on do not themselves have or exhibit mentality.⁹ The eliminativist and reductive physicalist strategies each deny that S is ontologically autonomous/distinct: the eliminativist denies this on grounds that S doesn't exist, and the reductive physicalist denies this on grounds that S is identical with P . Finally, the epiphenomenalist and reductive physicalist strategies each involve denying that S is causally autonomous/distinctively efficacious: the epiphenomenalist denies this on grounds that S isn't efficacious at all, and the reductive physicalist denies this on grounds that S and P are efficacious in just the same way, since they are identical.

⁷See, e.g., Hodgeson 1962 and Huxley 1874; see Robinson 2012 for contemporary literature.

⁸See, e.g., Lewis 1966, Smart 1958, and Kim 1993a.

⁹Recall that on the operative conception of the physical, the physical goings-on, whatever else they may exactly turn out to be by lights of our best physical theories, are not individually fundamentally mental.

2.1.2 The two ‘emergentist’ strategies for responding to the problem

The remaining strategies for responding to the problem of higher-level causation, and associated positions, do better by way of accommodating emergence. These are:

5. **Strong emergentism.** Deny *Physical Causal Closure*: avoid overdetermination by denying that every lower-level physically acceptable effect has a purely lower-level physically acceptable cause.
6. **Non-reductive physicalism.** Deny *Non-overdetermination*: allow that effects caused by S are overdetermined by P , but maintain that the overdetermination here is of an unproblematic non-double-rock-throw variety.

As I argue in the next two sections, these two strategies and associated positions are perspicuously seen as motivated by two conditions on the powers of a given special science feature, where satisfaction of one or other condition provides a *prima facie* plausible and principled basis for taking the feature to be emergent, in ways that standard proponents of the strategy/position would endorse. In each of these sections, treating Strong emergence and Weak emergence, respectively, I start by motivating the associated condition on powers by attention to standard versions of the position; I then show how satisfaction of the condition dovetails with the associated strategy for responding to the problem of higher-level causation; I then provide *prima facie* reasons for thinking that satisfaction of the condition provides a basis for taking higher-level special science features to be both synchronically materially dependent and ontologically and causally autonomous; finally, I use the condition to formulate the associated schema for metaphysical emergence.

Before getting started, three points of clarification are worth noting. First, as previously discussed, talk of ‘powers’ in what follows is simply shorthand for talk of what causal contributions possession of a given feature makes (or can make, relative to the same laws of nature) to an entity’s bringing about an effect, when

in certain circumstances. Anyone who accepts that the effects an entity causes (or can cause, relative to the same laws of nature) are in part a function of what features the entity has—effectively, all participants to the present debate—is in position to accept ‘powers’, in the shorthand, metaphysically neutral and nomologically motivated sense at issue here. Second, the qualifier ‘*prima facie*’ in the previous paragraph reflects that a full defense of the claim that the two conditions on powers serve as a basis for two viable conceptions of metaphysical emergence requires detailed treatment of the sort I’ll conduct in Chapters 3 and 4. Third, though it is an interesting question how to more specifically understand the forms of dependence at issue in the different schemas for emergence, entering into these details now would take us too far afield. Hence in formulating the schemas, the condition on dependence is expressed simply in terms of synchronic material dependence. Here again, more detailed discussion will be found in later chapters.

2.2 Strong emergentism and the *New Power Condition*

Strong emergentists maintain that some special science features are real, synchronically materially dependent, distinct, and distinctively efficacious as compared to their physically acceptable base features. The conception of higher-level efficacy at issue here is one that is intended to be incompatible with physicalism, and is characteristic of British Emergentism, associated with Mill, Alexander, Lewes, and Broad,¹⁰ as “the doctrine that there are fundamental powers to influence motion associated with types of structures of particles that compose certain chemical, biological, and psychological kinds” (McLaughlin 1992, 52), where the powers at issue are typically taken to be “powers to generate fundamental forces not generated by any pairs of elementary particles” (71).

A common concomitant of the Strong emergentist conception is the assumption that emergent features and powers arise in accord with fundamental ‘config-

¹⁰See, in particular, Mill 1843/1973, Alexander 1920, Lewes 1875, and Broad 1925. Evidently Lewes was the first to use the term “emergent” to characterize higher-level entities and features.

urational' or (as Broad put it) 'trans-ordinal' laws, connecting lower-level structures with higher-level entities and features, which laws are just as metaphysically and scientifically fundamental as the 'intra-ordinal' laws governing lower-level physical phenomena:

[T]he law connecting the properties of silver-chloride with those of silver and of chlorine and with the structure of the compound is, so far as we know, an unique and ultimate law.¹¹ (Broad 1925, 64-5)

As Lloyd Morgan (1923) writes:

I speak of events at any given level of the pyramid of emergent evolution as "involving" concurrent events at lower levels. Now what emerges at any given level affords an instance of what I speak of as a new kind of relatedness of which there are no instances at lower levels [...]. This we must accept "with natural piety" as Mr. Alexander puts it. [...] But when some new kind of relatedness is supervenient [...] the way in which the physical events which are involved run their course is different in virtue of its presence.

Here Lloyd Morgan's claim that higher-level events 'involve' concurrent lower-level events reflects the key assumption that emergent entities and features synchronically materially depend upon (complex configurations of) lower-level base entities and associated features; his claim that the 'new' emergents at a level do not admit of explanation reflects the assumption that Strongly emergent features are not just novel (e.g., in reflecting the occurrence of a novel aggregation of

¹¹It is worth registering that the "unique and ultimate law" at issue in such a case pertains not just to the bringing about of a given emergent feature, but also (as above) to various novel powers of the emergent feature, including those to affect lower-level, ultimately physical goings-on. Hence it would not be appropriate to characterize Broad's position as involving the posit of two distinct systems of laws: one 'emergent' law connecting lower-level configurations to emergent features understood as characterizable independent of any of the latter's effects, and another 'causal' law connecting emergent features to their potential effects. Broad and other British emergentists took the emergence of an entity or feature to be constituted by the emergence of fundamentally novel causal powers; it never occurred to them to think of scientific features as metaphysical posits (primitive non-causal quiddities?) that could be separated from their causal potentialities.

lower-level entities and associated physically acceptable features), but fundamentally so;¹² and his claim that emergent events are efficacious *vis-à-vis* physical events reflects the assumption that the fundamental novelty of an emergent entails its having at least one (fundamentally) novel power—in particular, he claims, to produce physical effects that lower-level entities and features cannot cause, or in any case cannot cause alone.

Contemporary accounts of Strong emergence also typically agree in taking emergent features to have or bestow fundamentally novel powers, not had (or had only in derivative fashion) by lower-level physically acceptable goings-on. For example, [Silberstein and McGeever \(1999\)](#) understand emergent features as having irreducible causal capacities:

Ontologically emergent features are features of systems or wholes that possess causal capacities not reducible to any of the intrinsic causal capacities of the parts nor to any of the (reducible) relations between the parts. (186)

[O'Connor and Wong \(2005\)](#) also make explicit that emergent features are “fundamentally new”, not just in being (perhaps epiphenomenally) different, but more specifically in having fundamentally new causal capacities:

[A]s a fundamentally new kind of feature, [an emergent feature] will confer causal capacities on the object that go beyond the summation of capacities directly conferred by the objects microstructure. (665)

And in [Wilson 2002a](#), I argue that naturalistic good sense can be made of the Strong emergentist posit of fundamentally novel powers, as reflecting novel fundamental interactions that come into play only at certain levels of compositional complexity.

¹²Some British emergentists put their view in terms of ‘in-principle’ failure of deducibility, but this reflected their assuming that such failures were indicative of fundamental novelty—that is, of Strong emergence. As [McLaughlin \(1992\)](#) notes, “the Emergentists do not maintain that something is an emergent because it is unpredictable. Rather, they maintain that something can be unpredictable because it is an emergent” (73). I’ll discuss this issue further in Chapter 4.

2.2.1 The New Power Condition

Summing up, Strong emergentists suppose that at least some higher-level special science features are fundamentally novel, specifically in having powers (“causal capacities”, etc.) not had by (“not directly conferred by”) their lower-level physically acceptable dependence base features, as per the following *New Power Condition*:

New Power Condition: Token higher-level feature S has, on a given occasion, at least one token power not identical with any token power of the token lower-level feature P on which S synchronically materially depends, on that occasion.

Three clarificatory points. First, here I make explicit the aforementioned supposition that at issue are tokens of features, and relatedly, token powers. Second, I do not make explicit the qualification that the novel power at issue is *fundamentally* (as opposed to non-fundamentally) novel, since this qualification follows from the fact of non-identity of token powers, given that the lower-level feature P is a synchronically occurring feature of a lower-level configuration or plurality. Third and relatedly, the qualification in Strong emergentist presentations that the novel power or capacity of a Strongly emergent feature is more specifically fundamentally novel is primarily aimed at ruling out as Strongly emergent entities or features whose novel causal capacities simply reflect the difference between unaggregated and aggregated lower-level entities and features. For example, the Strong emergentist does not claim that shape (a property of pluralities or configurations of atoms) Strongly emerges from features of individual atoms—not least because such “emergence” would not support the Strong emergentist’s intended contrast with physicalism. Accordingly, here and throughout, the *New Power Condition* should be read as involving a fundamentally rather than merely non-fundamentally novel power.

Moreover, Strong emergentists uniformly assume that among the (fundamentally) novel powers of a Strongly emergent feature are powers to influence lower-level physically acceptable goings on. So understood, *New Power Condition* entails the rejection of *Physical Causal Closure*; hence the Strong emergentist posi-

tion is naturally associated with a strategy of response to the problem of higher-level causation that proceeds by rejecting *Closure*.¹³

Let's now return to the problem of higher-level causation, starting with the second, simpler case, to see how satisfaction of the *New Power Condition* enters into implementation of the Strong emergentist strategy for responding to this problem. In the case where special science feature S causes a physically acceptable feature P^* (case 2), the Strong emergentist strategy involves, to start, the supposition that S satisfies the *New Power Condition* specifically in having a fundamentally novel power to bring about P^* . For example, S might be a Strongly emergent state of being thirsty, which depends on physically acceptable feature P , and which in the circumstances causes a physical reaching for a nearby glass of water P^* . On this assumption, P^* does not, contrary to the assumption of *Physical Causal Closure*, have a purely lower-level physically acceptable cause: as per the *New Power Condition*, P has no power identical with S 's power to cause P^* ; hence either P is not at all a cause of P^* (does not have any power to cause P^*), or else, if P can be understood to cause P^* (to have a power to cause P^*), P has this power only in a derivative sense, in virtue of P 's being a dependence base for S , which non-derivatively has the power at issue.¹⁴ Either way, P fails to be a purely sufficient lower-level physically acceptable cause of P^* ; and without loss of generality, it moreover follows that P^* has no purely sufficient lower-level physically acceptable cause, contra *Physical Causal Closure*, and overdetermination is avoided, as follows:

¹³More generally, as we'll shortly see, only against the backdrop of the rejection of *Closure* does satisfaction of the *New Power Condition* provide a basis for avoiding overdetermination.

¹⁴ S 's causing of P^* might be entirely independent of P , or it might be that S and P jointly cause P^* ; either route to the production of P^* is compatible with the denial of *Physical Causal Closure*. I'll revisit these options down the line.

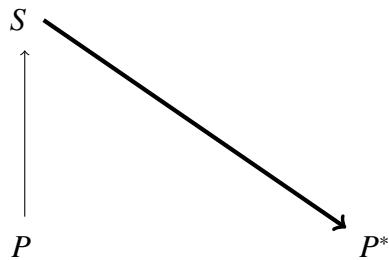


Figure 3: The Strong emergentist response to case 1

Next, suppose (as per case 1), that S causes another mental state S^* —say, a desire to drink some water. Here the Strong emergentist supposition is that S satisfies the *New Power Condition* specifically in having a fundamentally novel power to bring about S^* —that is, a power that P doesn’t have (either at all, or non-derivatively). Interestingly, even though the novel power at issue here is not directed at the production of a lower-level physically acceptable effect, it remains that satisfaction of the *New Power Condition* in this case requires the falsity of *Physical Causal Closure*. Why so? Because, if *Closure* held in this case, P would have a non-derivative power to cause S^* —by being a purely sufficient lower-level physically acceptable cause of P^* , which in turn nomologically necessitates S^* . This would contradict the assumption that S has a fundamentally novel power to cause S^* . In order to avoid such contradiction, the Strong emergentist must deny *Physical Causal Closure*, even when the novel power had by Strongly emergent S is for the causing of a special science feature S^* . In this case, the strategy is as follows:

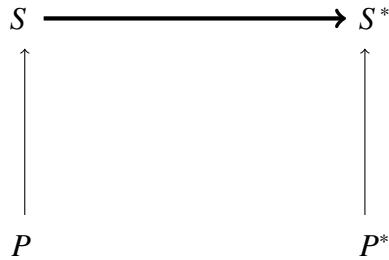


Figure 4: The Strong emergentist response to case 2

A remaining question about the Strong emergentist treatment of this case concerns what is responsible for S^* 's having the physically acceptable dependence base P^* that it does. Given that P is not (on this view) itself up to the task of causing P^* , there are two possibilities here: first is that S^* carries with it its own dependence base P^* , such that S , in causing S^* , also causes P^* (such that cases of the second type also involve cases of the first type— S causes P^*); another is that S and P jointly cause P^* (with S either independently causing S^* , or causing S^* jointly with P). We will explore some of these options in more detail in Chapter 4.

2.2.2 The schema for Strong emergence

Prima facie, satisfaction of the *New Power Condition* by a special science feature S avoids overdetermination while guaranteeing that S is both ontologically and causally autonomous from the lower-level physically acceptable feature P upon which S synchronically materially depends. First, since S has a token power (at a time or over a temporal interval) that P doesn't have (at that time or over that interval), S is distinct from P (by Leibniz's law); hence S is ontologically autonomous. Second, in having a novel token power, S can cause an effect that P can't cause, or that P can't cause in the same (non-derivative) way as S ; hence S is causally autonomous—that is, S is distinctively efficacious as compared to P . The *New Power Condition* at the heart of the Strong emergentist's strategy for resolving the problem of higher-level causation thus provides, in straightforward fashion, the basis for our first schema for metaphysical emergence:

Strong emergence: Token apparently higher-level feature S is Strongly metaphysically emergent from token lower-level feature P on a given occasion just in case, on that occasion, (i) S synchronically materially depends on P , and (ii) S has at least one token power not identical with any token power of P .

Some clarifications:

- The first condition minimally specifies synchronic material dependence, typically and comparatively neutrally understood as involving, in addition,

minimal nomological supervenience of Strongly emergent feature types on dependence base feature types.

- The second condition (effectively, the *New Power Condition*) captures the comparatively strong sense in which an emergent feature may be causally, hence ontologically, autonomous *vis-á-vis* the lower-level base feature upon which it synchronically materially depends.
- The schema is relativized to occasions, but it is worth noting that it would be reasonable to suppose that it suffices for the Strong emergence of S , *simpliciter*, that the condition is ever satisfied, and to suppose that it suffices for the Strong emergence of the feature type (of which S is a token), *simpliciter*, that any token feature S on any occasion satisfies (or would satisfy) the condition. These complications won't play a role in what follows.
- S is initially characterized as ‘apparently higher-level’, reflecting that the dialectical targets here are, in the first instance, special science or artifactual entities and features for which there is a *prima facie* case to be made (by appeal to distinctive special science types, laws, perceptual unity, compositionally flexible individuation conditions, etc.) that the features are in fact emergent.

2.3 Non-reductive physicalism and the *Proper Sub-set of Powers Condition*

Like Strong emergentists, non-reductive physicalists maintain that (some) special science features are real, synchronically materially dependent, distinct, and distinctively efficacious with respect to their base features. They too are non-substantival dualists or pluralists, maintaining that certain configurations (or pluralities) of entities whose constituents are ultimately wholly physical have lower-level features from which other features emerge, constituting emergent entities

and a novel level of natural reality; hence they are non-reductionists. But as physicalists, the sense in which higher-level features are real, synchronically materially dependent, distinct and distinctively efficacious cannot entail the rejection of *Physical Causal Closure*, which is core to the physicalist view that the physical goings-on are an existential and causal basis for all other broadly scientific phenomena. Rather, non-reductive physicalists reject *Non-overdetermination*, maintaining that distinct higher-level and base features can each be sufficient causes of a single effect, in virtue of standing in a relation that, while not identity, is intimate enough both to avoid overdetermination of the problematic double-rock-throw variety and to retain compatibility with *Physical Causal Closure*, hence with physicalism.¹⁵

In presenting their strategy, non-reductive physicalists typically endorse some or other ‘realization’ relation as holding between tokens or types of the features at issue. These relations include

- Functional realization (Putnam 1967, Fodor 1974, Papineau 1993, Antony and Levine 1997, Melnyk 2003, Polger 2007): higher-level types or tokens are functionally implemented by lower-level physically acceptable types or tokens.
- Mereological realization (Shoemaker 2000/2001, Clapp 2001): higher-level types or tokens are proper parts of lower-level physically acceptable types or tokens.
- The determinable-determinate relation (Yablo 1992, Wilson 1999 and 2009): higher-level types or tokens are determinables of lower-level physically acceptable types or tokens.
- Elimination in degrees of freedom (Wilson 2010b): the degrees of freedom (independent parameters needed to specify a composed entity’s law-governed properties and behaviours) needed to characterize higher-level en-

¹⁵How non-reductive physicalists aim to establish that higher-level features are not just efficacious, but distinctively so, will be addressed shortly.

tities E are strictly fewer than those needed to characterize the system of the entities e_i composing E .

The non-reductive physicalist then argues that on the assumption that their favored relation holds between higher-level and physically acceptable base features, problematic overdetermination is avoided, compatible with both physicalism and non-reduction. So, for example, Clapp (2001) says:

[M]ultiply realized mental properties, though real and causally efficacious, are better thought of as *parts* of their physical realizers. [...] Just as there is no causal and/or explanatory competition between a whole and its parts, so there is no causal and/or explanatory competition between instances of mental properties and instances of their physical realizers. (133; italics in text)

And Yablo (1992) suggests that if higher-level states are seen as determinables of lower-level determinate states, the concern that overdetermination must lead to one or other features' being excluded as efficacious dissipates:

[W]e know that [determinables and determinates] are not causal rivals. This kind of position is of course familiar from other contexts. Take for example the claim that a space completely filled by one object can contain no other. Then are even the object's parts crowded out? No. In this competition wholes and parts are not on opposing teams [...]. (183)

2.3.1 The *Proper Subset of Powers Condition*

The seeming diversity in these and other accounts of non-reductive realization hides a deeper unity of strategy, however, which again can be put in terms of a certain condition on powers (again, see Wilson 1999 and 2011; see also Shoemaker 2000/2001, Clapp 2001, and Shoemaker 2007).

To start, non-reductive physicalists maintain that every higher-level feature S stands in a relation to their lower-level base features P satisfying the following condition:

Token Identity of Powers Condition: Every token power of an apparently higher-level feature S , on a given occasion, is identical with a token power of a lower-level feature P on which S synchronically materially depends, on that occasion.

Note that *reductive* physicalists can and do accept the *Token Identity of Powers Condition*, so long as it is allowed that a feature can vacuously depend on itself. For example, if a given mental type is (as the reductive physicalist believes) identical to a given physical type, then tokens of these types, along with their associated token powers, will also be identical, and the *Token Identity of Powers Condition* will be satisfied.

Non-reductive physicalists moreover accept another, stronger condition, according to which at least some higher-level features S stand in a relation to their lower-level base features satisfying the following:



Proper Subset of Powers Condition: Token higher-level feature S has, on a given occasion, a non-empty proper subset of the token powers of the token lower-level feature P on which S synchronically materially depends, on that occasion.¹⁶

Motivating and seeing how the *Proper Subset of Powers Condition* supports a physically acceptable form of emergence, and how the non-reductive physicalist appeals to this condition in responding to the problem of higher-level causation, takes a bit more doing than in the case of the *New Power Condition* and associated form of Strong emergence. I'll start by arguing that each of the above standard accounts of non-reductive realization aims to ensure satisfaction of the *Proper Subset of Powers Condition*; I'll then argue that satisfaction of this condition provides a basis not just for a higher-level feature S to be distinct, but also for S to be distinctively efficacious; I'll then show how this condition provides an illuminating basis for understanding the non-reductive physicalist's response to the problem of higher-level causation; finally, I'll use the condition as the basis for the schema for Weak emergence.

¹⁶The requirement that the proper subset of powers be non-empty reflects the rejection of epiphenomenal features as metaphysically emergent, in the relevant sense.

The common strategy underlying non-reductive physicalist accounts

On functionalist accounts of realization, realized types are higher-order types associated with causal roles that, on a given occasion, are played by tokens of realizer types. A causal role is just a collection of powers. Hence if S is of a feature of functional type, then, on any given occasion, every token power of S will be numerically identical with a power of the base feature P that plays S 's causal role on that occasion. Functional accounts of realization thus satisfy the *Token Identity of Powers Condition*. Do such accounts also satisfy the *Proper Subset of Powers Condition*? One might think not, on grounds that instances of functionally realized features inherit *all* of the token powers of their realizing feature instances:

A functional reduction of pain has the following causal and ontological implications: Each occurrence of pain has the causal powers of its neural realizer [...] In general, if M occurs by being realized by N on a given occasion, the M -instance has the causal powers of the N -instance (Kim 2006, 554).

But in cases of multiple realizability, a functionally realized feature arguably has only a proper subset of the powers of its realizing feature(s), at both the type and token levels.

To start, recall the hardware/software analogy motivating functionalism, initially highlighted by Putnam (1967): the realizing systems are similar with respect to powers needed to implement a given software program, but different with respect to powers associated with their distinctive varieties of hardware. More generally, when a type of functionally characterized feature is multiply realizable, its realizing types will each have all the powers associated with the functional role, and more besides (where the further powers reflect differences between the multiple realizers). Hence the powers of the realized type will be a proper subset of those of each of its realizing types.

Moreover, this type-level proper subset relation between powers will arguably hold between token powers of the instantiated types, as per the *Proper Subset of Powers Condition*. It may make sense for a token feature to have fewer powers than its feature type, reflecting restrictions associated with circumstances in which

the feature occurs or is instantiated (see [Clarke 1999](#)). But as I've previously discussed (see [Wilson 2011](#) and [2015c](#)) it makes no sense for a token feature, whether functionally realized or not, to have more powers than its type; for if a token feature purportedly of a certain type had token powers not associated with the purported type, that would be reason for taking the instance not to be of the type. So functionally realized features satisfy the *Proper Subset of Powers Condition*.

Second, consider mereological (parthood-based) accounts of realization, according to which realized tokens (Shoemaker) or types (Clapp) are proper parts of base tokens/types. Proper parthood appears to satisfy the non-reductive physicalist's desiderata: proper parts are distinct from and yet in a sense nothing over and above a whole that is antecedently given,¹⁷ and may be efficacious without inducing overdetermination, as when both I, and my eye, cause a wink, or (as in Paul's [2002](#) case), both the plane and its wheels are causes of the runway's being touched. Both Shoemaker and Clapp suppose that mereological accounts satisfy the *Proper Subset of Powers Condition*; indeed, both see satisfaction of the condition as core to their accounts of non-reductive realization, with the appeal to mereology serving to illustrate their preferred way to satisfy the condition.

To start, [Shoemaker \(2000/2001\)](#) presents an account of realization based in a type-level version of the *Proper Subset of Powers Condition*:

Property X realizes property Y just in case the conditional powers bestowed by Y are a subset of the conditional powers bestowed by X (and X is not a conjunctive property having Y as a conjunct). (78)

He then claims that multiply realized feature types satisfy this condition:

Where the realized property is multiply realizable, the conditional powers bestowed by it will be a proper subset of the sets bestowed by each of the realizer properties. (78–9)

¹⁷As discussed in Chapter 1 ('Preliminaries'), and as more extensively discussed in [Wilson 2014](#), [Wilson 2016c](#), and elsewhere), specific metaphysical relations such as mereological part-whole serve as dependence relations against the backdrop of a supposed fundamental base. In the present case, the suggestion is that a lower-level physical feature is something like a more fundamental whole of which a higher-level feature is a non-fundamental proper part.

Shoemaker supports this claim by appeal to considerations similar to those canvassed for functional realization, with the main difference being that he takes all broadly scientific properties to be essentially characterized by distinctive sets of powers. When such a feature is multiply realized, its realizing types will share all the powers of the realized type, but will differ from each other in respect of further powers.

Shoemaker goes on to argue that this type-level proper subset relation between powers will plausibly hold between token powers of the instantiated types. In general, [Shoemaker \(2000/2001\)](#) notes, it is reasonable to suppose that if realized and realizer types are not identical, in virtue (at least in part) of bestowing different sets of conditional powers, then neither will be their instantiations:

[I]t seems doubtful that we should identify the mental property instance with the instance of the physical property that realizes it—or that we should identify the instance of red and the instance of scarlet. If we think of the instantiation of a property as the conferring on something of the conditional powers associated with that property, then when properties confer different sets of conditional powers, the instantiation of one of them is not identical with the instantiation of the other. (28)

These remarks suggest that in cases of multiple realization (whether mereological or not), the *Proper Subset of Powers Condition* is satisfied. Coupled with a view on which properties are not just essentially but also exhaustively individuated by their powers (as per [Shoemaker 1980](#)), realization is naturally interpreted as involving a proper parthood relation between tokens of realized and realizing features:

Likewise, the instantiation of a realizer property entails, and might naturally be said to include as a part, the instantiation of the functional property realized. ([Shoemaker 2000/2001](#), 81)

Alternatively, we can backwards-engineer the need to satisfy the *Proper Subset of Powers Condition* from a mereological approach. If features have non-causal aspects, that a realized state is a proper part of a realizing state need not indicate that the realized state has any powers at all (if the overlap concerns only the

non-causal aspect), much less that it is distinctively efficacious. Hence if proper parthood is to provide a basis for higher-level efficacy, the overlap must be specifically in respect of powers, as per the *Proper Subset of Powers Condition*.¹⁸

Next, consider accounts of non-reductive realization in terms of the determinable/determinate relation, the relation of increased specificity paradigmatically holding between colors and their shades. Yablo (1992) expected the suggestion that, e.g., mental features stand to their physical realizations in the relation that colors bear to their shades to be met with some incredulity. One way to make his conjecture more plausible is to put the point in terms of the causal powers of the properties involved (see Wilson 1999 and 2009). Consider a patch that is red, and more specifically scarlet. Sophie the pigeon, trained to peck at any red patch, is presented with the patch, and she pecks. The patch's being red caused Sophie to peck—after all, she was trained to peck at red patches. But the patch's being scarlet also caused Sophie to peck—after all, to be scarlet just is to be red, in a specific way. Nonetheless, Sophie's pecking was not problematically overdetermined. Plausibly, this is because each token power of the determinable red instance is numerically identical to a token power of its determining scarlet instance. Similarly, the proponent of this account of realization maintains, for the case of *S* and *P*, in which case the determinable/determinate relation satisfies the *Token Identity of Powers Condition*.

Again, one might doubt that the relation moreover satisfies the *Proper Subset of Powers Condition*, on grounds that determinable and determinate instances are identical (Macdonald and Macdonald 1995, Ehring 1996), such that an instance of a determinable feature inherits *all* of the powers of the determinate that realizes it on a given occasion. But again, the powers of a determinable feature are arguably

¹⁸Importantly, and notwithstanding that Shoemaker and Clapp each take a powers-based mereological account of realization to naturally flow from a causal theory of properties, on which properties are essentially and exhaustively constituted by sets of powers (as per Shoemaker 1980, Ellis 2001, and others), such an account of properties is not required in order for an account of non-reductive realization understood to be seen as ensuring satisfaction of the *Proper Subset of Powers Condition*. I'll return to this issue in Chapter 3, when considering the objection (pressed by Melnyk 2006, among others) that satisfaction of the schema for Weak emergence guarantees the physical acceptability of the associated emergent feature only on the assumption of a causal theory of properties.

only a proper subset of those of its determinate features, at both the type and token levels.

To start, note that given Sophie’s training, she would have pecked even if the patch had been a different shade of red (burgundy, say); but not so for Sophie’s cousin Alice, trained to peck only at scarlet patches. This suggests that the determinable type *red* has fewer powers than its determinate types (*scarlet*, *crimson*, etc.). More generally, since broadly scientific determinables are associated with distinctive sets of powers, and are typically “multiply determinable”, the powers of determinable feature types will typically be a proper subset of those of their determinate feature types.

Moreover, this relation will plausibly hold between token powers of determinable and determinate instances. Again, while it might make sense for an instance of a given feature type to have fewer powers than are associated with its type (reflecting restrictions on available circumstances of instantiation, or the like), it does not make sense for an instance of a given feature type to have more powers than are associated with the type. In particular, were a feature purportedly of a determinable type to have more token powers than are associated with the type, that would be reason to think that the instance was not of that type. Hence a determinable/determinate account of realization satisfies the *Proper Subset of Powers Condition*.

Finally, consider an account of non-reductive realization as involving an elimination in degrees of freedom (DOF). In Wilson 2010b, I offer a DOF-based characterization of what I call ‘Weak ontological emergence’:

Weak ontological emergence (DOF): An entity E is Weakly emergent from some entities e_i if

1. E is composed of the e_i , as a result of imposing some constraint(s) on the e_i .
2. For some characteristic state S of E : at least one of the DOF required to characterize a realizing system of E (consisting of the e_i standing in the e_i -level relations relevant to composing E) as being in S is eliminated from the DOF required to characterize E as being in S .

3. For every characteristic state S' of E : Every reduction, restriction, or elimination in the DOF needed to characterize E as being in S' is associated with e_i -level constraints.
4. The law-governed properties and behavior of E are completely determined by the law-governed properties and behavior of the e_i , when the e_i stand in the e_i -level relations relevant to their composing E .

(There is more to say by way of elucidating and defending this account, which I will revisit in Chapter 3 and expand on in Chapter 5; I take it that the general idea is clear enough for present purposes, and direct the interested reader to the more detailed discussion in my 2010b.) Having offered this account, I then argue that when an entity E satisfies *Weak ontological emergence*, then E will satisfy the *Proper Subset of Powers Condition*, even if E is only singly realized by a relational entity e_r :

E , being Weakly emergent, is treated by a theory extracted from a more fundamental theory treating of its composing e_i , as a result of certain constraints being imposed on the latter. The laws of the extracted theory express what happens when the e_i stand in relations associated with the e_i -level constraints, and the laws in the more fundamental theory express what happens when the e_i stand both in these and in other relations not associated with the constraints. For example, the laws of molecular physics express what happens in circumstances conducive to the existence of molecules, and the laws of atomic physics express what happens in these as well as in other circumstances involving, say, energies or temperatures too high for molecules to exist.

What does this mean for what powers should be assigned to E ? Plausibly, what powers an entity has are a matter of what it can do. And plausibly, the sciences are in the business of expressing what the entities they treat can do. It follows that, plausibly, what powers an entity has are expressed by the laws in the science treating it.

So again consider E and the relational entity e_r that singly realizes it. Given that what powers an entity has are expressed by the laws in the science treating it, the powers of E are those expressed by the laws

in the extracted theory treating E , while the powers of e_r are those expressed by the laws in the more fundamental theory treating the e_i (and any associated relational entities). It follows that E has a proper subset of the powers had by e_r .

For example, suppose e_r is a quantum relational entity, and E is a macro-entity singly realized by e_r . Then the powers of E include all those powers to produce, either directly or indirectly, effects that can occur in the constrained circumstances in which the quanta form macroscopic entities (in other words: in the macroscopic limit). The realizing entity e_r has all these powers, and in addition has all those powers to produce, either directly or indirectly, effects that can occur in circumstances that are not so constrained, and in which quantum physics is operative—for example, effects occurring in circumstances involving temperatures or energies in which atoms, but not molecules, can exist. Hence E has only a proper subset of the powers of e_r . (304–305)

In this passage I speak of the powers of the entities E and e_r as standing in the proper subset relation, reflecting that in the paper I cut out the feature middleman to focus on the emergence of an entity E ; but since this relation holds in virtue of some feature S of E 's having an eliminated set of DOF, E 's standing in the subset of powers relation to e_r will ultimately be a matter of E 's having a feature S standing in the subset-of-powers relation to some lower-level feature P of e_r upon which S depends.¹⁹

Summing up: a wide range of accounts of non-reductive realization arguably satisfy the *Proper Subset of Powers Condition*.²⁰

Distinctive power profiles as a basis for distinctive efficacy

In having only a proper subset of the token powers of the lower-level physical feature P upon which it depends, on a given occasion, a higher-level feature S

¹⁹ And, moreover, E 's not having any features with fundamentally novel powers, à la Strong emergence.

²⁰ See Wilson 1999 for discussion of how some other accounts—e.g., Pettit's (1995) account as appealing to a “dot-shape” analogy, according to which the shape is distinct from, but nothing over and above, a collection of dots—also arguably satisfy the *Proper Subset of Powers Condition*.

satisfying the *Proper Subset of Powers Condition* will clearly be distinct from P , by Leibniz's law. Might S also be causally autonomous—distinctively efficacious *vis-à-vis* an effect that P also (by *PhysicalCausalClosure*) causes, as required if S is to be genuinely metaphysically emergent?

Yes, supposing that a feature may be distinctively efficacious in virtue of having a distinctive set of causal powers, or distinctive power profile. The underlying promise of non-reductive physicalism as making sense of the distinctive efficacy of (at least some) special science entities and features lies in the claim that S 's causal autonomy does not require that S have a distinctive power: it is enough that S have a distinctive set (collection, plurality) of powers.

One case for taking the having of a distinctive power profile to be sufficient for causal autonomy appeals to difference-making or other “proportionality” considerations, in cases where S (or S 's type) is multiply realizable.²¹ Again, suppose that S is a state of feeling thirsty, which causes a physical reaching for a glass of water (effect E). Now suppose that S (or another instance of S 's type, etc.) were realized by P' rather than P , in circumstances relevantly similar to those in which S caused E . Would E (or an event of E 's type) have still occurred? Intuitively, yes, since the only powers that matter for the production of E are the powers associated with S : powers differing between P and P' (e.g., to produce different readings on a neuron detector) don't make a difference to, and hence are in this sense irrelevant for, E 's production. That S 's distinctive power profile contains just those powers relevant or “proportional” to E 's production provides a principled reason for taking S 's efficacy *vis-à-vis* E to be distinctively different from P 's, notwithstanding that (as per S 's satisfaction of the *Token Identity* and *Proper*

²¹The idea that proportionality considerations might potentially serve as a basis for distinctive higher-level efficacy stems from Yablo's (1992) discussion. There he suggests, in particular, that a candidate determinable cause (e.g., the patch's being red) might be more proportional to a given effect (e.g., Sophie's pecking), on a given occasion, than the associated candidate determinate cause (e.g., the patch's being scarlet), in that the determinable has an “essence” tracking both sufficiency and difference-making considerations (e.g., if the patch had been crimson rather than scarlet, Sophie would still have pecked). Here I focus on difference-making considerations, since (against the background assumption of *Physical Causal Closure* and associated satisfaction of the *Token Identity* conditions) it is difference-making rather than sufficiency which is distinctive of higher-level efficacy.

Subset conditions), S doesn't cause anything that P (or its other realizers, on other occasions) doesn't also cause.²²

Another case for causal autonomy reflects that distinctive power profiles are typically associated with distinctive systems of laws—for example, the special science treating entities of S 's type. Plausibly, systems of laws track causal joints in nature; hence when S is of a special science type, its distinctive power profile is similarly plausibly understood as tracking such a distinctive causal joint. Moreover, given the holding of the *Proper Subset of Powers Condition*, and consonant with a general understanding of special science phenomena as abstracting away from various lower-level details, the causal joints at issue here concern comparatively abstract goings-on. This is, I think, what [Antony and Levine \(1997\)](#) have in mind when they say that for causal autonomy of functional properties, “What we really need is a “realization-indifferent” regularity: a contingent regularity that essentially involves the second-order property, and that applies to any instance of the property, no matter the form of realization”. Moreover, as [Antony \(2003\)](#) notes, even in the absence of multiple realizers, one can still make sense of the presence of properties and laws that are, as she puts it “*at a higher level of abstraction*” (emphasis in text) as compared to lower-level properties and laws:

[M]ultiple realizability is something of a red herring. What matters, fundamentally, is not whether there could be minds embodied in things other than brains, but rather whether there is a level of reality beyond the level at which brains are normally studied—whether psychological kinds are “really there” [in addition to] the already recognized kinds in chemistry, biology and the other established sciences. If this is what is at stake, then it would not matter if brains turned out to be the only kinds of things that realize minds in any nomologically possible worlds. [...] The functional descriptions, and the generalizations given in terms of the psychological categories defined at the functional level would still, in this case, be autonomous from the de-

²²Note that nothing in this line of thought requires that one accept a “difference-making” account of causation or relatedly, that one reject P as being a cause of E . The suggestion is simply that attention to difference-making considerations provides a principled ground for S 's being distinctively efficacious as compared to P . I'll return to this issue in Ch. 3.

scriptions, generalizations and categories that turned up at the level of the realizers. (8)

Also worth noting is that causal joints may overlap—both in respect of a given token power and in respect of an associated effect E . If the joints as a whole are different, this provides a principled reason for taking S to be distinctively efficacious *vis-á-vis* E , in that S produces E as part of a different system of laws (different causal joint) than P . And if we understand causal relations as involving exercises of a given power, this is, I think, what Macdonald and Macdonald (1995) have in mind when they say that for mental properties may be causally autonomous, “any instance of a cause-effect relation can be an instance of more than one pattern” (71).

The key suggestion here is that *there are two ways for a higher-level feature to be distinctively efficacious* as compared to the lower-level feature(s) upon which it depends. One way, emphasized by Kim and others, is for the higher-level feature to be associated with a new power to produce the effect; here the distinctive efficacy (associated with Strong emergence) is located in the (novel) power itself. Another way—that at issue in the powers-based subset strategy—is for the higher-level property to be associated, with at least the strength of the laws of nature, with a collection of powers that are relevantly or distinctively proportional to the effect, in the ways indicated by difference-making considerations and comparatively abstract special science laws; here the distinctive efficacy reflects, in part, facts about which power profiles are associated with which features.

The non-reductive physicalist response to the problem of higher-level causation

Let’s now see how satisfaction of the *Proper Subset* condition enters into the non-reductive physicalist’s response to the problem of higher-level causation, and in particular into their rejection of *Non-overdetermination* (according to which, with the exception of double-rock-throw cases, effects are not causally overdetermined by distinct individually sufficient synchronic causes).

In case 1, special science feature S again depends on a lower-level physically acceptable feature P , and S causes another special science feature S^* , which depends on a lower-level physically acceptable feature P^* ; again, we might suppose that S is a state of feeling thirsty, and S^* is a desire to reach for a nearby glass of water. Here the non-reductive physicalist's strategy involves, to start, the supposition that S satisfies the *Proper Subset of Powers Condition* in such a way as to have the power, on a given occasion, to bring about S^* . As per the condition, this token power is identical to one had by P ; hence when S causes S^* , so too does P . S and P are each sufficient causes of S^* ; they are distinct, by Leibniz's Law, since S , in satisfying the *Proper Subset of Powers Condition*, has fewer token powers than P ; and S and P , being synchronic, are not parts of a diachronic causal chain. Consequently, S^* is overdetermined, contra *Non-overdetermination*. Yet, the non-reductive physicalist maintains, the overdetermination here is not of the variety that is supposed to be problematic. In a double-rock-throw case, distinct token powers and associated causal chains converge on a single effect (a window breaking). In cases where the higher-level cause satisfies the *Proper Subset of Powers Condition*, however, two distinct but synchronically materially dependent features are each associated with the same token power and hence the same causing. The overdetermination here is benign—indeed, is no more problematic than cases where both the plane and its wheels are causes of the runway's being touched.²³ Hence it is that *Non-overdetermination*, at least in fully generality, must be rejected, leaving the way clear for higher-level efficacy (about which more anon). Here it's worth representing the features at issue as having overlapping sets of powers, with each power represented as a dot:

²³Note that, contra a guiding supposition of Morris 2011b, the motivations for thinking that satisfaction of the *Proper Subset of Powers Condition* blocks problematic overdetermination do not hinge on the illustrative analogy to cases of benign part-whole overdetermination. Rather, that the overdetermination is benign follows just from the fact that, if the condition is satisfied, only one token power (however understood) is manifest on the occasion in question, in which case the overdetermination here is nothing like that at issue in double-throw cases.

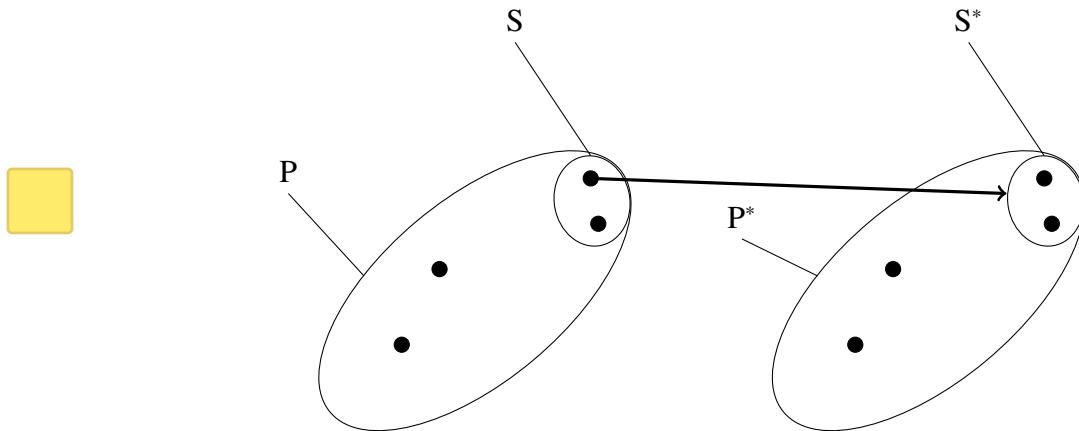


Figure 5: The non-reductive physicalist's response to case 1

In case 2, S rather causes a physical feature P^* . In the first instance, the treatment of this case is a variation on the same theme:

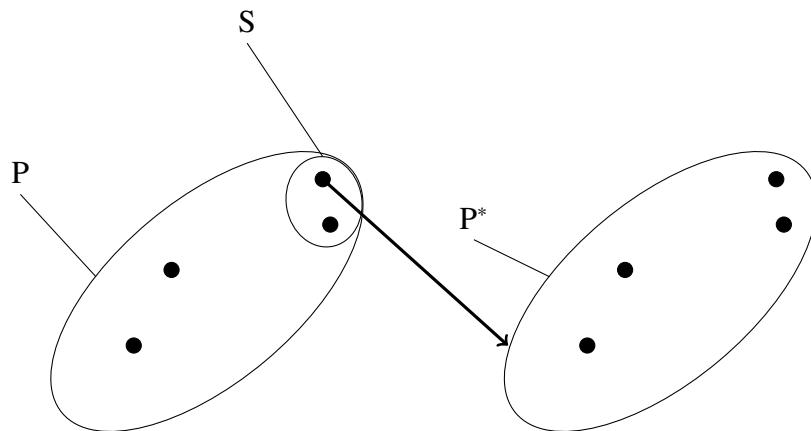


Figure 6: The non-reductive physicalist's response to case 2, version 1

That said, there is a subtlety here that must be addressed, associated with the possibility that no other lower-level feature besides P can, in the relevant circumstances, cause P^* . In this case, the supposed distinctive autonomy of S cannot rely on difference-making considerations, according to which if S had been realized by some lower-level physical property other than P , S would still have caused P^* .

Nor can S 's distinctive efficacy *vis-à-vis* P^* be a matter of S and P^* mutually occupying a distinctively abstract level of grain, since P^* is, by assumption, a lower-level physical feature.

In such a case, we can still accommodate the seeming efficacy of S *vis-à-vis* physical goings-on by taking these appearances to concern, not lower-level goings-on (at, e.g., the quantum level), but rather physically acceptable goings-on P' at some level lower than that at which S is properly located, but higher than that at which S 's ultimate lower-level physical realizer P is located. For example, the non-reductive physicalist can accommodate mental state S 's being distinctively efficacious *vis-à-vis* some sort of physical behaviour P' —say, reaching for a glass—which is also realized by P^* but for which difference-making considerations *vis-à-vis* S would be present. In this case, a more accurate picture would be the following:

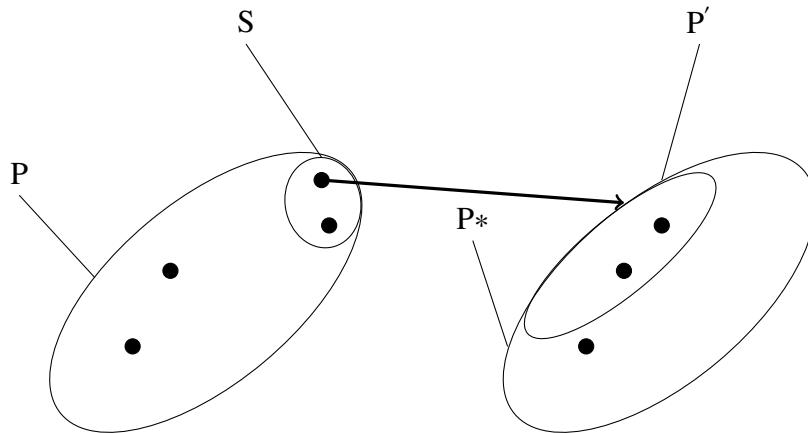


Figure 7: The non-reductive physicalist's response to case 2, version 2

Indeed, irrespective of whether P is the only lower-level physical feature capable of causing P^* , this sort of model might more generally make better sense of the seeming efficacy of special science features *vis-à-vis* physically acceptable goings-on.

It is also worth noting that the strategy here is compatible with physicalism. The main concern about S 's physical acceptability, given its reality, synchronic

material dependence on, and distinctness from its dependence base feature P , turns on the possibility that S might be Strongly emergent from P —that is, that S might have, as per the *New Power Condition*, a (novel, fundamental) power not had by P , of the sort undermining *Physical Causal Closure*. But S 's satisfaction of the *Proper Subset of Powers Condition* is incompatible with S 's satisfaction of the *New Power Condition*, ruling out S 's being Strongly emergent.²⁴

2.3.2 The schema for Weak emergence

Prima facie, satisfaction of the *Proper Subset of Powers Condition* makes room for S 's being metaphysically emergent. Satisfaction of this condition guarantees that S is ontologically autonomous from P : since S has a proper subset of the token powers of P , S is distinct from P , by Leibniz's law. It moreover makes room for S to be distinctively efficacious *vis-à-vis* P , in virtue of S 's having a distinctive power profile, tracking difference-making or proportionality considerations, or a distinctive level of causal grain.

We have thus arrived at our second schema for metaphysical emergence:

 *Weak emergence*: Token apparently higher-level feature S is Weakly metaphysically emergent from token lower-level feature P on a given occasion just in case, on that occasion, (i) S synchronically materially depends on P ; and (ii) S has a non-empty proper subset of the token powers had by P .

Some clarifications:

²⁴As discussed in Wilson 1999, Satisfaction of the *Proper Subset of Powers Condition* also appears to block other live routes to physical unacceptability, associated with S 's being non-natural (see Moore 1903) or supernatural (as per Malebranchean occasionalism). Moore used the term ‘non-natural’ as indicative of epistemological irreducibility (more specifically: indefinability), which is arguably compatible with physicalism (see Wilson 2002a). In any case, if S 's being epistemologically irreducible is deemed naturalistically (hence physically) problematic, this must be because such irreducibility indicates that S 's existence involves something metaphysically new relative to ('over and above') P ; on the non-reductive physicalist's operative assumption that S is efficacious, the problematic addition in question would presumably either be or entail S 's having of a non-natural or supernatural causal power, not had by physically acceptable P . But the having of such a power is ruled out if the *Proper Subset of Powers Condition* is satisfied.

- Again, the first condition minimally specifies synchronic material dependence, understood as involving, in addition, minimal nomological supervenience of the higher-level feature on lower-level features.
- The second condition (effectively, the *Proper Subset of Powers Condition*) captures the comparatively weak sense in which an emergent feature can be causally, hence ontologically, autonomous *vis-à-vis* its base feature.
- Again, the condition is relativized to occasions. If one wants to maintain that token feature S is Weakly emergent, *simpliciter*, one needs to generalize the condition to apply to all occasions on which S exists, as follows:

Weak emergence: Token higher-level feature S is Weakly metaphysically emergent, *simpliciter*, from lower-level physically acceptable features just in case for every occasion on which S exists (i) S synchronically materially depends on some token lower-level physically acceptable P , and (ii) S has a proper subset of the token powers had by P .

Further quantification over all tokens of S 's type would be required to establish that S 's type is Weakly emergent from (any) lower-level physically acceptable feature types. These generalizations won't make a difference in what follows.

The main objection to non-reductive physicalism, diagnosed

The problem of higher-level causation poses a general challenge to anyone wanting to make sense of the seemingly layered structure of natural reality, but as Kim presents the problem, it is in the first instance directed at non-reductive physicalist accounts of this structure. More specifically, Kim maintains, non-reductive physicalism cannot avoid problematic overdetermination without collapsing either into reductive physicalism (contra non-reduction) or expanding into Strong emergentism (contra physicalism). But, as above, the *Proper Subset of Powers Condition* and associated schema for Weak emergence provide a schematic basis for avoiding problematic overdetermination in a way that is compatible with both

physicalism and non-reduction; moreover, several contemporary accounts of non-reductive realization appear to guarantee satisfaction of the *Proper Subset of Powers Condition*, and so are appropriately seen as implementing the powers-based subset strategy at the heart of Weak emergence. In Chapter 3, I will consider various second-pass objections to the strategy, either in general or in certain of its specific implementations; but first I want to highlight the core presupposition giving rise to Kim's dilemma, the denial of which makes room for the strategy.

The first horn turns on the need for the non-reductive physicalist to endorse the *Token Identity of Powers Condition* (Kim's “causal inheritance principle”), which is indeed plausibly required for physicalism. In early discussions of the dilemma, Kim supposes that realized instances inherit *every* power of their realizing instances. So, for example, in his (1993a), he assumes that any higher-level property that is both efficacious and irreducible must have irreducible powers. But this supposition neglects the viability of locating irreducibility—both ontological and causal—in the having of a distinctive power profile. More recently, Kim allows that the *Token Identity of Powers Condition* may be satisfied via the *Proper Subset of Powers Condition*; but still fails to consider reasons for thinking that satisfaction of the latter condition suffices for ontological and causal autonomy.²⁵

Pereboom (in Pereboom and Kornblith 1991 and Pereboom 2002) similarly fails to consider these reasons, in assessing versions of non-reductive physicalism which satisfy the *Token Identity of Powers Condition*. He says, concerning a case where the higher-level feature is mental, that if a “token-identity thesis for these causal powers” is held,

...then the causal powers to which the psychological explanation refers would in the last analysis, in fact, be microphysical. Psychological explanations might then presume a classification that clusters

²⁵Kim does say (see <http://brainbrain.blogspot.com/2006/07/kim-vs-subset-view-of-higher-level.html>) that, with respect to the powers associated with a given higher-level feature, “the Subset View doesn’t justify that [S] indeed does have those powers, it merely stipulates it, and showing how [S] could have those powers just is the problem of mental causation. So the Subset View doesn’t solve the problem of mental causation”. But the task of a response to the problem of mental causation is not to establish *that* a given mental feature has powers; it is to establish that, *given* that a mental feature has certain powers to produce certain effects, these would not problematically overdetermine the powers to produce the same effects had by its dependence base feature.

microphysical causal powers in a way distinct from how microphysics sorts them, but this would not compromise the microphysical status of those causal powers (2002, 500).

(Here the reference to an alternative classification might be seen as gesturing towards satisfaction of the *Proper Subset of Powers Condition*, in addition to the *Token Identity of Powers Condition*.) And Pereboom concludes that accounts requiring token identity of powers are incompatible with “robust non-reductive materialism” (502). But such remarks describe, rather than undermine, the powers-based strategy. Why, then, do Kim and Pereboom, among others, think that higher-level features cannot be genuinely (physically) non-reductive if they satisfy the *Token Identity of Powers Condition*—even if they moreover satisfy the *Proper Subset of Powers Condition*?

The reason, I speculate, is that these philosophers assume that there is only a single form of causal efficacy: that associated with the having and manifestation of a power. Consequently, they assume that the only way for a higher-level feature to be distinctively efficacious is for it to have a power that its base feature doesn’t have. If distinctive efficacy requires new powers, then satisfaction of either the *Token Identity of Powers Condition* or the *Proper Subset of Powers Condition* would rule out the distinctive efficacy of higher-level features, contra the intended understanding of non-reductive physicalism. Moreover, if distinctive efficacy requires new powers, it is natural to suppose, as Kim does, that if the physicalist insists on non-reduction, they will be pushed to the Strong emergentist horn of his dilemma, since new powers are indeed the hallmark of such emergence.

But there is another route to distinctive causal efficacy, associated not with distinctive powers, but with distinctive *collections* of powers. Again, this alternative conception may be grounded either in proportionality or difference-making considerations, or in distinctive systems of laws or causal joints; at a more abstract level, the alternative conception locates efficacy in facts about collections of powers that are partly extrinsic to facts about the individual powers manifested in the production of a given effect. Given this alternative, the dilemma dissolves—at least so it seems.

2.4 Merricks's overdetermination argument

Merricks (2003) argues that attention to considerations of causal overdetermination undermine certain views of the ontological status of ordinary objects such as baseballs, rocks, and tables; in this respect his argument is similar to Kim's overdetermination argument. Indeed, as we'll shortly see, Merricks's argument can be presented in a form that is very close to that standardly associated with Kim's argument.

Merricks's overdetermination argument differs from Kim's in three respects:

1. Merricks focuses on a threat of overdetermination attaching to higher-level objects as causes; Kim focuses on a threat of overdetermination attaching to higher-level features as causes, on the common assumption that objects are efficacious in virtue of certain features rather than others. Nothing deep turns on this difference, since one could translate talk of an object's causing an effect into talk of the effect being caused by the property of being that object (or an object of that type).
2. Unlike Kim, Merricks does not take the truth of physicalism (via *Physical Causal Closure*) to be a premise in his argument; however, he does suppose that with certain few exceptions—notably, persons—the effects of ordinary objects are completely determined by the lower-level physical goings-on upon which they depend. As such, while Kim (at least in his original expositions) takes his argument to apply to any purportedly higher-level feature, Merricks takes his argument to apply only to what we might call material ordinary objects, whose powers are plausibly seen as completely determined by the lower-level physical goings-on upon which they depend.
3. Merricks presupposes that the lower-level dependence base is a plurality, such that the dependence relation at issue, were it to be instantiated, would be a many-one relation holding between a plurality of lower-level entities or features and a single higher-level entity or feature. As previously, Kim assumes that a one-one relation is at issue; for Kim, the lower-level de-

pendence base entity is a relational lower-level aggregate—i.e., a micro-configuration—or a feature of such a micro-configuration. This difference is the most interesting of the three. Kim, who wants to preserve the truth of physicalism, responds to his version of the problem of causal overdetermination by endorsing reductionism—that is, by identifying seeming higher-level entities or features with micro-configurations or lower-level features of such configurations. Since Merricks thinks that micro-configurations would also be subject to causal overdetermination concerns, he could be a reductionist about objects such as baseballs only if he allowed that identity could hold between one and many—which he doesn’t.²⁶ Hence Merricks takes the conclusion of his argument to motivate eliminativism rather than reductionism about ordinary objects (and their would-be features).

Merricks’s overdetermination argument (which takes a baseball as a case-in-point, then generalizes) is along the following lines. To start, on the assumption that a given baseball exists (assumed for reductio), the baseball must be able to enter into producing an effect: at least for ordinary objects, Merricks reasonably supposes, something like Alexander’s Dictum (‘to be is to have causal powers’) is true. Now, it is reasonable to suppose, Merricks maintains, that each power purportedly had by the baseball to produce a given effect is already a power of the plurality of atoms arranged baseball-wise, upon which the baseball depends. (This premise corresponds roughly to Kim’s assumption of *Physical Causal Closure*.) For example, consider the power of a baseball to shatter a window. This power, Merricks suggests, is already a power of the atoms arranged baseball-wise—each exerting its own bit of force in joint operation with the others.²⁷ Now, as Merricks notes, in cases where some x s cause a given effect, it might nonetheless be that some y is causally relevant to the x ’s causing the effect in ways that do not induce

²⁶Again, see Cotnoir 2013 for reasons to think that one can make sense of a generalized one-many identity relation.

²⁷In fact, since windows are also composed higher-level objects, Merricks doesn’t think the operative effect is really ‘a shattering of a window’; rather, the effect consists in multiple atomic interactions between atoms entering into the plurality of atoms arranged baseball-wise and the atoms entering into the plurality of atoms arranged window-wise. He uses the ordinary expression, however, to keep discussion simple.

overdetermination, either by (i) y being one of the xs , (ii) y causing the xs to cause the effect, (iii) the xs causing y to cause the effect, or (iv) y and the xs being jointly sufficient for the effect. But, he argues, the baseball is not relevant to the causing of the shattering of the window in any of these ways. (Kim basically considers and rejects these same routes to a higher-level feature's being efficacious in ways not inducing overdetermination.) As such, Merricks maintains, were the baseball to exist and (in particular) shatter the window, this would overdetermine that effect, contravening the assumption that such events are not appropriately seen as overdetermined. Generalizing, the same would be true of any purported effect or associated powers of the baseball, violating the non-overdetermination premise. And here, Merricks suggests, the right response is to reject the supposition that the baseball is real. Generalizing still further: since similar considerations would apply to any ordinary object, no such objects exist.²⁸

As prefigured, Merricks's overdetermination argument can be put into a form that is very similar to that characteristic of Kim's argument. Recall the six premises of Kim's argument:

1. *Dependence.* Special science features synchronically materially depend on lower-level physically acceptable features.
2. *Reality.* Both special science features and their base features are real.
3. *Efficacy.* Special science features are causally efficacious.
4. *Distinctness.* Special science features are distinct from their base features.
5. *Physical Causal Closure.* Every lower-level physically acceptable effect has a purely lower-level physically acceptable cause.
6. *Non-overdetermination.* With the exception of double-rock-throw cases, effects are not causally overdetermined by distinct sufficient causes that are

²⁸More precisely, Merricks takes his argument to generalize to any material ordinary object; he allows that some composed objects exist—namely, ones that have powers that the plurality of their composing objects don't have. Here we see something like the supposition that only Strongly emergent composed objects exist. I'll return to this issue down the line.

not part of a single causal chain.

Merricks's argument is, *mutatis mutandis*, a special case of Kim's argument, directed in the first instance at a baseball (or the feature of being a baseball):

1. *Dependence*. Baseballs synchronically materially depend on atoms arranged baseball-wise.
2. *Reality*. Baseballs as well as atoms arranged baseball-wise are real (assumed for reductio).
3. *Efficacy*. Baseballs are causally efficacious.
4. *Distinctness*. Baseballs are distinct from atoms arranged baseball-wise.
5. *Physical Causal Determination*. Every effect caused by a baseball is also caused by atoms arranged baseball-wise.
6. *Non-overdetermination* The effects of atoms arranged baseball-wise are not systematically causally overdetermined.

Now, as above, the reductionist conclusion of Kim's overdetermination argument can be blocked, in principle, by appeal to either a Strong or a Weak emergentist strategy, according to which overdetermination is either denied or unproblematically accommodated by taking the higher-level entity (feature) to have either more powers (Strong emergence) or fewer powers (Weak emergence) than its dependent base entity (feature). Can these strategies for avoiding problematic overdetermination be applied in response to Merrick's argument, notwithstanding the differences between Merricks's and Kim's argument?

I think so. The first and second differences above clearly pose no barrier to implementing either strategy, since the strategies can be implemented if powers are directly associated with objects (rather than with features of objects), and in a way sensitive, as Merricks's discussion is, to whether or not the higher-level object (features) are taken to be physically acceptable or not. As for the third difference: here what is in the first instance required is that it make sense for pluralities to

have powers, or to have features associated with powers, which sets of powers can then be compared with those associated with ordinary objects. But this does make sense: we can talk of the powers of the lower-level dependence base, whether this base is understood as a configuration or as a plurality. And Merricks agrees that this makes sense—indeed, that pluralities can have powers is a crucial premise in his overdetermination argument. As such, we can consider, *vis-à-vis* the powers of a lower-level plurality or its associated lower-level features, whether a given entity or feature which is dependent on the this plurality has more, or fewer, token powers as compared to these goings-on—a project I will embark on shortly. And if the conditions in the schemas for either Weak or Strong emergence are satisfied in a given case, then in light of the considerations discussed here or in the next chapters, this would serve to support not just the existence but also the emergence of certain ordinary objects.

Indeed, Merricks is happy to allow that some higher-level composed objects do not invoke overdetermination, and hence are not candidates for elimination—namely, those having powers that are new as compared to the plurality of lower-level objects which compose them and which they correspondingly to some extent depend. As such, Merricks arguably appeals to a ‘new power’ conception of emergence—that is, Strong emergence—as a way of gaining the existence and efficacy of at least some composed objects—namely, persons. We’ll revisit this aspect of Merricks’s view in Chapter 6, where I’ll offer some reasons for thinking that Merricks’s commitment to the Strong emergence of persons and their mental states may undermine his supposed commitment to artifacts such as baseballs failing to have any powers that their physical base pluralities don’t have.

It remains, however, that on the assumption that *Physical Causal Determination* does apply to such artifacts, Merricks goes wrong in failing to see how the ‘proper subset’ approach at issue in the Weak emergentist’s strategy provides a basis for sidestepping overdetermination. Like Kim, Merricks supposes that, on the assumption that every power of the higher-level entity is already had by its lower-level dependence base entity, there is no way to make sense of higher-level efficacy without incurring problematically redundant efficacy and associated

powers. But this supposition is incorrect, for the Weak emergentist strategy provides a route to avoiding overdetermination (via the token-identification of every power of the higher-level entity with a power of its base entity) while making sense of higher-level efficacy as tracking not the having of new powers, but rather the having of a distinctive power profile, containing a proper subset of the powers of the dependence base entity—configuration or plurality, no matter—tracking difference-making considerations and abstract causal joints.²⁹

2.4.1 The status of configurations

As above, Merricks takes his overdetermination argument to support not only the elimination of any higher-level ordinary objects that fail to have new powers (i.e., fail to be Strongly emergent), but also the elimination of lower-level configurations (relational aggregates), whose powers *vis-à-vis* lower-level pluralities would be, he maintains, similarly problematically redundant. Given that many (perhaps most) accounts of emergence, of either Weak or Strong varieties, presuppose that such configurations exist, it is worth considering what sort of responses might be available to those many philosophers (including myself) who have endorsed one-one realization relations as at least potentially holding between lower-level and higher-level goings-on. For simplicity, in what follows I revert to the usual focus on special science goings-on.

Two strategies for responding to Merricks's concern are salient. One would be to allow that there can be metaphysical emergence of an intra-level variety, and to consider motivations for taking lower-level configurations to be genuinely metaphysically emergent in this way from lower-level pluralities. On the assumption that both configurations and pluralities are governed by the same (ultimately physical) laws, the status of the configuration as Strongly emergent is presumably not at issue; but perhaps a case could be made that lower-level configurations are Weakly emergent from lower-level pluralities, in satisfying the conditions in the

²⁹That said, as we will see in Chapter 6, certain motivations for taking artifactual ordinary objects to be Weakly emergent differ from motivations for taking special science entities (features) to be emergent.

associated schema. That said, the usual motivations for thinking that these conditions are satisfied would not be available for the case of intra-level dependence. For example, under the operative assumptions we cannot motivate the configuration's having a distinctive power profile on grounds that it enters into different laws (since the same laws govern both plurality and configuration); nor can this be motivated by appeal to difference-making considerations according to which a configuration of a given token or type would have been able to cause a given effect even had some member of the plurality upon which it depends been absent or substantively changed, for on the usual understanding, lower-level configurations are individuated, not just nomologically but essentially, by their composing parts.

Attention to the holding of relational bonds in a lower-level configuration suggests a second, more promising strategy of response to Merricks-style overdetermination concerns about such configurations; namely, to deny that composed lower-level configurations really are appropriately seen as synchronically materially dependent on associated pluralities, in any sense in which configuration and plurality would be competitors for the production of a given effect. After all, when, e.g., the atoms a_i in a lower-level plurality come to form a lower-level configuration composed of the a_i , this is as a result of the a_i coming to stand in relations that are typically *causal*—e.g., in our toy case, atomic bonding relations between the a_i . But on the usual understanding, causal relations take time, in which case a lower-level configuration would be better seen as a complex, diachronically produced effect of the uncomposed plurality. Supposing so, then since effects typically have different powers than their causes, then there would be no clear danger of configurations' problematically overdetermining the effects of their associated pluralities. I see something like this ‘same-level’ strategy for ontologically and causally distinguishing configurations from pluralities as operative in Goldwater’s (2015) discussion of whether what he calls ‘arrangements’ are subject to Merricks’s overdetermination concern:

Merricks’ argument [against configurations, or ‘arrangements’] is this: since simples—ultimate parts—are efficacious, there is no need for the higher level whole—which is thereby expendable or eliminable. But [...] as the whole exists at the same level as the parts, it is not

the case that all the work is done on one level with the next level up being (potentially) epiphenomenal. Instead, just as the gravitational force will differ between two objects ten instead of nine meters away, so too will a tablewise arrangement have different powers than some arrangement which is not quite tablewise. (375)

Moreover, even if a given plurality were appropriately seen as having the power to produce any effect produced by an associated configuration, the production of this effect would proceed by way of first causing the configuration, and so the configuration would be causally relevant to the plurality's causing the effect, and so fail to be an overdetermining cause, by Merricks's own lights.

Though obviously more could be said about how best to understand the relation between pluralities and configurations, this much is, I think, enough to motivate continuing on with the usual supposition that the dependence base goings-on in many cases of purported emergence are configurations (and ultimately, micro-configurations). That said, nothing deep turns on whether the lower-level goings-on serving as a dependence base for higher-level goings-on are configurations or pluralities, and those who are inclined to reject configurations can, in what follows, substitute talk of these with talk of lower-level pluralities.

2.5 The schemas for Strong and Weak emergence as core and crucial to metaphysical emergence

Let's sum up the results so far. Attention to the problem of higher-level causation points toward two strategies of response to this problem, associated with Strong emergentism and with non-reductive physicalism, which are, I have argued, able to accommodate the metaphysical emergence of higher-level entities and features, understood as coupling synchronic material dependence with ontological and causal autonomy (that is, with distinctness and distinctive efficacy). Each response involves a (different) condition on the token powers of a higher-level feature *vis-á-vis* the token powers of the lower-level feature upon which it depends on a given occasion, which condition, along with a minimally specified

condition on synchronic material dependence, is encoded in the associated schema for Strong and Weak emergence, as follows:

Strong emergence: Token apparently higher-level feature S is Strongly metaphysically emergent from token lower-level feature P on a given occasion just in case, on that occasion, (i) S synchronically materially depends on P , and (ii) S has at least one token power not identical with any token power of P .

Weak emergence: Token apparently higher-level feature S is Weakly metaphysically emergent from token lower-level feature P on a given occasion just in case, on that occasion, (i) S synchronically materially depends on P , and (ii) S has a proper subset of the token powers had by P .

I conclude that satisfaction of the conditions in either schema is, as I put it, core and crucial to metaphysical emergence of the sort at issue in the target cases. As earlier noted, I prefer this terminology to the usual though to my mind overly coarse-grained terms of necessary and sufficient conditions, since any schematic account needs to be sensibly filled in. But modulo this caveat, the results of this chapter can also be seen as providing *prima facie* reason to think that the conditions in the schemas are both necessary and sufficient for metaphysical emergence of both physically acceptable and physically unacceptable varieties—a bold claim, but one that, as I argue in ensuing chapters, is surprisingly robust.

In any case, attention to the problem of higher-level causation and the conditions on token powers in the strong emergentist and non-reductive physicalist responses makes clear the limited ways in which a higher-level feature can be causally autonomous *vis-á-vis* its base feature, as the operative conception of metaphysical emergence requires. First, the feature may have *more* powers than its base feature; second, the feature may have *fewer* powers than its base feature. In terms of effects: the higher-level feature may be distinctively efficacious in potentially contributing to causing *more* effects than its base feature, or it may be distinctively efficacious in potentially contributing to *fewer* effects than its base feature. Since complete coincidence of token powers doesn't make room

for causal autonomy (distinctive efficacy), these routes to metaphysical emergence exhaust the available options. I conclude that satisfaction of the conditions either in *Weak* emergence or in *Strong emergence* is *prima facie necessary* for metaphysical emergence of the sort at issue in the target cases.

We thus have a preliminary schematic answer—rather, two answers—to the first key question, ‘What is metaphysical emergence’? The answers are schematic, since as we have already seen and will see further in the chapters to follow, there are a number of ways in which either schema might be (or might aim to be) implemented. The answers are also at this point preliminary, since a number of objections have been raised to the viability of physically acceptable or physically unacceptable emergence, either in general or as specifically directed at the schemas for *Weak* or *Strong* emergence or the conditions on powers therein. In the next two chapters, I’ll consider these objections, and show that they can be answered.

Chapter 3

The viability of Weak emergence

In Chapter 2, I provided *prima facie* reasons for thinking that satisfaction of the conditions in the schema for Weak emergence is core and crucial to a broadly scientific higher-level feature’Weak emergences being metaphysically emergent from—synchronously materially dependent on, yet ontologically and causally autonomous relative to—a lower-level feature, in such a way that, if the lower-level feature is physically acceptable, then so too will be the higher-level feature. Again, the schema is as follows:

Weak emergence: Token apparently higher-level feature S is weakly metaphysically emergent from token lower-level feature P on a given occasion just in case, on that occasion, (i) S synchronically materially depends on P , and (ii) S has a (non-empty) proper subset of the token powers had by P .

In terms of necessary and sufficient conditions, the working hypothesis here is that conditions in the schema are necessary for Weak emergence of the physically acceptable variety, and modulo the caveat that the schema must be sensibly filled in, are also sufficient for this. Moreover, as I previously argued (§2.1.1), a representative range of existing accounts of non-reductive realization, including functional role accounts, mereological accounts, determinable-based accounts, and degree of freedom-based (DOF-based) accounts, are each plausibly seen as aiming to implement the schema for Weak emergence.

In this chapter, I consider and respond to a range of objections that have been or could be made to the viability of Weak emergence.¹ Some of these directly target higher-level features or entities on grounds that are different from the concern about causal overdetermination addressed in the last chapter; some question motivations for thinking that the conditions in the schema are ever satisfied; some target the sufficiency of the conditions on grounds that their satisfaction is compatible with reduction; some rather argue that satisfaction of the conditions is compatible with physical unacceptability; others target the necessity of the conditions, on grounds that there are alternative routes to physically acceptable emergence for the cases in question. The focus of many of the objections is on condition (ii) in the schema—i.e., the *Proper Subset of Powers Condition*—and so for simplicity I will sometimes present an objection as pitched at this condition; unless otherwise indicated, however, S 's satisfaction of the dependence condition in (i) is also assumed to be in place.

As we'll see, each of these objections admits of at least one response that could be endorsed by proponents of any of the aforementioned implementations of Weak emergence. Upon occasion, however, a response to a given objection is available that relies on a specific implementation of the schema. In particular, certain attractive responses appeal to either a determinable-based account or a DOF-based account, of the sorts I have previously endorsed. Supposing one endorses responses that presuppose specific implementations of Weak emergence, this accordingly restricts one's options for filling in the schema (and relatedly, the sufficiency claim); but again, this is a choice point.

Here as elsewhere, I assume that the dependence base features at issue are physically acceptable, and will speak of the ‘Weak emergentist’ as the proponent of the view that conformity of an apparently higher-level feature to the conditions in the schema for Weak emergence is core and crucial—and when sensibly filled-in, both necessary and sufficient for—emergence of the sort compatible with physicalism (i.e., ‘non-reductive realization’).

¹Prior to publication, I plan to address certain further objections—e.g., the objection, due to [Yates \(2016\)](#), according to which physical realization does not require satisfaction of the *Proper Subset of Powers Conditions*.

3.1 Objection: anti-realism about higher-level features

The first line of objection I will consider is that, notwithstanding the *prima facie* appearances of higher-level features satisfying the *Proper Subset of Powers Condition*, such features should be given a deflationary anti-realist interpretation, either on pragmatic/abstractionist or on epistemological grounds.

3.1.1 Pragmatic/abstractionist anti-realism

Certain anti-realists maintain that seeming satisfaction of the *Proper Subset of Powers Condition* reflects not an independently existing higher-level feature, but rather a (mere) mind-dependent abstraction from, or pragmatically motivated way of conceiving of, lower-level physically acceptable goings-on.

For example, Ney (2010) argues, after discussing how Shoemaker (2000/2001, 2003, and 2007) takes satisfaction of the *Proper Subset of Powers Condition* to serve as the basis for non-reductive realization, that an alternative way of understanding seeming satisfaction of the condition is available:

[One] can [...] think of things in the following way (nothing has been said to rule out this way of thinking): on this view, really the mental event (and realized tokens more generally) are just abstractions from concrete microphysical situations. They are abstractions in the sense that they are what we attend to when we focus only on a proper subset of a microphysical state's causal powers. (442)

Ney sees this reading as indirectly supported by Shoemaker's (2007) claim that realization relations form a hierarchy, where "those higher in such a hierarchy will be realized by those further down" (23), and where properties whose instantiations lie at the bottom of the hierarchy are, unlike other properties, "self-constituted". As she says:

Plausibly, [one] can tolerate the existence of entities that are not reducible, i.e., identical to, physical entities, so long as these things are

mere abstractions. Since realized events can very naturally be seen on Shoemaker's account as mere abstractions from their realizer events, since they are not self-constituted, I don't see why [one] cannot endorse this approach. (443)

On such an interpretation, "The only entities with genuine, (mind-)independent existence here are the microphysical states of affairs" (444). Though there might be a sense in which, on Ney's interpretation, mental states understood as 'mere' mind-dependent abstractions exist, in the sense relevant to our investigations here this interpretation is properly deemed 'anti-realist'.²

Heil (2003b) also argues for a kind of anti-realist conceptualism about higher-level features. As Heil sees it, embrace of non-reductionism reflects an uncritical tendency for philosophers and others to suppose that "we can 'read off' features of reality from our ways of speaking about it" (207), in a way that involves the following 'Picture Theory', whereby representation always corresponds to reality:

When a predicate applies truly to an object, it does so in virtue of designating a property possessed by that object and by every object to which the predicate truly applies (or would apply). (210)

By way of illustration of what is problematic about this principle, Heil argues that notwithstanding that the predicate 'is red' truly applies to some objects—e.g., tomatoes, stoplights, apples—it would be a mistake to uncritically suppose that this is in virtue of these objects' possessing the "*very same* property". Rather, he suggests, one may take the predicate and associated concept to here be serving the broadly pragmatic function of enabling us to categorize objects that are inexactly similar.³

²Ney herself presents such abstractionism as compatible with reductionism, but in my view this muddies the terminological waters, insofar as standardly, and in any case as I am understanding it here, ontological reductionism is the view that higher-level goings-on are identical to some or other lower-level goings-on. Similar remarks apply to Heil's view, which he sometimes presents as being compatible with a form of realism about higher-level features.

³Berkeley (1710) argues similarly against the assumption that general terms denote abstract ideas: "[I]t is thought that every name has, or ought to have, only one precise and settled signification, which inclines men to think there are certain *abstract, determinate* ideas, which constitute

3.1. OBJECTION: ANTI-REALISM ABOUT HIGHER-LEVEL FEATURES 103

The concept expressed by the predicate ‘is red’ [...] seems tailor-made for picking out a range of objects that are, in a particular way, less-than-perfectly-similar to one another. The concept applies to objects by virtue of properties possessed by those objects, presumably an extremely complex and diverse class of physical properties. There is, I gather, no prospect of defining or analyzing redness in terms of these physical properties. This is due, in some measure, to the fact that the properties in question are salient—to us—partly owing to the nature of our perceptual system. Were we built differently, were we made of different materials, the diverse collection of properties that satisfy our concept of redness could well fail to stand out. In that case we should have no use for the concept. (215)

The same holds more generally, Heil continues, for special-science predicates such as ‘is in pain’, ‘is a tree’, or ‘is a planet’. Rather than taking diverse applications of a given such predicate as indicating the existence of a common property—whether reducible or irreducible, no matter—we should see such applications as reflecting a merely conceptual means of tracking inexact similarity among lower-level physical properties.

Heil here endorses a form of anti-realism about seeming higher-level features, at least in cases of seeming multiple realizability. And his conceptualist account of cases of the sort frequently appealed to in support of the holding of the *Proper Subset of Powers Condition* (with the proper subset of powers associated with the higher-level feature being those in the intersection of the sets associated with its multiple realizers), thus constitutes an objection to the viability of Weak emergence, to the effect that the embedded condition is not in fact satisfied.

My response to the anti-realist begins by recalling the present dialectic. As argued in the first chapter, the apparently “leveled” structure of the sciences, and the associated characteristics of the special sciences as involving distinctive types and laws, provide *prima facie* support for taking there to be higher-level entities and features; so too do our perceptual and introspective experiences and practices

the true and only immediate signification of each general name. [...] whereas, in truth, there is no such thing as one precise and definite signification annexed to any general name, they all signifying indifferently a great number of particular ideas” (§18).

of individuation of macro-entities. As such, there is no quick route to anti-realism based in parsimony principles such as Ockham's razor, according to which "one should not multiply entities beyond necessity". For such principles are to be applied only "other things being equal", and an anti-realist conception of apparently higher-level entities and features would be costly, since seriously revisionary. Such a view would entail, among other things, that the motivating scientific and experiential beliefs and practices which appear to support the existence of higher-level entities and features are mainly or always in error, as false, meaningless, or (even if true and meaningful) in any case about entities and features different from those that the motivating beliefs and practices appear to be about. For example, on Heil's approach, special science laws which appear to make no reference to lower-level entities and features would turn out to be about such entities and features and their inexact similarities. Given the revisionary consequences, we need good reason to depart from the *prima facie* appearances of higher-level reality and instead embrace whatever revisionary metaphysics is put in its place; but, as I'll shortly argue, neither Ney nor Heil provide any such good reasons.

Heil's anti-realism is primarily motivated by Kim-style concerns that higher-level properties are causally excluded by their lower-level realizers, as when he says

I need not remind you of difficulties a levels conception breeds. Consider just the problem of the 'causal relevance' of higher-level properties. Suppose that mental properties are higher-level properties realized in the nervous systems of sentient creatures. How could such properties affect the behavior of creatures possessing them? The potential causal contribution of any higher-level property would seem to be preempted by its lower-level realizing property. [...] This worry about the causal relevance of mental properties extends smoothly to higher-level properties generally. If you like to think of the special sciences as occupied with higher-level properties and events, then you will need some accounting of how these properties and events could make a causal difference in our world. (Heil 2003b, 213)

In Heil's remarks, we again see operative the assumption, following Kim, that the only way for a feature to be distinctively efficacious is by having a distinctive

power. But as previously, the Weak emergentist has provided an “accounting of how [higher-level] properties and events could make a causal difference in our world”, in spite of not having any distinctive powers—namely, by having distinctive power profiles, reflecting either causal difference-making considerations and/or comparatively abstract, metaphysically real, levels of causal grain. What is needed at this point, then, is anti-realist reason to think that the strategy of locating distinctive efficacy in the having of distinctive power profiles is somehow problematic.

Ney offers such a reason, but in focusing solely on Shoemaker’s discussion, her objection misses the mark. After presenting her mental abstractionist alternative, she says, on Shoemaker’s behalf:

But maybe this is too quick. At times Shoemaker has tended to suggest that in cases of mental causation, it is only the mental event S that is efficacious with respect to the effect. In *Physical Realization* [Shoemaker 2007], he makes the slightly weaker but similar suggestion that in cases of mental causation, only the mental event is directly efficacious, with the physical or microphysical realizer events being efficacious only in virtue of containing the causal powers of the mental event as parts (53). My speculation is that the reason Shoemaker makes claims like these is in order to emphasize that he is really providing a nonreductive account of mental causation [...] The goal is to do this by securing some kind of distinctive causal efficacy for the mental, claiming that in cases of mental causation, it is only mental events (not underlying physical or microphysical events) that are efficacious, or that in cases of mental causation, it is only mental events that are directly efficacious.⁴ (443)

Ney goes on to say that, insofar as every power of a mental feature on a given occasion is supposed to be identical with those of its physical realizer, and insofar as satisfaction of this token identity condition is crucial to preventing overdetermination, “it doesn’t make sense to say that the realizers are only indirectly efficacious

⁴Morris (2011b) similarly suggests “perhaps we could take Shoemaker to be saying that [...] when a property M is subset realized by a property P on some occasion, and M and P seem to overdetermine an effect, in fact it is the M -instance that is the ‘real’ cause of that effect” (370).

vis-a-vis the efficacy of the mental event. There is only one causal relation here [...] there is no reason to try to secure any distinctive causal efficacy for mental events. Assuming we are all physicalists, the challenge is to secure nonredundant causal *efficacy* for mental events, not causal distinctiveness" (444). Hence, she concludes, Shoemaker's strategy for securing the distinctive efficacy of higher-level features, via satisfaction of the *Proper Subset of Powers Condition*, fails.

It is true that Shoemaker has several times suggested that a higher-level feature satisfying the *Proper Subset of Powers Condition* might be appropriately deemed 'the' cause of a given effect. So, for example, in his (2000/2001), Shoemaker says

[W]here only the causal features of a property *P* that play a role in producing an effect are ones that belong to a property *M*, of which *P* is a [...] realizer property, there seems a good sense in which considerations of proportionality [including sufficiency and difference-making] favor the instantiation of *M* over the instantiation of *P* as the cause of the effect. (436)

In his (2003), he confirms:

One advantage of this approach is that it provides a basis for saying that in some cases it is the instantiation of a mental property, rather than the instantiation of one of its realizer properties, that caused a certain effect, or contributed to causing it. (3)

But here, in my view, Shoemaker's discussion goes awry. I agree with Ney that the supposition that a Weakly emergent feature might be appropriately deemed 'the' cause (or a 'more direct' cause) of a given effect is problematic: she is right that it doesn't really make sense to deny that the lower-level realizer is a cause (and moreover a direct cause) of the effect, given that the supposition that the *Proper Subset of Powers Condition* is satisfied entails that the realizer has the power to produce (or contribute to producing) the effect, and this power is manifested (or the associated regularity instanced, etc.) on the occasion in question. Relatedly, as I note in Wilson 1999, 2011, and elsewhere, the supposition that a higher-level property causes something that its lower-level physical realizer does not cause

is in direct or indirect tension with physicalism, and the supposition of *Physical Causal Closure*.

Luckily, there is no need to follow Shoemaker in suggesting that realized features can cause effects that their realizing features don't cause, since what the Weak emergentist requires is not that higher-level features be *uniquely* efficacious but just that they be *distinctively* efficacious.⁵ Once it is seen that distinctive efficacy can be achieved via difference-making and/or abstract causal joint/law considerations, one can allow that both realizing and realized features are causes—distinctive causes—of a given effect: one can have one's physicalism and one's distinctive higher-level efficacy, too.

Having gotten clear about the strategy operative in Weak emergence and the associated commitments of the non-reductive physicalist, it can also be seen that Ney's discussion fails to accurately register the distinctive form of efficacy associated with the power profile resulting from satisfaction of the *Proper Subset of Powers Condition*. This distinctive efficacy is not, even for Shoemaker, a matter of a higher-level feature's being able to cause things that its realizer(s) cannot cause; rather, it is a matter of its profile's being such as to track relevant difference-making considerations, or distinctive levels of causal grain and associated laws. Relatedly, Ney is wrong to claim that for physicalists “the challenge is to secure non-redundant causal *efficacy* for mental events, not causal distinctiveness”, for ensuring the former without ensuring the latter will still fail to accommodate the pretheoretic appearances, leading to the need to drastically revise our understanding of either the truth or content of a large range of scientific and ordinary beliefs. So far, then, pragmatic/abstractionist anti-realism remains unmotivated.

⁵For similar reasons, the Weak emergentist can and should reject Yablo's (1992) claim that difference-making and other proportionality considerations can support identifying a determinable feature (e.g., red) rather than an associated determinate feature (e.g., scarlet) as *the* cause of a given effect (e.g., the pecking of a pigeon at a scarlet patch).

3.1.2 Explanatory gap eliminativism

I turn now to another motivation for anti-realism about (some) higher-level features, associated with eliminativist physicalism. Eliminativist physicalists maintain, as per *Physical Causal Closure*, that every physical effect has a purely lower-level physical cause; they moreover maintain that if we cannot explain a seemingly existent higher-level feature in physical terms—if an insuperable explanatory gap exists between the seeming feature (or associated laws and theories) and lower-level physical features (laws, theories)—then we should reject the seeming existence of the feature as genuine (see, e.g., Feyerabend 1963, Paul Churchland 1981, and Patricia Churchland 1986). Finally, eliminativist physicalists maintain that there are in-principle barriers to explaining certain mental features—in particular, qualitative (or ‘phenomenal’) features such as *seeing red* or *being in pain*—in lower-level physical terms, thus warranting the eliminativist conclusion.

Insofar as the eliminativist target has mainly been qualitative features, this objection is not generally directed against the viability of Weak emergence, since many special science features, of the sort that would be candidates for satisfying the schema for Weak emergence, are uncontroversially explainable in lower-level physical terms; and as I’ll shortly argue (in §3.2.2), such explainability, including deducibility, is compatible with Weak emergence. Even the limited inference from explanatory gaps to the elimination of some apparently higher-level features can be challenged, however, in three ways.

First and perhaps most basically, idealism aside, epistemology—what we do or can believe, know, understand—is one thing, and metaphysics—what is the case—is another; as such, it is frequently unclear whether any metaphysical import should be assigned to a seeming explanatory gap, and if so, what this import should be taken to be. As is historically familiar, metaphysical conclusions drawn from explanatory failures are often subject to later refutation, due to an explanation’s becoming available, or to the explanatory gap’s itself being explained by reference to representational or cognitive limitations having no clear bearing on the metaphysical status of the explanandum, or both.⁶ Even

⁶I’ll discuss a case-in-point of such metaphysical revisionism in Chapter 5, where explanatory-

granting that a given explanatory gap has metaphysical import, it is often unclear what this should be taken to be. It has been suggested, for example, that explanatory gaps in the cases of qualitative mental features push not towards anti-realism about such features but rather towards one or other anti-physicalist view. Hence Descartes (1641–7/1984) takes these to motivate Cartesian dualism; James (1950/1890) and Nagel (1979) argue that explanatory discontinuities motivate pan- or proto-psychism; British emergentists such as Mill (1843/1973) and Broad (1925), and more recently, van Cleve (1990), take these to motivate Strong emergence; and Nagel (1974), Jackson (1986), and Chalmers 1996 take these to motivate the rejection of physicalism, while remaining broadly neutral on exactly which anti-physicalist option should be endorsed. So there is little general support for inferences from the existence of explanatory gaps to anti-realism about the unexplained phenomenon.

Second, given our introspective access to qualitative mental features, we have more reason to believe that such features exist, one way or another, than we have to believe the (highly speculative, theoretical) premises in the eliminativists' arguments to the contrary (for arguments in the ballpark of this objection, see, e.g., Kitcher 1984 and Fodor 1987). Eliminativists might respond by telling an error-theoretic story according to which our seeming experience of qualitative mental features could be an illusion, if certain complex lower-level physical laws and circumstances were in place. But again, given our seemingly direct experience of such features, any strategy for ‘explaining away’ this experience would appear to rely on a premise that we have less reason to believe than the claim that qualitative mental features exist.

Third, the Churchlands suppose (as panpsychists, Strong emergentists, and Cartesian dualists also typically do) that, were there to be existent yet physically inexplicable higher-level features, then these would be incompatible with physicalism. This is incorrect, however; for as I'll discuss in Chapter 4, there are in-principle empirical tests of physical acceptability of a feature that could be per-

gap-based claims that chemical goings-on were Strongly emergent from physical goings-on were withdrawn upon the advent of quantum-mechanical explanations of chemical phenomena.

formed even if the feature were not, for whatever reason, explicable in lower-level physical terms. To prefigure: on my preferred way of interpreting the schema for Strong emergence, the new power associated with a Strongly emergent feature reflects the coming into play of a new fundamental interaction, and new fundamental interactions are posited in response to apparent (empirically observed) violations in conservation laws. As such, if it turned out that the occurrence of qualitative mental features did *not* give rise to any such apparent violations, that would be motivation for denying that such features were Strongly emergent, compatible with maintaining that the features existed, even assuming the features were not explicable, even in principle, in physical (or physically acceptable) terms. Hence even granting that an explanatory gap is in place between phenomenal and physical features, eliminativism is presently unmotivated.

3.2 Objection: non-satisfaction of the *Proper Subset of Powers Condition*

I'll next consider a line of objection according to which the conditions in the schema for Weak emergence are not, or in any case have not been shown to be, satisfied.



3.2.1 Failure to motivate satisfaction of the *Proper Subset of Powers Condition*

I previously argued (Ch. 2) that a representative range of non-reductive physicalist accounts are reasonably seen as aiming to implement the schema for Weak emergence, including functional realization accounts and determinable-based accounts. Morris (2011b and 2013) argues that the motivations offered for thinking that the relations at issue in these accounts fail to successfully motivate the holding of the *Proper Subset of Powers Condition*. I present and address his arguments in turn.

Functional realization and multiple realizability

As previously, functional realization accounts take higher-level features to be associated with functional roles which can be played by multiple lower-level realizers; insofar as the multiple lower-level realizers share the powers associated with the functional role, but (reflecting differences between these realizers) other powers besides, it is natural to suppose that functionally realized features will have a proper subset of the powers of each of their lower-level realizers. As Morris (2013) puts the motivations for what he calls ‘Subset Inheritance’:

The background idea [...] which may grant [Subset Inheritance] some plausibility, is the presumption that a multiply realized property is causally unified across its diverse realizations—that instances of a multiply realized property have the same powers regardless of how they are realized. (207)

He goes on to note that someone rather maintaining that instances of higher-level features inherit *all* of the token powers of their realizers, as per what he calls ‘Full Inheritance’, can accommodate a form of such unity:

But note, for one, that there is a sense in which the defender of Full Inheritance can accept the causal unity of multiply realized properties—namely, in the sense that all instances of a property *M* have a certain set of causal powers. A functionalist about pain who endorses Full Inheritance may, for example, allow that all instances of pain have the power to bring about avoidance behavior. It is just that the defender of Full Inheritance will then contend that insofar as pain is multiply realized, there may be many other powers that are possessed by some instances of pain but not others. (207)

Morris goes on to consider what motivations there might be for endorsing ‘Subset Inheritance’ as opposed to ‘Full Inheritance’, but argues that each of these motivations ultimately presupposes the falsity of Full inheritance, and so will be unlikely to convince someone who is inclined not to accept that, in particular, an instance of a property has a causal power just in case all instances of that property have that power.

Here I have two responses. The first echoes the methodological point made at the end of the first chapter. The question at issue *vis-à-vis* Weak emergence here is: can we make sense of the *prima facie* appearances of metaphysical emergence as genuine and moreover as compatible with physicalism? The dialectical burden of the Weak emergentist is not to force reductionists into accepting non-reductionism, but to establish that a positive answer can be given to the question at hand, by showing how such emergence can be intelligibly and unproblematically modelled. To push back against this result it is not enough to simply observe that there are anti-realist or reductionist strategies for accommodating the appearances; rather, one must show, somehow or other, that the Weak emergentist's model is unsuccessful. Morris doesn't do this, and it is at least reasonable, following the usual analogy to comparatively abstract functional roles which may be multiply implemented, to take multiply realizable features to be individuated (only) by their common powers, not only at the level of types but also at the level of tokens (as per Subset Inheritance and the *Proper Subset of Powers Condition*).

In any case, given the usual supposition (which Morris apparently accepts) that functional types are associated with comparatively abstract causal roles, a more direct response is available. To start, Morris's suggestion that functional realization can be understood in terms of Full Inheritance entails that instances of a given type can have token powers not associated with the type. But as previously discussed, that some token feature has powers that are not associated with a given type is good reason to think that the feature is not of that type.⁷ Coupled with the usual supposition that functional types are associated with comparatively abstract roles, this observation provides independent good reason to endorse Subset rather than Full inheritance as characteristic of functionally realized higher-level features.

⁷I will offer a similar objection to accounts on which higher-level features are taken to be token-identical but not type-identical to lower-level features; see §5.1 of this chapter.

Determinables, determinates, and pecking pigeons

As previously, a determinable-based account takes higher-level features to be determinables of determinate lower-level features, and that higher-level features would satisfy the *Proper Subset of Powers Condition* on a determinable-based account is often motivated by attention to cases of causation involving determinables and determinates along lines of Yablo's (1992) case of Sophie the pigeon, who has been trained to peck at red things (see also Wilson 1999 and Shoemaker 2000/2001). Again, if Sophie pecks at a red patch which happens also to be scarlet, then it is natural to think that she did so (as per her training) in virtue of the patch's being red, in which case the instance of red is plausibly taken to have the power to get Sophie to peck. Of course, since to be scarlet is to be red, in a specific way, the patch's being scarlet also has this power; and similarly for any other token power the instance of red might be thought to have. Now consider Alice, Sophie's picky cousin, who has been trained to peck only at scarlet things. If Alice pecks at the patch, then it is natural to think that she did so in virtue of the patch's being scarlet and *not* in virtue of the patch's being (merely) red, since if the patch had been burgundy or crimson, then while it would still have been red, Alice wouldn't have pecked. In that case it appears that the instance of scarlet has a power that the associated instance of red doesn't have. So here is one case where features appear to satisfy the *Proper Subset of Powers Condition*; similar considerations would appear to be available to support thinking that determinable instances and the determinate instances upon which they depend, on a given occasion, always satisfy this condition. Yet more generally, one might suppose, either on grounds that realization just is the determinable-determinate relation or on grounds of structural similarities between the latter relation and (other cases of) realization,⁸ that cases of non-reductive realization generally satisfy the

⁸In particular, one might naturally suppose that a feature's satisfying the *Proper Subset of Powers Condition* reflects that being comparatively ontologically unspecific (as determinables are supposed to be, *vis-à-vis* their determinates) goes hand-in-hand with being comparatively causally unspecific; and insofar as many seemingly higher-level features are supposed to be comparatively ontologically unspecific *vis-à-vis* lower-level physical realizers, the case would provide a generalizable model and associated motivation for thinking of higher-level features as satisfying the

Proper Subset of Powers condition.

Morris (2013) objects to this case-based line of motivation for taking intuitively realized instances to have fewer token powers than those of their realizing features. He starts by noting that if what powers are associated with a given instance of a feature is a matter of what the feature can cause, and if (as both Yablo 1992 and Shoemaker 2000/2001 suggest, by appeal to proportionality considerations) the patch's being red is the best candidate for being the cause of Alice's pecking and the patch's being scarlet is the best candidate for being the cause of Sophie's pecking, then we have symmetry here, with each feature having at least one power that the other doesn't have, contra the supposed satisfaction of the *Proper Subset of Powers Condition*.

The assumption underlying Morris's symmetry concern is similar to that which played a role in Ney's criticism of Shoemaker's powers-based account of realization—namely, that in the cases at hand, the assignment of a power to a given (instance of a) feature depends on taking the feature to be *the cause* of the effect in question: *red* but not *scarlet* causes Sophie to peck, while *scarlet* but not *red* causes Alice to peck. Morris is correct that this reading of the cases is problematic, not just in that (as I noted previously) taking a higher-level feature to be ‘the’ cause of a given effect is in clear tension with *Physical Causal Closure* and hence with physicalism, but also because if powers track causes—as they do—then the supposed satisfaction of the *Proper Subset of Powers Condition* in these cases is indeed undermined.

The assumption at issue—namely, that the assignment of a power to a given (instance of) higher-level feature depends on taking the feature to be *the cause* of the effect in question—can and should be rejected, however, for one can maintain that the cases are intended to illuminate certain facts about causation (rather than explanatory relevance or some other notion) without taking on board this assumption.⁹ On the contrary, the whole point (in my view) of attention to the

condition.

⁹Relatedly, I reject Morris's (2011b) claim that in order to avoid problematic overdetermination, “What the subset theorist needs is the claim that the manifestation of a single power on an occasion entails that there is only a single cause of an event, rather than multiple causes” (372).

determinable-determinate relation as a model in this case is as showing how *each* feature can be a (distinctive) cause of the effect in question. Again, and *pace* the presentation in Yablo and Shoemaker, gaining the causal autonomy of higher-level features does not require that these be univocally efficacious, but only that these be distinctively efficacious.¹⁰

More generally, the Weak emergentist can and should reject the assumptions generating Morris's symmetry concern—namely, that powers can be assigned to (instances of) higher-level features only if these can be made out to be ‘the’ causes of certain effects. Having done so, the Weak emergentist can continue to maintain, following my reading of the pigeon cases in particular, that these cases are reasonably seen as modeling determinable-based satisfaction of the *Proper Subset of Powers Condition*.

3.2.2 Token multiple realizability as a barrier to satisfaction of the condition

Pereboom (2002) argues that token mental states S are typically multiply realizable, and that such a possibility is incompatible with taking the powers of S to satisfy the *Proper Subset of Powers Condition*. Though his focus is on mental states, the considerations he raises, were they to go through, would seem to more generally apply to block the satisfaction of the schema for Weak emergence for many, most, or even all of the seemingly higher-level features of the sort treated by the special sciences.

Pereboom first argues that if a token higher-level state S is multiply realized,

¹⁰It is also worth noting that what is taken to cause what in these cases is not solely a product of difference-making or proportionality considerations. Indeed, that *red* is a cause of Sophie’s pecking follows from her being trained to peck at red patches, and that *scarlet* is also a cause of Sophie’s pecking reflects further metaphysical assumptions about *scarlet*—namely, that to be scarlet is to be red, in a specific way. Proportionality considerations primarily come into play just in rendering plausible that *red* is a *distinctive* cause of Sophie’s pecking. Similarly, even granting that proportionality considerations are sometimes used to support taking higher-level (e.g., mental) features to be efficacious *vis-à-vis* some effect, that their lower-level physical features are also efficacious *vis-à-vis* these effects (if physicalism is correct) follows from *Physical Causal Closure*, not attention to proportionality.

S cannot be identical with a token base state *P*:

Suppose that *S* is realized by a complex neural state [*P*]. It is possible for *S* to be realized differently only in that a few neural pathways are used that are token distinct from those actually engaged. [...] [I]t is evident that this alternative neural realization is itself realized by a microphysical state *P'* that is token distinct from *P*. It is therefore possible for *S* to be realized by a microphysical state not identical with *P*, and thus *S* is not identical with *P*. (503)

Pereboom continues:

[T]his reflection would also undermine a token-identity claim for mental causal powers—should they exist—and their underlying microphysical causal powers. For if the token microphysical realization of *S* had been different, its token microphysical causal powers would also have been different. We therefore have good reason to suppose that any token mental causal powers of *S* would not be identical with the token microphysical causal powers of its realization. (503)

Again, since these considerations would, if compelling, more generally apply to any multiply token-realizable special science features, Pereboom's claim might be seen as constituting a case against taking Weak emergence to be core and crucial to physically acceptable emergence, on grounds that the *Proper Subset of Powers Condition* in that schema is rarely, if ever, satisfied. The Weak emergentist can, however, respond in either of two ways.

First, they can deny that tokens of higher-level features are ever multiply realizable. What is (fairly) uncontroversially true is that *types* of higher-level features are multiply realizable, in the sense that different tokens of the realized type can be realized by different tokens of the lower-level type. Pereboom's case for taking token higher-level feature *S* to be multiply realizable isn't compelling, and may be rejected. Compare: could that very instance of red, currently realized by an instance of scarlet, have been realized by an instance of burgundy? One might reasonably judge not, and continue to reasonably judge not even supposing the alternative shade to be only "slightly" different from the original. One may similarly reasonably deny that *S* (as opposed to another token of *S*'s type) could be

realized by a base feature other from P , whether this alternative feature be very different from P , or only different with respect to “a few neural pathways” or other lower-level physical details.

Second, the Weak emergentist can allow that a token feature S may be multiply realizable, but maintain that S ’s token powers are relativized to realizers (or occasions). Pereboom presupposes that S has its token powers essentially; but why think this? As he later observes, “stable tokens [...] often retain their identity over certain changes in their constitutions and configurations” (529). One might similarly maintain that token feature S can retain its identity across changes in its realizers and associated token causal powers. The *Proper Subset of Powers Condition* requires only that S ’s token powers on a given occasion be numerically identical with the powers of the lower-level physical feature realizing it on that occasion; hence the *Proper Subset of Powers Condition* can be satisfied even if S is token multiply realizable.

3.3 Objection: compatibility with reductionism

In this section, I consider objections that target the sufficiency of Weak emergence on grounds that even granting that the conditions in the schema are satisfied by S and lower-level physically acceptable P , such satisfaction is compatible with S ’s being ontologically or metaphysically reducible to—that is, identical with—some other lower-level physically acceptable feature P' . The line of thought is as follows. Let the level of physically acceptable base property P be the fundamental physical level—the only level the reductionist thinks exists. To be sure, if S is Weakly emergent from P on a given occasion, then S has, on that occasion, a proper subset of the token powers of P . Still, the reductionist maintains, given that S satisfies the *Proper Subset of Powers Condition*, surely S must be reducible to some *other* lower-level physically acceptable feature P' . Whether both P and P' should be considered ‘realizers’ of S is a topic for further debate.

There are three different strategies for accomplishing the proposed reduction of a feature S : one according to which P' is a conjunctive feature having S as

a conjunct; one according to which S is a disjunctive feature having P' as a disjunct; and one according to which S , if it is to be a lawful metaphysical consequence of lower-level physical goings-on (as physicalists uniformly assume), must be reducible to some or other lower-level physical goings-on P' , even if we are not in position to characterize such a P' in any detail. Recalling that the resources typically given the reductionist are circumscribed by attention to how levels are individuated, one can see the first two reductive strategies as drawing on the ‘lightweight combination’ approach to levels, according to which conjunctions and disjunctions of lower-level features are also lower-level physical features. The third reductive strategy might be seen as drawing on a coarse-grained conception of the ‘law-consequence’ approach, according to which any entities or features that are consequences of laws at a level L are also appropriately placed at that level. I address each strategy, in turn.

3.3.1 Reduction to a conjunct of a lower-level conjunction?

On the first reductive strategy, satisfaction of the *Proper Subset of Powers Condition* is compatible with seemingly higher-level feature S ’s being a conjunct of a lower-level conjunctive feature P' , such that S , though different from P , would nonetheless also be appropriately considered a lower-level physical feature. Indeed, one might reasonably suppose that a conjunct feature satisfies the *Proper Subset of Powers Condition* *vis-á-vis* associated conjunctions, insofar as conjunctive features, being more specific, can contribute to producing more effects (in the same circumstances) than their conjunct features can individually do, for reasons similar to those motivating taking determinates to have more powers than their associated determinables. So, for example, if P is a conjunctive feature consisting in *being massy* and *being charged*, then, one might think, P has powers to produce effects that *being massy* alone does not have (e.g., attract pins) and that *being charged* alone does not have (e.g., attract other massy objects). But if P is a conjunctive lower-level physical feature having S as a conjunct, then plausibly, S will

also be a lower-level physical feature, and so reducible rather than emergent.¹¹

That conjuncts of conjunctions appear to satisfy the *Proper Subset of Powers Condition* is especially pressing for Shoemaker, who characterizes realization just in terms of satisfaction of the *Proper Subset of Powers Condition*. It is also a difficulty for accounts of Weak emergence involving functional realization, since if the higher-level functional role is associated with a lower-level conjunct feature, one might naturally suppose that any conjunction containing that conjunct will count as implementing or realizing the role.

Shoemaker is sensitive to this line of thought, and in order to rule out that lower-level conjuncts of lower-level conjunctions ever count as emergent, stipulates that for a feature to count as realized, it must not be a conjunct of a conjunctive realizer:

Property P realizes property S just in case the conditional powers bestowed by S are a subset of the conditional powers bestowed by P (and P is not a conjunctive property having S as a conjunct).¹² (78)

The need to stipulatively rule out conjunction-conjunct cases as genuine cases of Weak emergence is perhaps a bullet one could bite. There are, however, two non-stipulative strategies for ways to rule out such cases.

The first strategy, due originally to [Baysan \(2014\)](#), is compatible with any implementation of Weak emergence. Here one rules out conjunct/conjunction cases as candidates for Weak emergence by imposing a further condition, or (as I would put it, in the context of the schema) by saying more about the intended notion of synchronic material dependence operative in condition (ii) of the schema for Weak emergence. However exactly the notion of dependence is cashed out, in a context where physicalism is at issue this notion should be one according to which synchronically materially dependent entities and features are less fundamental than dependence base entities and features. But on the operative understanding of the

¹¹Note that the concern here remains even if, as [Gibb \(2013\)](#) compellingly argues, it is not always the case that conjunctive features have more token powers than each of their conjunct features.

¹²Here I have replaced Shoemaker's variables with S and P for consistency with my discussion.

physical entities and features as compositionally basic, conjunct features are plausibly more, not less, fundamental than associated conjunctive features. As such, a conjunct feature S would not be appropriately taken to satisfy the relevant criterion of dependence in the schema for Weak emergence, notwithstanding that a conjunctive feature (at least nomologically) necessitates its conjuncts; hence such features pose no threat to the sufficiency of this schema, properly interpreted. As Baysan puts it in an abstract:

[A] property P realizes a property Q if and only if the causal powers of Q are a proper subset of the causal powers of P , and P is more fundamental than Q . Thanks to the requirement that a realized property is less fundamental than its realizers, two things that the original version of the subset view cannot explain are guaranteed: first, fundamental properties are not realized; second, arbitrary conjunctions of properties do not realize their conjuncts.

I like the idea of putting some requirements on the dependence relation at issue, since the notion of synchronic material dependence at issue in condition (i) of the schema for Weak emergence, even understood (as per usual) as involving minimal nomological supervenience, is as it stands fairly weak. Some—notably, Shoemaker, in his 2000/2001, 2007, and elsewhere, have offered the *Proper Subset of Powers Condition* as doing double duty in ensuring, by itself, satisfaction of the relevant dependence condition. But this supposition is dispensable: nothing prevents Shoemaker or any other proponent of Weak emergence from offering an independently plausible characterization of the synchronic material dependence at issue, with the *Proper Subset of Powers Condition* serving mainly to ensure the autonomy that is characteristic of emergence. It would, after all, be unsurprising that two conditions would be needed to characterize metaphysical emergence of whatever variety: dependence is one thing, autonomy another.

Second, one might appeal to an account of Weak emergence on which conjunct/conjunction cases are non-stipulatively excluded as satisfying the *Proper Subset of Powers Condition* in the schema. This would be the case on a determinable-based account of Weak emergence, on which lower-level features are determinates of higher-level determinables; for it is definitive of this relation that it is

not properly metaphysically characterized in terms of anything like the conjunct-conjunction (or relatedly, genus-species) relations (see Wilson 2017 for historical and other discussion). In mereological terms: determinates, unlike classical wholes, do not satisfy ‘weak supplementation’, according to which a whole having one proper part must also have at least one other proper part disjoint from the first. In Karen Bennett’s memorable terms (p.c.): determinates are not determinables with frosting on top. As such, on a determinable-based implementation of Weak emergence, the nature of determinate features alone non-stipulatively rules out that conjuncts of lower-level conjunctions would count as Weakly emergent.

An implementation of Weak emergence in terms of an elimination in degrees of freedom might also non-stipulatively rule out conjunct/conjunction cases, if no features satisfying the condition on degrees of freedom at issue in this account also stand in the conjunction/conjunct relation. Since which degrees of freedom are associated with which features is an empirical matter, the exclusion here would be contingent as opposed to following (as in the case of the determinable/determinate relation) just from features of the specific relation taken to satisfy the *Proper Subset of Powers Condition* itself; but even so it would not be stipulative.

Summing up: that conjuncts can satisfy the *Proper Subset of Powers Condition* vis-á-vis associated conjunctions poses a *prima facie* problem for taking satisfaction of the conditions in Weak emergence to be sufficient for physically acceptable emergence. The concern can be addressed in a number of ways, however, either by stipulating the exclusion of such cases (a less than optimal but non-fatal addendum), by denying that conjunct features properly satisfy the synchronic material dependence condition, since it is not plausible to take a presumed fundamental conjunction to have non-fundamental conjuncts (a reasonable addendum in cases where the dependence at issue is supposed to conform to physicalism), or by endorsing an instantiation of Weak emergence (along determinable-based and perhaps also DOF-based lines) that non-stipulatively rules out conjunct/conjunction cases as cases of such emergence.

3.3.2 Reduction to a disjunction of lower-level disjuncts?

The second reductive strategy takes as its starting point an account on which higher-level features are identical with disjunctions of lower-level physically acceptable features. The strategy is commonly motivated by a seeming objection to reductionism, according to which the multiple realizability of a given seemingly higher-level type rules out the realized type's being identical with any one of its realizing types. As [Antony \(2003\)](#) expresses the objection:

[M]ultiple realizability has *ontological* consequences. Clearly, a property P cannot be identical with a property Q if there can be instances of P that are not instances of Q . But to say that a property S is multiply realizable [by types P_1 and P_2] is to imply that [...] there can be instances of S that are not instances of P_1 , and instances of S that are not instances of P_2 , and S cannot be identical with either P_1 or P_2 .
(3)

A popular reductionist response implements the ‘disjunctive strategy’ or ‘disjunctive move’ (first anticipated in [Fodor 1974](#); see [Jaworski 2002](#) and [Dosanjh 2014](#) for discussion), according to which multiply realizable types may be identified with a lower-level physically acceptable *disjunctive* type, where each disjunct is a type of lower-level realizer of S ’s type. As [Heil \(1992\)](#) presents the position (without endorsing it):

Multiple realizability [...] need not deter a determined identity theorist [that is, reductionist]. It is open for such a theorist, for instance, to argue that the relevant [...] characteristic is, in fact, disjunctive in character. That is, it might be that, in you, mental feature S is realized in neural structure N , whereas in an octopus, S is realized in a different sort of neural structure N' . Would this undermine type identity? It would not, unless we assume that S [could] not be identical with the disjunctive characteristic $\langle N \vee N' \rangle$. (64)

Now, a number of objections have been raised against the disjunctive strategy.¹³ But even putting these concerns aside, there is as yet no clear concern for

¹³These include that disjunctive features don’t exist (as per [Armstrong 1978](#), 1923), that dis-

the sufficiency of Weak emergence here; for on the usual understanding of what it is for a disjunctive type to be instanced or tokened, the disjunctive strategy is incompatible with satisfaction of the *Proper Subset of Powers Condition*. On the usual understanding, what it is for a disjunctive feature type D to be tokened on a given occasion is for one of the disjunct types to be tokened on that occasion. In the case at hand, the disjunct types (P_1, P_2, \dots, P_i) correspond to the physically acceptable realizers of S 's type. It follows that if S 's type is identified with disjunctive type D , then any token of S 's type would, on a given occasion, be identical with a token of one of its realizer types, on that occasion. In that case, however, the *Proper Subset of Powers Condition* would fail to be met: S 's token powers would be, on any given occasion, the same as, rather than a proper subset of, the token powers of the lower-level physically acceptable feature realizing S on that occasion. But by assumption, S satisfies the condition. It follows that on the assumption that S satisfies this condition, as Weak emergence requires, S 's type cannot be identified with a type consisting of a disjunction of S 's lower-level realizers.¹⁴

Granting that the *Proper Subset of Powers Condition* is met, a disjunctivist gambit remains. Drawing in part on discussion in Clapp 2001 and Antony 2003, Dosanjh (2014) argues that it is not generally the case that what it is for a disjunctive property to be instanced is for one of its constituent disjunct properties

junctive features are not available for purposes of physicalist reduction of scientific properties, because they are too heterogenous to form a natural kind (as per Putnam 1967, Fodor 1974, and Kim 1992b), that disjunctions are “open-ended” (having an indefinite or infinite number of disjuncts), rendering them unsuitable to enter into type-identities (as per LePore and Loewer 1987 and Pereboom and Kornblith 1991), and that among the metaphysically possible realizers of a multiply realizable feature type such as S will be some physically unacceptable types (as discussed, though not endorsed, by Dosanjh 2014).

¹⁴One might think that the disjunctive strategy in any case highlights a potential difficulty with both functional role and determinable-based implementations of Weak emergence—namely, that while these accounts are typically forwarded in service of a non-reductionist account of such realization, at the same time it is common to give functional features an analysis in terms of disjunctions of those features playing the functional role, and to give determinable features an analysis in terms of disjunctions of their associated determinate features. If those accounts are successful, then the associated implementations of Weak emergence would not, in fact, make sense. As argued in the previous chapter, however, there are cases to be made that functional role or determinable features should not be analyzed in or reduced to such disjunctive terms.

to be instanced. To be sure, Dosanjh allows, this is the case when the disjuncts of disjunctive properties are gerrymandered or otherwise dissimilar (as, e.g., the property of being red or round, or of being sour or prime). But in some cases, he maintains, the instantiation of the disjunction is *not* the same as the instantiation of one or other disjunct. In particular, when the disjuncts are relevantly similar, in sharing powers associated with the seemingly higher-level feature, and when the disjunction contains all and only such disjuncts (that is, when the disjunctive property has disjuncts that “exhaustively overlap”), then, Dosanjh maintains, there is a case to be made that the powers of the disjunction should be seen as a proper subset of the powers of whichever disjunct is instanced. And this proper subset relation between powers will, as per usual, be inherited by the tokens of the disjunction and the disjunct types. So, Dosanjh suggests, there is no clear barrier to identifying the type of an apparently higher-level feature S with the disjunction of its realizer types, since tokens of both types will satisfy the *Proper Subset of Powers Condition*.

My initial response is to deny that this gambit is available to the reductionist.

First, whether or not the disjuncts of a disjunct are relevantly similar, there is a case to be made that disjunctive features do not have a proper subset of the powers of their disjuncts. Consider the powers of $P \vee Q$ to produce effects when in circumstances C (restrict attention to these, for simplicity). For the cases of multiple realization at issue, we can without loss of generality assume that different realizers P and Q cannot be co-instantiated, in which case there will be two ways for $P \vee Q$ to be instanced, and so (at least) two powers associated with C :

1. If in $C \wedge P \wedge \neg Q$, then E_1 ; and
2. If in $C \wedge Q \wedge \neg P$, then E_2 .

(There may well be others, but that won’t matter for making the point.) What powers will P have, in C ? It will have at least one of the powers of $P \vee Q$, in C —namely,

If in $C \wedge P \wedge \neg Q$, then E_1 .

(Here the conjunct P is redundant, but no matter.) However, P will not have another of the powers of $P \vee Q$, in C —namely,

If in $C \wedge Q \wedge \neg P$, then E_2 .

P would have such a power only if it could be both instanced and not instanced in C at the same time (powers being relativized to times or temporal intervals), which it can't. So in this case (and more generally) $P \vee Q$ does not have a proper subset of the powers of P , or any other of its disjuncts.

Second, the disjunctive strategy is most naturally implemented against the assumption that the reductionist can help themselves to any ‘ontologically lightweight’ combinations of characteristically physical entities and features, including—all parties agree—any combinations resulting from boolean operations (again, with the possible exception of negations). Hence it is that, were it possible to identify a seemingly higher-level feature S with a disjunction of lower-level physical features, that would suffice for S ’s being ontologically reducible to—that is, identical with—some lower-level physical feature. As Dosanjh (2014) himself notes, “If there are any ontologically innocent combinations of properties, boolean logical combinations are among them. For an important example: when we talk about a disjunction of properties, we are committed to little beyond the properties that serve as the disjuncts” (17). But if the disjunctive property D with which S is identified is *not* such that its instantiation consists just in the instantiation of its disjuncts, then resources for constructing D have gone beyond mere boolean combination, and it is no longer clear that D is appropriately taken to be a lower-level physically acceptable property. On the contrary: to the extent that D , like S , is supposed to satisfy the *Proper Subset of Powers Condition* vis-à-vis S ’s lower-level realizer, one might rather suppose that D (hence S) is a higher-level feature, contra reductionism.¹⁵

¹⁵Dosanjh recognizes the concern here, and in response, attempts to undermine the reasons for believing that higher-level properties are not ontologically innocent, and in particular attempts to “problematisize the claim that higher-level properties are causally autonomous with respect to their realizers” (69), by arguing that neither difference-making nor law-based considerations support the causal autonomy of higher-level features.

3.3.3 Reduction to a metaphysical consequence of lower-level laws?

The third strategy for reductively treating a feature S satisfying the conditions for Weak emergence adverts to the metaphysical consequences of the laws governing entities at the level of S 's base entity—that is, to the fundamental physical laws governing entities at the one level that the reductionist thinks exists.¹⁶ Here the reductionist starts by observing that the Weak emergentist is committed to taking higher-level features and laws to be metaphysical consequences of—indeed, perhaps even to be theoretically deducible or predictable from—lower-level physical features and laws—after all, it is this which ensures that S is physically acceptable. But in that case, the reductionist continues, what prevents S , even with its reduced set of powers, from itself being a lower-level physical feature?

Variations on the theme of this concern are common. Consider these remarks by Klee (1984), directed against an account of purportedly physically acceptable emergence on which emergent entities and laws simply involve new relational structures:

[I]n what sense are these new regularities emergent? To be sure, they may be regularities and structures of a type not found on lower-levels

Dosanjh's strategy for showing that difference-making considerations need not be seen as supporting the causal autonomy of a seeming higher-level feature is complex and a full treatment would take us too far afield. For present purposes, however, it suffices to note that his strategy requires that true counterfactual claims of the form “had the realizer [P] not been instantiated, the effect would still have been instantiated” (69) be interpreted as involving “a conditional with a disjunctive antecedent” and second, that “the defining properties of the effect can also be identified with exhaustively overlapping disjunctions” (69). Such revisionary reinterpretations of the counterfactuals at issue are, however, both implausible and deniable.

Dosanjh's strategy for showing that attention to distinctive systems of law and associated comparatively abstract levels of causal grain need not be seen as supporting higher-level causal autonomy is to “deny that such laws are distinct enough to ground causal autonomy”, insofar as “every law statement about a property that satisfies [the *Proper Subset of Powers Condition*] is entailed by law statements about physical properties” (73). This strategy is of a piece with, and admits of the same response to, the more general concern that lawful metaphysical consequences of lower-level physical laws should be considered reducible to lower-level goings-on, to be next considered.

¹⁶Recall that the qualifier ‘metaphysical’ when applied to consequences of laws is intended to sidestep concerns about the lower-level laws containing vocabulary local to the special sciences: the notions of ‘consequence’ or ‘deducibility’ here are metaphysical, not representational.

of organization, but it has seemed to some (Nagel 1961, 367–74) that this fact by itself would not justify the label of ‘emergent’ if they had been predictable on the basis of a thorough understanding of those lower-levels of organization. If the new relational structure which grounds the new regularities could have been predicted on such a basis, then the new regularities could have been predicted and the force of any emergence claim, at least partially, compromised. (46)

Indeed, it might seem practically definitional that (in-principle) theoretical deducibility entails ontological reducibility. As Owens (1989) put it:

Reductionism is sometimes expressed as the thesis that the laws of the non-physical sciences can be deduced from those of the physical sciences together with certain bridging generalizations [...]. (63)

To be sure, some higher-level goings-on—in particular, certain complex systems of the sort we will discuss in Chapter 5—might be thought not to be even in-principle deducible from lower-level goings-on, at least given a wide but empirically restricted purview of the available resources; but the deeper concern here remains even so. For so long as the non-reductive physicalist maintains that the higher-level goings-on are metaphysical consequences of the lower-level goings-on, as they are committed to doing, the concern remains that metaphysically, if not representationally, the higher-level goings-on must be reducible to lower-level goings-on, after all.

My response here takes as its starting point the response previously given (in Chapter 1, ‘Preliminaries’) to the concern that a law-consequence account of what entities and features are properly located at a level L might rule the possibility of Weak emergence out of court. Recall, first, that attention to the role played by degrees of freedom (DOF)—independent parameters needed to specify an entity’s law-governed properties and behaviours—in characterizing the laws operative at a given level indicates that we must distinguish between two sorts of law-based “consequences”. It is correct that non-reductive physicalists will allow, *qua* physicalists, that all special science entities and features are metaphysical consequences

of the laws governing lower-level physical goings-on. However, they can reasonably deny that as a result, special science entities and features are thereby themselves lower-level physical entities or features. As previously, and as is reflected in the physical laws, the specification of the law-governed properties and behaviours of lower-level physical entities and features requires all the information needed for the lower-level physical laws to operate, including quantum-mechanical DOF and associated values—e.g., spin and quark colour charge. By way of contrast, the specification of the law-governed properties and behaviours of special science entities and features does not require specification of quantum-mechanical DOF such as spin or colour charge: such DOF are eliminated as unneeded to characterize higher-level goings-on.

This elimination in quantum DOF explains, in part, why special science entities and features are insensitive to certain micro-level details, in a way that makes room (for example) for their being multiply realizable: higher-level macro-entities are typically insensitive to spin-theoretic details, among other quantum features. More importantly for present purposes, that higher-level goings-on are insensitive to quantum-level details, with the resulting loss in quantum-level information, means that, even though special-science entities and features are metaphysical consequences of physical laws, it is not appropriate to place these entities and features at the physical level: the quantum laws wouldn't know what to do with them!

As such, one must distinguish two sorts of metaphysical consequences of the lower-level physical laws. First are the entities, features and laws which are consequences in the broadest sense—which, if physicalism is correct, will include any and all special science entities, features and laws. Second are those consequences which retain all the DOF and associated information (pertaining to spin, color charge, etc.) needed for the lower-level physical laws to operate. If there is Weak emergence, then some entities and features which are metaphysical consequences of the physical laws in the first sense will not be metaphysical consequences in the second sense. Such entities and features will have specifications that fail to include all the DOF and associated information needed for the lower-level physical

laws to operate, and so these entities and features will not be appropriately seen as identical with any lower-level physical goings-on—as the non-reductionist maintains.

This line of thought—that laws require certain kinds of information in order to operate, and that among the entities and features that are consequences of level- L laws, only those that preserve the information needed for the level- L -laws to operate are appropriately placed at L —is explicitly encoded in a DOF-based implementation of Weak emergence (see Wilson 2010b), and it is some advantage of a DOF-based account that it clearly has the resources to appeal to this objection. That said, there does not appear to be any barrier to proponents of other accounts of Weak emergence appealing to the aforementioned general scientific facts about laws and associated degrees of freedom by way of responding to the threat of law-consequence reducibility.

3.4 Objection: compatibility with physical unacceptability

Another general line of objection to the sufficiency of Weak emergence is that a feature S satisfying the conditions in this schema might nonetheless be, one way or another, ‘over and above’ its dependence base feature P , and so physically unacceptable. There are at least four strategies for developing this concern. One common strategy aims to show that features have aspects going beyond their powers, including primitive identities (‘quiddities’), phenomenal aspects, historical or spatiotemporal features, or ‘backwards-facing powers’ (registering all the different ways in which the feature can be caused), which are supposed not to be fully incorporated in the ‘forward-facing’ powers at issue in the *Proper Subset of Powers Condition*. Another strategy appeals to a case involving a conjunctive feature that has a proper subset of the token powers of one of its conjunct features, and so satisfies the *Proper Subset of Powers Condition*, but where the conjunctive feature is intuitively not realized by the conjunct feature having the superset of powers. Another strategy raises the possibility that a physically unacceptable

process is responsible for making it the case that some proper subsets of powers, rather than others, correspond to a genuine feature. A final strategy aims to show that satisfaction of the condition is compatible with a form of pan-psychism. In what follows I present and address these concerns, in turn.

Before presenting and treating these objections, a clarification is called for. Several of these objections are pitched as follows: first, it is claimed that if a feature S that synchronically materially depends on a lower-level physical feature P is to be itself physically acceptable, then features of P 's type must (perhaps along with a specification of the operative physical laws and relevant circumstances) metaphysically necessitate or entail (where the entailment here is metaphysical, not conceptual) features of S 's type. Hence Melnyk (2006, 141-143) supposes that physical realization requires that “the physical properties of an object (perhaps together with other physical conditions, including physical laws) necessitate in the strongest sense the object's non-physical properties” (138); Morris (2010) maintains that the primary physicalist constraint on a theory of realization is that it should non-trivially imply that instances of physically realized properties are necessitated, in a modally strong sense, by how things are physically; and Gibb (2013) maintains that while “the notion of nothing over and aboveness is notoriously unclear”, nonetheless “regardless of how one interprets it, presumably for the having of property X to be ‘nothing over and above’ the having of property Y , having Y must at least entail having X (553); see also Walter (2010, 211), LePore and Loewer (1989, 179), Clapp (2001, 112–113), and others.

The clarification involves observing that there is a sense in which the entailment requirement is appropriate, and a sense in which such a requirement would be too strong (such that its failure wouldn't count against the schema for Weak emergence). The sense in which the requirement is appropriate is one closer to Melnyk's expression of the requirement, as specifying that realized (Weakly emergent) features are necessitated “in the strongest sense” by base properties “together with other physical conditions, including physical laws”. This caveat is required, since as some believe, properties can enter into different laws and thus have different powers; but whether a property P is associated with property

S in worlds with different laws of nature is neither here nor there for purposes of motivating *S*'s being realized by *P* in worlds with laws relevantly similar to ours.

In any case, as we'll see, concerns about *S*'s physical acceptability can be raised without proceeding via a requirement on metaphysically necessary correlations (supposed to hold even in worlds with different laws of nature) that the Weak emergentist can reasonably deny need to be in place. In what follows I will aim, in presenting such objections, to extract the underlying deeper concerns about physical acceptability that the objections aim to identify.

3.4.1 Quiddities

Melnyk (2006, 141-143) supposes that for a feature *S* that synchronically depends on some lower-level physical feature *P* to be physically acceptable, *S* must satisfy not only the aforementioned "necessitation" condition, but also the "constitution" and "truthmaking", conditions, according to which realized features are constituted by realizing features, and truths about realized features are made true by truths about realizing features. Properly restricted to worlds with physical laws of nature similar to those actually governing *P*, Melnyk's suppositions are plausible, since it is commonly supposed that *S*'s physical acceptability requires that *S* be necessitated by *P* in worlds with (only) physical laws. And although what constitution comes to is disputed, the idea is a clear variant on the 'nothing over and above' theme, and the truthmaking condition might be thought to follow from the general physicalist commitment to the lower-level physical goings-on providing a basis, hence a truthmaking basis, for all else.

Does satisfaction of the schema for Weak emergence satisfy these conditions on physical acceptability? Melnyk is willing to grant that the answer is 'yes', if one endorses a 'causal' account of properties along lines of that endorsed by Shoemaker (1980) and (1998), on which features are essentially and exhaustively constituted by the powers they have or bestow. The resulting version of a powers-based account of realization, Melnyk observes, "has the important virtue of meeting the necessitation, constitution, and truthmaking conditions" (144), with "[t]he key move [being] to identify property-instances with something like clusters of

causal power-tokens of particular types [...]” (140).

Granting for the moment Melnyk’s supposition that implementing the schema for Weak emergence requires endorsement of a causal account of properties (features), the problem remains, in Melnyk’s view, that for some properties, “it’s implausible to identify their instances with clusters of causal power-tokens”.¹⁷ The most pressing case of those Melnyk considers has to do with properties whose individuation involves some sort of primitive identity or ‘quiddity’—effectively, the property equivalent of a haecceity, or primitive identity, serving to individuate objects or other particulars.¹⁸ By way of illustration, Melnyk refers to Hawthorne’s (2001) case of properties which are intuitively distinct but which play the same nomic role (perhaps positive and negative charge are actual such properties). If *S* has a non-causal quiddity, then, it seems, *S*’s satisfaction of the *Proper Subset of Powers Condition* won’t guarantee that *S*’s quiddity is constituted by or otherwise ‘nothing over and above’ *P* (or *P*’s quiddity, as the case may be), or that truths about *S* are made true by truths about *P*, or even that instances of *P*’s type (physically) necessitate instances of *S*’s type. As such, Melnyk suggests, the possibility of non-causal quiddities poses a dilemma for the Weak emergentist: either properties may have non-causal quiddities, in which case satisfaction of the *Proper Subset of Powers Condition* doesn’t ensure satisfaction of the constitution, truthmaking, and necessitation conditions as required for *S*’s physical acceptability; or quiddities are rejected and a causal account of properties maintained, which ensures *S*’s physical acceptability, but only at the price of endorsing a controversial and (given the seeming possibility of Hawthorne-style cases) not obviously satisfactory account of properties (more generally, features).

The Weak emergentist can sidestep this dilemma, as follows (see Wilson 2011).

¹⁷A similar concern applies to variations on the theme of Shoemaker’s account, associated with views on which properties and laws are essentially intertwined (as per Swoyer 1982 and Bird 2001 and 2007), and which would also (at least provisionally) entail that satisfaction the conditions in Weak emergence would ensure satisfaction of the constitution, truthmaking, and necessitation conditions.

¹⁸Effectively, a haecceity makes room for the identity and individuation of an entity to float free of any of the entity’s features, and a quiddity makes room for the identity and individuation of a feature to float free from any of the feature’s powers.

3.4. OBJECTION: COMPATIBILITY WITH PHYSICAL UNACCEPTABILITY 133

To start, they can observe that the individuation of scientific features is neutral on the presence or absence of quiddities: in scientific contexts, the occurrence of scientific features, and any truths about such features, does not depend on or otherwise track whether such features have quiddities, much less track how the non-causal quiddities of seemingly distinct features are related. This is true, in particular, for properties such as positive and negative charge, which at some level of abstraction play the same causal role. Indeed, that there is more than one charge property reflects global considerations pertaining to the structure of the laws as requiring that there be two or more distinct properties playing what is in some sense the ‘same’ role; the posit of primitive quiddities plays no role in this story. As such, the Weak emergentist can reasonably maintain that whether S and/or P have quiddities, shared or not, is irrelevant to whether or not S satisfies the constitution, truthmaking, or necessitation conditions conditions, and more generally, is irrelevant to whether or not S is physically acceptable.

This response does not require endorsing a causal theory of properties; and indeed, it is reasonably endorsed even by those taking properties to have non-causal quiddities. For the primary reason for endorsing property (feature, etc.) quiddities concerns not a supposed need to distinguish properties with shared roles at a world but rather a supposed need to identify or individuate properties in worlds with different laws of nature, on the assumption that properties may have completely different powers (see, e.g., Lewis 1983a and Schaffer 2004). But if properties and powers can come apart in different worlds, it’s unclear why they couldn’t come apart in the *same* world. In that case, and insofar as we do not have any access to non-causal quiddities, if properties are individuated within a world even partly by their quiddities, then all bets are off so far as actual property individuation is concerned (see Shoemaker 1980 for a similar epistemological point). As such, those taking properties to have quiddities should agree that the individuation of broadly scientific features proceeds by reference to their powers in a way that is neutral on the presence or absence of quiddities.

The upshot is that the possibility of non-causal quiddities poses no threat to whether a feature S satisfying the conditions of Weak emergence is physically

acceptable, given the physical acceptability of S 's base feature P .

3.4.2 Phenomenal aspects

A second concern about the physical acceptability of features satisfying the conditions of Weak emergence adverts to the phenomenal or qualitative aspects of certain mental features. As Walter (2010) summarizes the concern, “phenomenal properties just cannot be characterized in terms of their causal role, and thus they cannot be individuated in terms of the causal powers of their bearers to which they contribute” (220). Indeed, that phenomenal aspects of mental features “cannot be characterized in terms of their causal role” is a fairly common claim. As I'll now argue, however, none of the motivations for this claim withstand scrutiny.

Consider, to start, Walter's (2010) stated reasons for believing the claim:

Functionalism has replaced the type identity theory not the least because it allowed for the possibility that creatures with a (radically) different biological make-up (mammals, reptiles and mollusks, to take Putnam's (1967) examples) that are in pain have, despite their differences, a property in common, viz., the property having pain. Yet, to the very degree that having pain manifests itself differently in creatures of different species, or in different members in the same species, it evades a characterization in terms of a set of causal powers had by all and only the creatures that have pain. Humans having pain wince and call the doctor, dogs having pain mewl and scratch themselves, and reptiles and mollusks having pain do something else instead. It is not for nothing that functionalism is often said to be plausible for intentional mental properties but implausible for phenomenal properties: phenomenal properties just cannot be characterized in terms of their causal role, and thus they cannot be individuated in terms of the causal powers of their bearers to which they contribute. (220)

This line of thought doesn't show that phenomenal or qualitative aspects of mental states go beyond powers, however, since it interprets the functional role at issue in overly specific terms, as encoding different behaviours associated with different kinds of creature. That's not right: the whole point of looking to functional roles

3.4. OBJECTION: COMPATIBILITY WITH PHYSICAL UNACCEPTABILITY 135

as characterizing higher-level features, whether or not these involve phenomenal aspects, is as encoding the role of the higher-level feature (again, whether phenomenal or not) in such a way as to make sense of the seeming multiple realizability of the feature. Hence just as the functional role associated with *being a mousetrap* must be general enough to encode widely diverse kinds of mousetraps, so too must the functional role associated with *being in pain* be general enough to encode widely diverse realizations of pain. To be sure, as above, there are various deflationary strategies for accommodating seemingly multiply realizable features in terms, e.g., of disjunctions of specific lower-level physical features; but as I argued above, such strategies are ruled out on the assumption that the *Proper Subset of Powers Condition* is satisfied.

A better reason to be concerned about whether phenomenal aspects go beyond powers focuses more directly on phenomenal ‘character’ or qualitative ‘feel’, sometimes described as there being “something that it is like” to have the feature in question. Call a feature having phenomenal or qualitative aspects a ‘qualitative feature’. As Jonas Christensen notes (p.c.), qualitative features needn’t be epiphenomenal—an important qualification for present purposes, since on the operative conception of emergent entities and features these are efficacious (indeed, distinctively efficacious) as well as distinct.¹⁹ Still, Christensen suggests, attention to qualitative feel provides reason to think that phenomenality is not exhausted by the powers associated with it. The intense feel of pain might indeed cause one to cry out and/or behave in certain pain-attending or pain-mitigating ways; nonetheless, Christensen suggests, the qualitative feel of being in pain is reasonably taken to be an extra feature of reality, in addition to the powers associated with it. Supposing so, then even if a qualitative feature *S* satisfied the *Proper Subset of Powers Condition*, *S* would, in virtue of being associated with a distinctive phenomenal aspect, be over and above its base feature *P*, and hence physically unacceptable.

Moreover, Christensen observes, the means of resisting taking the possibility of quiddities to pose a problem for Weak emergence do not carry over to the case

¹⁹That said, as I’ll discuss below, reasons for thinking that phenomenal or qualitative features might be entirely epiphenomenal are in short supply, and arguably reflect an overly mechanistic conception of causation.

of phenomenal aspects, for two reasons. First, unlike quiddities, scientific theory and practice does appear to be concerned with phenomenal aspects of mental features—hence, e.g., pain and how to allay it are part of the subject matter of pharmacology. Second, unlike quiddities, it is commonly thought that if mental features have rich phenomenal aspects that are not fully explained or (more weakly) somehow metaphysically accommodated in terms ultimately involving lower-level physical powers, then such mental features would be physically unacceptable. As such, even granting the success of the previous response to the possibility of quiddities, this response doesn't carry over to the case of phenomenal aspects.

I see at least two strategies of response to Christensen's concern. On the first, one maintains that any phenomenal aspects of features there may be are reducible to non-phenomenal representational aspects; here one might follow reductive representationalists (e.g., Harman 1990, Dretske 1995, Tye 1995, Byrne 2001; see Chalmers 2004 for discussion) or expressivists (e.g., Hellie 2014) in thinking that phenomenal aspects are reducible to non-phenomenal features of reality, which non-phenomenal features in turn are standardly seen as amenable to treatment in terms of powers.

On the second strategy of response, which I prefer, one rather maintains that phenomenal aspects of mental features are fully incorporated into the powers of these features (compatible with a view on which powers are contingently associated features, relative to a given set of laws). In other words, phenomenal aspects are ‘causally loaded’, so to speak. After all, as Christensen notes, qualitative mental features are plausibly taken to enter into causal relations—in virtue, at least in part, of their phenomenal aspects. In my view, it is reasonable to believe that, as our immediate introspective access to the phenomenal aspects of mental features suggests, any (discernible) differences in phenomenality would result in causal differences—including, if one wants a systematic hook to hang this point on, differences in what sort of qualitative experience the bearer of the mental feature will have and (upon reflecting) take themselves to be having. In that case, however, it is reasonable to believe that the powers of a qualitative feature *S* fully incorporate

its phenomenal aspects, in such a way that, were each power of S to be (on an occasion, etc.) token-identical with a power of its lower-level physical feature P , that would indeed suffice for S 's being physically acceptable.²⁰

Call the thesis that phenomenal aspects of a feature are fully incorporated into powers of that feature, perhaps in a way that varies with different laws, ‘Phenomenal Incorporation’.

What reason might there be to reject Phenomenal Incorporation? One purported reason might advert to arguments (see Chalmers 1996, 2009, and 2003) for the (suitably ideal) conceivability of zombies, according to which there could be a world physically and functionally (including causally) identical to the actual world, but with a complete absence of phenomenal character. Supposing that zombies are (suitably ideally) conceivable, and supposing also that (suitably ideal) conceivability is a guide to metaphysical possibility, then one might reasonably think that higher-level qualitative mental features are, even if nomologically connected to lower-level physical features and powers, nonetheless wholly distinct from such features and powers in a way at odds with Phenomenal Incorporation. A related scenario involves the (suitably ideal) conceivability of persons who are physically and functionally identical, yet spectrally inverted (with one seeing green where the other sees red, and so on); here again one might see such a possibility, if genuine, as indicating that qualitative mental features float free of lower-level physical features and powers, contra Phenomenal Incorporation.

There is a huge contemporary literature on whether the zombie and/or inverted spectrum scenarios are genuinely (ideally) conceivable, and on whether (suitably ideal) conceivability is a guide to metaphysical possibility.²¹ In Ch. 7, I will revisit Chalmers's conceivability argument, in particular, in more detail. For now, I am to lay out two ways of maintaining Phenomenal Incorporation, and the associated

²⁰Here it is worth recalling that the physicalist is under no obligation to deny that there are phenomenal aspects of natural reality—they merely maintain that any such aspects are nothing over and above suitably complex lower-level physical goings-on.

²¹Re zombies, see the discussion and references in Kirk 2015; re inverted spectrum cases, see the discussion and references in Byrne 2016; re whether conceivability is the best way to implement ‘epistemic two-dimensionalism’ (by way of regaining, post-Kripke, some a priori access to modal truths), see the discussion and references in Biggs and Wilson 2017.

viability of Weak emergence, in the face of these sorts of scenarios.

First, and perhaps most importantly, even if the scenarios are taken to be ideally conceivable and metaphysically possible, they do not undercut Phenomenal Incorporation *per se*; rather, they undercut Phenomenal Incorporation understood as coupled with physicalism. This much is compatible with qualitative features' satisfying Phenomenal Incorporation, but being Strongly emergent. In that case, however, the scenarios pose no difficulty for the in-principle viability of Weak emergence, for what they establish is not the falsity of Phenomenal Incorporation but rather (at best) the failure of qualitative features to satisfy the *Proper Subset of Powers Condition*. Again: the zombie and inverted spectra scenarios do not show that qualitative aspects float free of powers; they show, at best, that such aspects float free of physical powers.

Second, though it is not the present order of business to defend the live possibility of a Weak emergentist account of qualitative features (a possibility to which we will return in Ch. 7), it is worth observing that a Weak emergentist (non-reductive physicalist) has a fairly straightforward reason to deny that the scenarios are genuinely possible. As Perry (2001) correctly notes *vis-á-vis* the zombie case, physicalists (reductive or non-reductive) taking qualitative features to be efficacious will maintain that the absence of such features at a world would entail the absence of the corresponding powers or associated effects at that world, resulting, contra Chalmers's assumption, in a physical or functional causal difference (see Wilson 2002b for discussion). Similarly, physicalists accepting Phenomenal Incorporation may reasonably deny that either zombie or inverted spectrum cases are genuinely possible. Indeed, independent of whether physicalism is true, it is implausible that inverted spectral differences would fail to make a physical or functional causal difference—one has only to look at a spectrally-inverted image of food, for example, to appreciate how comparatively unappetizing the represented contents seem. Given the availability of this physicalist response, it appears that the zombie and spectral inversion scenarios presuppose rather than establish that qualitative aspects float free of physical goings-on (and so are physically unacceptable). Relatedly, only if one antecedently rejects Phenomenal Incorporation,

either in general or as compatible with physicalism, will one be inclined to think that persons in worlds physically and functionally identical to ours might have mental features entirely lacking in or spectrally inverted with respect to their phenomenal character. As such, the scenarios pose no clear threat to Phenomenal Incorporation, either in general or as compatible with physicalism, and so pose no threat to the claim that satisfaction of the conditions in Weak emergence is insufficient for physical acceptability.²²

How else might one argue against Phenomenal Incorporation in such a way as to target the sufficiency of Weak emergence for physical acceptability? What seems to be required is that it be genuinely possible that qualitative features be epiphenomenal—capable of being instantiated without any powers whatsoever, physical or otherwise. But as above, since the zombie and inverted spectra scenarios are compatible with the truth of Strong emergence, they do not establish the genuine possibility of epiphenomenal qualitative features; and given that the having of a qualitative feature crucially involves the having of a qualitative experience—a seemingly causal affair—one might surmise that no other scenario is going to establish this, either. At present, I can say this much: there are *prima facie* reasons, based in introspective experience, for thinking that any differences in phenomenality would result in causal differences, both in what is experienced and in what one would, on reflection, take oneself to be experiencing. As such, one might also reasonably suppose, as per Phenomenal Incorporation, that the phenomenal aspects of a qualitative feature *S* are fully incorporated in the powers of the feature. I have here considered certain scenarios that some (in particular, Chalmers) have taken to establish that qualitative aspects are not so incorporated, and argued that the scenarios do not establish this. I conclude that, modulo the presentation of some new and better reasons to reject Phenomenal Incorporation, phenomenal aspects of qualitative features pose no problem for the physical acceptability of features satisfying the conditions in Weak emergence.

²²In Ch. 7 we will consider a more sophisticated reason to take the conceivability of zombies to have anti-physicalist import, and will present a more sophisticated response to this line of thought on behalf of the physicalist.

3.4.3 Historical aspects and ‘backwards-facing’ powers

In addition to non-causal quiddities and phenomenal aspects, several other cases have been advanced as showing that satisfaction of the *Proper Subset of Powers Condition* fails to ensure that P (when instantiated in worlds with only physical laws, etc.) necessitates/entails S , and so fails to ensure that S is physically acceptable. As prefigured above, these cases are offered as showing that a supposed metaphysical necessitation or entailment condition on physical acceptability isn’t satisfied, but the deeper concern—that properties have aspects going beyond powers—can be appreciated even if all that is appropriately required for such acceptability is nomological necessitation.

Two such cases are again due to Melnyk (2006), who suggests that if S ’s possession requires having a causal history of some sort (e.g., *being a member of the species Homo sapiens*, or *being a mother*), or requires standing in non-causal (e.g., spatiotemporal) relations (e.g., *being to the right of a rock*), then even if S satisfies the conditions in Weak emergence, S might fail to be necessitated by P , and hence not be guaranteed to be physically acceptable.

A third sort of case, highlighted by McLaughlin (2007), concerns a view of properties on which they are individuated not just by what effects they may contribute to causing (powers, properly speaking), but also by how they may be caused (as per what Shoemaker calls ‘backwards-facing powers’). Insofar as the *Proper Subset of Powers Condition* makes no reference to backwards-facing powers, then satisfaction of this condition by S vis-á-vis P won’t ensure that S is entailed by (the occurrence of) P . What is additionally needed to ensure that P entails S , McLaughlin suggests, is that the backwards-facing powers of P entail the backwards-facing powers of S (as the *Proper Subset of Powers Condition* does for the forward-facing powers of S vis-á-vis P); but there doesn’t seem to be any clear way of doing this. In particular, Shoemaker’s (2007) revision of his account of realization to incorporate reference to backwards-facing powers doesn’t ensure that P even nomologically entails S , for this revision requires that the token backwards-facing powers of an instance of a higher-level realized feature S be a proper *superset* of those of the feature P that realizes it on a given occa-

sion. In that case, as McLaughlin observes, the occurrence of P will have *fewer* backwards-facing powers than S , in which case P will not entail S (at least, not without further conditions).

Following Shoemaker, both Melnyk and McLaughlin direct their cases against an account of non-reductive realization that only appeals to the *Proper Subset of Powers Condition*. Given that Weak emergence also involves a dependence condition (which among other things requires that Weakly emergent features are minimally nomologically supervenient on base features) is also operative in Weak emergence, one response to these cases is to maintain, first, that as discussed above, metaphysical necessitation or entailment is too strong a requirement on physical acceptability, and second, that precisely because the cases assume that (nomological) necessitation or entailment isn't in place, they don't satisfy the dependence condition and so aren't counterexamples to Weak emergence.

Another response, applying to Melnyk's cases, is to maintain that what the cases show is that if the instantiation of a higher-level feature S requires that certain spatiotemporal or causal-historical features or facts be in place, then establishing that S is Weakly emergent will require identifying a lower-level feature P capable of encoding such facts in such a way as to satisfy the dependence condition. For example, features such as *being to the right of a rock* will require realizing features covering the relevant spatiotemporal extent; and if S is a historically sensitive species feature—say, being human—then if P is to minimally nomologically necessitate S , then P might well have to be a ‘broad’, presumably massively complex, extrinsic lower-level physical feature.²³ Such extensions of the base entities and features are familiar from the literature on supervenience as a potential realization relation (to be discussed down the line), where, it is suggested, accommodating extrinsically constituted higher-level features requires taking the physical supervenience base feature to be regional or even global (see, e.g., Horgan 1982, Kim 1984, and Paull and Sider 1992).

The previous response doesn't clearly apply to McLaughlin's case(s), since

²³Perhaps there are other strategies for encoding such historical facts, but this will do for purposes of illustration.

even a highly complex lower-level feature P might not encode all the diverse ways in which a given higher-level feature S might be caused; and more generally, in being more ontologically and causally specific it is likely that P will typically have fewer potential causes than S . A different sort of response is available as regards this case, however. To start, the Weak emergentist can reject as clearly incorrect a view of properties on which they are always individuated (in part) by reference to all the ways in which they can be caused. For example, scientific properties do not seem to be individuated in this way; for example, it isn't any part of what it is to be an H_2O molecule (even given the actual laws) that such molecules might arise either as the result of natural or artificial processes. If higher-level features are not individuated by all the ways in which they may be caused, then the fact that higher-level features can typically be caused in more ways than lower-level features poses no concern for either the necessitation of or the physical acceptability of higher-level features satisfying the *Proper Subset of Powers Condition*. Indeed, and perhaps for such reasons, Shoemaker has since given up individuating properties in terms referencing ‘backwards-facing powers’, and reverted to characterizing realization in terms referencing only powers of the usual, forward-facing variety. To be sure, it is plausible that some higher-level properties are individuated in part by reference to how they were in fact caused (as in the case of species kinds); but in that case the Weak emergentist can offer a response similar to that just given to Melnyk, according to which such cases show only that lower-level feature P must be broad enough to ensure satisfaction of the dependence condition.

3.4.4 Conjunctions and conjunctions

Gibb (2013) offers another sort of case of seeming failure of entailment which again raises the threat of physical unacceptability. She starts by arguing that it is not generally the case that the powers of conjunct features are a proper subset of the powers of their associated conjunctive features, and that on the contrary, cases can be constructed where the powers of a conjunctive feature S are a proper subset of the powers of one of its conjuncts P . As she correctly observes, con-

juncts do not entail conjunctions; so here again a feature S satisfying the *Proper Subset of Powers Condition* might fail to be entailed by P , and hence fail to be physically acceptable—if, in particular, the additional conjunct of the higher-level conjunction involves a non-physical feature.

Here I think the proper response on the part of the Weak emergentist is, in the first instance, to maintain that Gibb's case either fails to satisfy the conditions in the schema, or else poses no threat to the viability of Weak emergence. To start, in general conjunct features will fail to necessitate, in any relevant sense, associated conjunctive features. At least this will be true when the occurrence of the conjunct feature does not necessitate, nomologically or otherwise, the occurrence of the other conjunct(s) in the conjunction. Indeed, in the specific cases that Gibb discusses, the base conjunct feature does not nomologically necessitate the conjunctive feature.²⁴ as such, the case does not satisfy the dependence condition and so is not a counterexample to the sufficiency of Weak emergence for physically acceptable emergence. On the other hand, suppose that the base conjunct feature does (at least) nomologically necessitate the higher-level feature, as the dependence condition requires. In that case, and given the satisfaction of the *Proper Subset of Powers Condition*, there seems no reason to think that the entailed conjunction is physically unacceptable. So here again, properly understood, a case of seeming failure of entailment poses no threat to the physical unacceptability of a feature satisfying the conditions of Weak emergence.

3.4.5 Physically unacceptable individuation

Physically acceptable emergence, according to the schema for Weak emergence, is ultimately a matter of a higher-level feature S having, on a given occasion, only a proper subset of the token powers of the lower-level feature P upon which S depends, on that occasion. But presumably not *every* proper subset of powers had by such a P corresponds to a distinct higher-level feature. So, one might ask, what distinguishes subsets of powers that are associated with a given higher-level

²⁴The reader will have to take my word for this; Gibb's cases involve circuits and are relatively complicated, and it would take us too far afield for me to set these out.

feature from those that aren't? The immediately following concern is that there is room here for an answer that would render the existence or instantiation of the higher-level feature physically unacceptable. For example, Melnyk (2006) says:

[N]ot just any old subcluster of a given cluster of causal power-tokens constitutes a genuine property-instance [...] Hence, some further condition must be met by those subclusters that do (see Shoemaker, 2001, pp. 8586). And it is presumably a task for metaphysics to say what this further condition is. [...] the meeting of the further condition must be a purely physical affair [...]. (146)

He then goes on to suggest a “frivolous” example, in which a sub-cluster of a given cluster of causal power-tokens constitutes a genuine property-instance only if it is divinely classified as natural, and to observe that such a scenario would call into question the physical acceptability of the higher-level feature at issue—even were such a feature to satisfy the conditions in Weak emergence.

Before responding to this concern, we need to clarify just what it is. At least as I read Melnyk, the concern isn't that any property instantiation that might be caused by a divine being is thereby physically unacceptable. Suppose that in fact a divine being said ‘Let there be the big bang and the laws of physics’ and in so saying brought our universe into being, to evolve henceforth according to those initial conditions and laws. This sort of scenario would not obviously falsify physicalism—some physicalists are theists—and relatedly, it would not obviously render all natural features (including physical ones!) physically unacceptable. Rather, the concern is that whatever makes it the case that some proper subsets of token powers of a given lower-level physical feature correspond to (instantiated) features, while other subsets do not do so, had better itself be physically acceptable if the associated historical property is to be physically acceptable; yet satisfaction of the conditions in Weak emergence is silent on why the higher-level feature *S* has the distinctive power profile it has, and so is compatible (one might think) with the instantiation of a higher-level feature being, somehow or other, the outcome of a physically unacceptable process.

Here the first thing the Weak emergentist can say is that, contra Melnyk's claim above, it is not “a task for metaphysics” to say when a given set of powers

is associated with a property; rather, this is a task for the sciences (or whatever domain—e.g., ordinary experience—is at issue). Moreover, in the sciences, which entities and properties are posited as potentially or actually existing or instantiated reflects which such posits are needed in order to make systematic good sense of the phenomena at issue—typically, as entering into the associated fundamental physical or special scientific laws. So again, what broadly scientific entities and features are taken to actually or potentially exist or be instantiated is a question for science, not metaphysics.

A second thing to say is that the sort of problem case Melnyk offers, as involving a supernatural being whose divine classification makes it the case that a given subset of powers corresponds to a genuine feature, doesn't need to be taken seriously. The point here is not one directed at the supposed supernaturalist element of the case; the point is rather, and more generally, that we do not need to take seriously the suggestion that which features (or associated entities) are genuine is a matter of any kind of primitive designation. A similar point would apply to cases in which what subsets of powers are associated with a genuine feature is a matter of primitive ‘naturalness’ (along lines of Lewis 1983a, Sider 2011, and many others).

That said, we can refine Melnyk’s objection, by noting that—and notwithstanding that it is not the job of a Weak emergentist to say which higher-level features are genuine—in many cases the posit of higher-level entities and features corresponds to the presence of certain constraints, broadly speaking. As the qualifier ‘special’ suggests, the special sciences concern broadly natural goings-on when restricted to certain ‘special’ circumstances—corresponding, e.g., to energies and associated temperatures conducive for the formulation of atoms or stable molecules, or where conditions are favorable for life, or where mental creatures exist, and so on. Effectively, such restricted circumstances reflect the presence of certain constraints, broadly speaking, which are operative in individuating the associated special scientific entities, features, and laws—including what such entities and features have the power to do. Now, entities and features that can exist only in some circumstances will be able to do less than, hence will have fewer

associated powers than, systems of entities and features that can exist under the restricted circumstances as well as other circumstances. For example, a system of atoms (the property of being such-and-such system of atoms) will have more powers than any molecule (the property of being such-and-such a molecule) for which it may serve as a lower-level base entity (feature), since the system of atoms may exist in, and contribute to the production of effects in, circumstances in which the molecule cannot exist. And one might reasonably suggest that many special science features which are good candidates for physically acceptable emergence are ones reflecting the holding of certain constraints.

To return to Melnyk's case: that constraints may be operative in the existence and individuation of higher-level entities or features satisfying the *Proper Subset of Powers Condition* provides a way for Melnyk's concern about physical acceptability to be pitched—namely as the concern that these constraints arise as a result of physically unacceptable processes.

Though this concern is a live one, the Weak emergentist has two ways to respond. First, they can maintain that, as per the historical dispute between physicalists (materialists) and their rivals, what distinguishes ‘over and above’ features is just that they have powers their base features don’t have. So even if some supernatural entity were to be operative in making it the case that, e.g., temperatures were within the range needed for stable molecules to exist, this wouldn’t in itself show that molecules were physically unacceptable. This response seems to be generally available to proponents of any implementation of Weak emergence.

A second response would be to make explicit that part of the story about Weak emergence or non-reductive realization concerns the occurrence of constraints, broadly construed—that is, to build in to the operative implementation of Weak emergence that any constraints underlying or responsible for the holding of the proper subset condition must occur as a result of physically acceptable processes. Condition 3 of my DOF-based implementation of Weak emergence ([Wilson 2010b](#)) does just that:

Weak ontological emergence (DOF): An entity E is weakly emergent from some entities e_i if

1. E is composed by the e_i , as a result of imposing some constraint(s) on the e_i ,
2. For some characteristic state S of E : at least one of the DOF required to characterize a realizing system of E (consisting of the e_i standing in the e_i level relations relevant to composing E) as being in E is eliminated from the DOF required to characterize E as being in S .
3. For every characteristic state S of E : Every reduction, restriction, or elimination in the DOF needed to characterize E as being in S is associated with e_i -level constraints;
4. The law-governed properties and behavior of E are completely determined by the law-governed properties and behavior of the e_i , when the e_i stand in the e_i -level relations relevant to their composing E .

If there is a well-motivated revision of the schema for Weak emergence, I am inclined to think that the inclusion of something like clause 3 above is it.

3.4.6 Fundamentally mental powers

A final concern about physical acceptability is due to Baltimore (2013). Baltimore correctly notes that a traditional concern of non-reductive physicalists is to avoid identifying higher-level with lower-level physical features, and he grants that satisfaction of the *Proper Subset of Powers Condition* avoids such identifications, and as such avoids reductionism about higher-level features. But, he suggests, it is compatible with satisfaction of this condition that a single fundamental physical entity—an electron, say—is associated with powers some proper subset of which is associated with a higher-level mental feature:

[A]ccording to Wilson, when the set of causal powers associated with a mental property is a proper subset of the set of causal powers associated with its physical base, the mental property, although distinct from its physical base, will still be physicalistically acceptable [...] There is reason to question, however, the safety of such a retreat. Consider [...] a micro-object at the fundamental level of the micro-macro

hierarchy [;] channel the spirit of panpsychism and endow the fundamental micro-object with mentality. However, instead of it having a mental property that is identical with one of its physical properties, suppose that it has a mental property that non-reductively supervenes on one of its physical properties. Suppose further that [...] each causal power associated with the mental property is identical with a causal power associated with its physical base. But the physical base here is a fundamental property and, so, mental causal powers are thereby associated with a fundamental property, which seems physically unacceptable. (20)

Baltimore goes on to surmise that the main difficulty here is that the requirement that mental causal powers be identical with physical causal powers is itself insensitive to whether or not the lower-level physical powers are themselves fundamentally mental, or not.

This is an interesting case, and I agree with Baltimore that it is not ruled out just by satisfaction of the *Proper Subset of Powers Condition* (or more generally, the conditions in Weak emergence). It is ruled out, however, by (what most will take to be) the operative conception of the physical. As discussed in the Preliminaries, this account is along lines of that I have proposed and defended in [Wilson 2006a](#):

The physics-based NFM (no fundamental mentality) account: An entity or feature existing at a world w is physical if and only if (i) it is treated, approximately accurately, by current or future (in the limit of inquiry, ideal) versions of fundamental physics at w , and (ii) it is not fundamentally mental (that is, does not individually either possess or bestow mentality) (72)

This account aims to encode the historical philosophical motivations for characterizing the physical—namely, in order to express the felt pressure of the mind-body problem and its generalizations—as well as certain facts about the target subject matter of physics (as not comprising everything, in particular) and the current state of that discipline (as presumably on the right track, though presently imperfect),

while avoiding Hempel's dilemma. And as I've argued, the account (or variations on this theme) achieves these aims, and does so better than any competing accounts.

Baltimore considers this response, and replies that the NFM constraint should not be imposed, since it "builds too much metaphysics into the notion of the physical", and in particular, rules out that the physical entities might turn out to be as panpsychists suppose. Given the desiderata on an account of the physical that I've identified, however, more would have to be done in order to show that the NFM account wasn't better, all things considered. In the absence of a superior alternative, I am inclined to see Baltimore's case as supporting the imposition of the NFM constraint—since after all, it seems reasonable to want the physical to be physically acceptable, even if this is not (as I argue; see also [Dowell 2006](#) and [Ney 2008](#)) analytic.

3.5 Objection: non-necessity

I turn now to considering objections each of which can be seen as challenging the claim that satisfaction of the conditions in Weak emergence is necessary for physically necessary emergence. There are three main lines of objection to this claim, each consisting in the presentation of an alternative account of such emergence in terms of, first, token identity; second, constitution; and third, primitive Grounding. (A fourth line of objection to the necessity claim appeals to supervenience, with the general idea being that it suffices for physically acceptable emergence that a higher-level feature asymmetrically metaphysically supervene on lower-level physical features; however, since appeals to supervenience in this context are primarily aimed at distinguishing physically unacceptable emergence from any sort of physically acceptable relation, I reserve discussion of supervenience as a basis for metaphysical emergence in the next chapter.) In what follows, I provide reasons to think that each of these alternative approaches to physically acceptable emergence is unsatisfactory.

3.5.1 Token identity

On a token identity approach, a non-reductively realized feature S is token but not type identical to the base feature P upon which it depends, on a given occasion (see, e.g., [Davidson 1970](#), [Macdonald and Macdonald 1995](#), [Ehring 2003](#), and [Robb 1997](#)). Such an approach entails that every token power of S , on an occasion, is identical to a token power of base feature P , on that occasion, and hence avoids token-level overdetermination between S and P . It doesn't gain S 's (token-level) ontological autonomy, but one might think that this isn't as important as gaining S 's reality and efficacy. It remains, however, that the causal autonomy characteristic of emergence (of whatever variety) requires not just that S be efficacious, but distinctively so.

How might S be distinctively efficacious, on a token identity account? It cannot be so in virtue of its distinctive token powers, or in virtue of having a distinctive cluster of token powers, since by assumption S and P share all their token powers. The remaining option (as per [Macdonald and Macdonald 1995](#) and [Ehring 2003](#)), is that S is distinctively efficacious in virtue of falling under a distinctive type.

One concern here is that appeal to non-identical types as the ground of the causal autonomy of S vis-à-vis P reintroduces a (or the) threat of causal overdetermination, and associated threat of exclusion. As [Ehring \(2003\)](#) puts it: "Since mental types are not identical to physical types (because of multiple realizability) even if mental tokens are identical to physical tokens, there are no causes of physical effects that are efficacious in virtue of mental property types" (364). To gain S 's causal autonomy, the proponent of a token identity account must provide an account of the relation between S and P 's associated types, and show that the associated means of gaining autonomy does not reintroduce problematic overdetermination.

Ehring provides an account of the relation between types aiming to show that this threat is avoided. To start, he takes S and P to be tropes—particularized properties, such as *this redness*, or *that complex configuration of charges*—and their associated types to be collections of resembling tropes. He then argues that

S and *P*'s types are related as part to whole. Here the order of the part/whole relation is reversed from Clapp's (2001) understanding: for Clapp, a realized type is part of each of its realizing types; for Ehring, a realized type is a whole, having as parts the subclasses (of resembling tropes) of its realizing types.

What is interesting for present purposes is that Ehring takes appeal to a type-level powers-based strategy as required (by the token identity theorist) to establish the requisite causal autonomy without inducing problematic overdetermination. He first motivates the view for the type *red* and associated shade types:

It seems clear that for the class of red tropes as a whole, the type “red”, has certain causal powers. [...] [W]e are still left with the question of how the causal powers of this class as a whole are related to the causal powers of the subclasses of each determinate shade of red. I believe the answer is that the causal powers of the type “red” are those exactly similar causal powers shared by each of these subclasses. [...] For any causal power of a shade of red not matched by an exactly similar causal power belonging to each of the other shades of red, “red” lacks any such power. (374)

In other words, the powers of the type *red* are a proper subset of those of its constituent determinate types. Ehring takes similar considerations to indicate that the powers of mental types are a proper subset of the powers of their realizing physical types. Macdonald and Macdonald (1995) also plausibly implement a subset-of-powers-based strategy at the level of types, for they take mental state types to be relevantly analogous to determinables, and as previously, determinable types arguably have a proper subset of the powers of their realizing determinate types.

Such hybrid approaches, combining token identity of features with a proper subset relation between powers of associated types, are problematic, however, for a reason that we have already touched on. On the hybrid view, *S*'s type does not have powers that differ between its realizer types; but a token of *S*'s type can, when identical with a token of *P*'s type, have such powers. Hence token feature *S* can have powers that *S*'s type doesn't have. But it arguably makes no sense for a token feature to have more powers than its type, at least if types are supposed

to track similarities among associated tokens. If a token feature has more powers than a given type, that is itself compelling reason to think that the feature is not of that type, or so it seems to me.

Avoiding this difficulty requires that the proper subset relation between powers hold at the level of tokens as well as types—that is, that the *Proper Subset of Powers Condition* be imposed. More precisely, it requires imposing the *Proper Subset of Powers Condition* if the account is to be a version of non-reductive physicalism. Alternatively, the proponent of a token identity account could endorse reductionism at both token and type levels, or else endorse Strong emergentism at the type level. At the end of the day, token identity accounts of non-reductive realization either do not establish autonomy of higher-level features, contra non-reduction; or else impose the *Proper Subset of Powers Condition*, and so are not really token identity accounts, after all.

3.5.2 Constitution

On Pereboom's (2002) account of "robust" non-reductive physicalism—i.e., of physically acceptable emergence—a higher-level feature S is neither type- nor token-identical with the lower-level physical feature P upon which it depends; and contra both the *Token Identity Condition* and the *Proper Subset of Powers Condition*, S 's token powers are "irreducible to" powers of P : "robust non-reductive physicalism affirms various token-diversity claims for mental causal powers" (500). Such a view will clearly make sense of S 's ontological and causal autonomy. But how are S 's token-irreducible powers supposed to avoid problematic overdetermination while retaining compatibility with physicalism?

According to Pereboom, this is because S 's powers are "constituted" by P 's powers, in a way piggybacking on the notion of token feature constitution:

Token Power Constitution: The causal powers of a token of kind F are constituted of the causal powers of a token of kind G just in case the token of kind F has the causal powers it does in virtue of its being constituted of a token of kind G . (504)²⁵

²⁵See Pereboom and Kornblith 1991, 131.

The notion of constitution of one token feature by another is broadly primitive, but is (as per Pereboom 2011), to be grasped as relevantly analogous to the “made up of” relation holding between one particular and another (e.g., a statue and a lump of clay). The account of feature constitution, coupled with *Token Power Constitution*, is intended to motivate taking the powers of a non-reductively realized feature *S* to be, while irreducible to, still nothing over above the powers of *P*:

[Though *S*'s token powers are irreducible to *P*'s] there would be a sense in which the token causal powers of *S* would be “nothing over and above” the token causal powers of *P* [...] *S*'s causal powers would nevertheless be “absorbed” or “swallowed up” by *P*'s causal powers. But there are importantly distinct modes of this sort of absorption: identity and constitution without identity. [...] token mental causal powers are wholly constituted by token microphysical causal powers. (503–4)

(I have changed Pereboom's notation for continuity with my discussion.) But such appeals to token feature and power constitution do not establish that rejection of the *Token Identity Condition*, hence *Proper Subset of Powers Condition*, is compatible with physicalism. “Constitution” is a term of art, applied mainly (as Pereboom notes) to objects. Where token features are at issue, and where conformity to physicalism is presumed, “constitution” is usually just another name for “realization”. But as previously argued, standard accounts of non-reductive realization presuppose satisfaction of the *Token Identity Condition*. The expression “in virtue of” entering into the account of token power constitution is also a term of art, compatible with many underlying relations, including identity (satisfying a token identity condition on powers) and the determinable/determinate relation (satisfying the *Proper Subset of Powers Condition*).

Pereboom offers further considerations in support of irreducible mental powers' being compatible with physicalism and with the avoidance of overdetermination, but these also fail to establish his case. In re compatibility with physicalism, he says that “correlated with the possibility of this sort of constitutional explanation is the fact that the existence and nature of token higher-level causal powers

would be predictable in principle from their microphysical constituents together with the laws governing them” (504). But if the powers of S , at either the type or token level, are not identical with the powers of P , what guarantees that the powers of S would be so predictable? Perhaps such predictability could be guaranteed if every power of S was type-identical (though token-distinct) with a power of P . But this understanding looks to give rise to the worst kind of overdetermination.

Pereboom resists this conclusion, saying that “no competition arises in the case of mere constitution”:

For if the token of a higher-level causal power is currently wholly constituted by a complex of microphysical causal powers, there are two sets of causal powers at work which are constituted from precisely the same stuff [...] and in this sense we might say that they coincide constitutionally. (505)

To the extent that I understand why constitutional coincidence blocks overdetermination, however, this is because (the relevant sort of) coincidence would entail identity of the token powers had by features of the “stuff”. Pereboom acknowledges “that they now coincide in this way might tempt one to suppose that these causal powers are token-identical, but, as we have just seen, there is a good argument that they are not” (505). Here Pereboom is referring to his argument from the supposed token multiple realizability of S to the non-identity of S ’s token powers with P ’s powers; but as I discussed in §2.3 of this chapter, that argument is unsuccessful.

Finally, Pereboom provides an indirect argument that problematic overdetermination can be avoided even if the *Token Identity Condition* is rejected. He starts by saying, “If identity [of powers] and not just constitutional coincidence were necessary for [...] noncompetition, then there would be features required for non-competition that identity has and current constitutional coincidence lacks” (506). He goes on to say that the features possessed by identity but not constitutional coincidence are coincidence at all other times and worlds; then notes that lack of such features surely could not induce overdetermination. But what identity of powers has and constitutional coincidence of powers lacks, and which is necessary

for non-competition, is not a matter of goings-on at other times or worlds. It is rather simply that identity of powers on an occasion guarantees noncompetition on that occasion, whereas constitutional coincidence on an occasion without identity of powers on that occasion seems not to avoid competition, but rather to invite it (if *S* and *P* are token distinct, but type-identical). Summing up: Pereboom doesn't provide reason to think that failure to satisfy a token identity condition on powers (hence, more specifically, failure to satisfy denying the *Proper Subset of Powers Condition*), is compatible either with physicalism or with avoiding problematic overdetermination, much less jointly so compatible.

3.5.3 Grounding

It has recently been suggested that complete metaphysical dependence should be understood in terms of a primitive relation or notion—call it ('big-G') 'Grounding', supposed to be operative in any context where some goings-on or facts are 'nothing over and above' or hold 'in virtue of' some others (see [Fine 2001](#), [Schaffer 2009](#), [Rosen 2010](#), [Audi 2012](#), [Raven 2015](#); see [Bliss and Trogdon 2014](#) for an overview and other references). Moreover, proponents of Grounding typically stipulate that Grounding is a partial order—asymmetric, irreflexive, and transitive—such that Grounding supports taking Grounded goings-on to be ontologically autonomous from (distinct from) Grounding goings-on. In particular, [Schaffer \(2009\)](#), [Rosen \(2010\)](#), [Dasgupta \(2014\)](#), that Grounding provides an apt basis for formulating physicalism, as follows:

Physicalism (Grounding): All broadly scientific goings-on are Grounded in lower-level physical goings-on.

[Schaffer \(2009, 364\)](#), [Rosen \(2010, 111–112\)](#), and [Dasgupta \(2014, 557\)](#) each motivate a Grounding-based formulation of physicalism by appeal to the following form of argument:

1. Physicalism is the thesis, schematically speaking, that all broadly scientific goings-on are nothing over and above lower-level physical goings-on.

2. The operative notion of “nothing-over-and-aboveness” cannot be successfully characterized in semantic/representational, epistemic, or purely modal (i.e., supervenience-based) terms.
 3. No other non-primitive approach to characterizing nothing over-and-aboveness is available.
- ∴ The operative notion of “nothing-over-and-aboveness” in physicalism should be characterized in terms of primitive Grounding.

Such an argument might be seen as offering an alternative means of characterizing a form of non-reductive realization relation making no reference to the *Proper Subset of Powers Condition* in the schema for Weak emergence, and so to suggest that the latter condition, and associated schema, are not necessary for emergence of a physically acceptable variety.

As I've discussed in a series of papers ([Wilson 2014, 2016c, 2016b](#)) there are many ways in which this line of thought goes wrong. A full discussion of all of these would take us too far afield; here I'll simply flag some of the more obvious.

First, as I note in [Wilson 2014](#), arguments like that above for positing Grounding are unsound, since premise (3) is false. To start, and as will already be clear from previous discussion of the many varieties of realization relations, during the past several decades philosophers working on physicalism have identified and explored numerous specific accounts of metaphysical dependence, explicitly assumed to go beyond merely modal, representational, or epistemic notions. These accounts fill in the schematic reference to ‘nothing over and-aboveness’ (or other rough-and-ready idioms of complete metaphysical dependence of the sort entering into schematic formulations of physicalism) with specific familiar metaphysical relations—what I call in my 2015 ‘small-‘g’ relations—including type and token identity, functional realization, the determinable-determinate relation, the causal composition relation, the part-whole relation, the proper-subset-of-powers relation, and so on, which serve, against the backdrop of the specified lower-level physical base, to characterize diverse forms of metaphysical dependence in an explanatory and illuminating way. Given all these highly articulated, metaphysically

substantive suggestions for how to fill in the operative understanding of complete metaphysical dependence at issue in physicalism, there is not even a *prima facie* route from the failure of modal/epistemic/representational conceptions of such dependence to a primitive Grounding-based understanding of this notion (much less to a primitive understanding tracking just metaphysical dependence of a non-reductive variety).

Moreover, and even putting aside the initial enthymematic argumentation for positing Grounding, the absence of content associated with this primitive renders it incapable in itself of shedding any illuminating light on the notion of ‘nothing over and above’ it is introduced as explicating. For example, we saw in Ch. 2 that a primary issue in debates over how to understand higher-level goings-on concerns whether, if these are synchronically materially dependent on lower-level physical goings-on, any effects purportedly caused by a higher-level entity or feature are causally overdetermined (since already brought about by lower-level dependence base entities or features; and large part of the action concerning this debate has involved exploration of what import the holding of a specific relation (e.g., functional realization, the determinable-determinate relation, token identity, and so on). But the holding of a primitive Grounding relation says nothing—at least, nothing non-stipulative—about whether and how causal overdetermination might be avoided as between higher- and lower-level entities and features, or even about whether a Grounded entity has any powers at all.

Here and elsewhere, metaphysical investigations cannot proceed just by appeal to Grounding—appeal to the specific metaphysical (as I call them: ‘small-‘g’’) relations is unavoidable if any insight into the nature of the metaphysical dependence is to be gained. Correspondingly, rendering the motivation for a Grounding-based formulation of physicalism sound requires that proponents provide reason for thinking that primitive Grounding serves some useful purpose that is not already accomplished by the specific metaphysical relations that are standardly operative in the physicalism and other debates.

So far two strategies have been advanced, neither of which is compelling. The first appeals to unity considerations, with the basic idea being that Grounding is

needed in order to formally and otherwise unify the small-‘g’ relations. But as I and others have argued (see Wilson 2014, 2016c; Koslicki 2012, Koslicki 2016), the small-‘g’ relations are not formally or otherwise unified. Moreover, even if they were unified, more would be required to establish that such unity motivates a generic worldly posit, as opposed to merely motivating a schematic or generic concept.

The second strategy appeals to a purported need for Grounding in order to fix the direction of priority of the small-‘g’ relations, since (as the dispute between Monists and atomists over whether the Cosmos *qua* whole is prior to or posterior to its proper parts illustrates) these relations do not fix priority on their own. But as I have previously argued, even granting that more is needed for small-‘g’ relations to determine a direction of priority, what more is needed is not a primitive pointer (i.e., Grounding) but rather a specification of what is or serves as fundamental. For example, given that the Cosmos is fundamental, proper parts of the Cosmos are non-fundamental; given that the atomic parts are fundamental, fusions of the parts (including the Cosmos) are non-fundamental.²⁶

In sum: a primitive notion of non-reductive dependence is insufficiently contentful to itself serve as a basis for characterizing emergence of a physically acceptable variety. There is no avoiding appeal to one or other of the specific non-reductive realization relations which have been the focus of non-reductive physicalist attention for the past thirty years, and beyond—which relations, I have argued, have in common the intended satisfaction of the conditions in Weak emergence. Moreover, the holding of these relations, against the backdrop assumption that the lower-level physical goings-on are fundamental, is sufficient to ensure that realized higher-level goings-on are less fundamental than their lower-level physical realizers, without the additional posit of Grounding to act as a primitive priority pointer. Correspondingly, attention to Grounding provides no reason to deny that satisfaction of the conditions in the schema for Weak emergence (plausibly implemented, as per usual) is necessary as well as sufficient for emergence of a physically acceptable variety.

²⁶This is a first pass on the more complex story discussed in Wilson 2014, 2016c, and 2016b.

3.6 Concluding remarks



In this chapter I have defended the claim that satisfaction of the conditions in Weak emergence is core and crucial to—and when sensibly filled in, necessary and sufficient for—a higher-level feature’s being metaphysically emergent from a lower-level feature, in such a way that, if the lower-level feature is physically acceptable, then so will be the higher-level feature. I have argued that the many objections to the viability of Weak emergence can be answered. Again, each objection admits of at least one response relying only on general considerations, which could be offered by proponents of any of the diverse implementations of the schema, including accounts pitched in terms of functional realization, the part-whole relation, the determinable-determinate relation, or degrees of freedom. In a few cases, additional responses are available which rely on features specific either to a determinable-based account or a DOF-based account. In particular, a determinable-based account provides a non-stipulative basis for ruling out the Weak emergence of conjuncts of lower-level conjunctions, and a DOF-based account explicitly includes a condition specifying that the holding of any constraints entering into or responsible for the holding of the *Proper Subset of Powers Condition* (as is common in the special sciences) must be a matter only of physically acceptable processes. My sense is that if any of the considered objections requires tweaking the schema for Weak emergence, it is the one (due to Melnyk) pertaining to the need to require the physical acceptability of any operative constraints associated with the higher-level feature and its distinctive power profile. If so, addition of such a condition would be straightforward, and hardly fatal to the overall approach; but again, this is a choice point, and how exactly a Weak emergentist chooses to respond to the various objections may depend on further of their commitments. In any case, these results collectively indicate that Weak emergence is not just a viable and indeed attractively robust means of accommodating physically acceptable emergence, but moreover captures what is core and crucial to this notion.

Chapter 4

The viability of Strong emergence

I now turn to a project similar to that of the last chapter, only as directed at the schema for Strong emergence. Again, the schema is as follows:

Strong emergence: Token apparently higher-level feature S is strongly metaphysically emergent from token lower-level feature P on a given occasion just in case, on that occasion, (i) S synchronically materially depends on P , and (ii) S has at least one token power not identical with any token power of P .¹

In Chapter 2, I provided *prima facie* reasons for thinking that (given that the dependence base entities and features are physically acceptable) satisfaction of the conditions in Strong emergence is core and crucial to—and when sensibly filled-in, both necessary and sufficient for—for metaphysical emergence of a physically unacceptable (more generally, ‘over and above’) variety. I also argued that a representative range of accounts of physically unacceptable emergence satisfy the schema for Strong emergence.

In this chapter, I will consider and respond to a range of objections that can and have been made to the viability of Strong emergence. Some of these target

¹More generally, as previously discussed, for a feature’s being Strongly emergent it plausibly suffices that the conditions are satisfied on at least one occasion by at least one instance of the feature in worlds with the laws relevantly similar to the actual laws; for continuity with the schema for Weak emergence I stick with the schema expressed in terms of occasions.

physically unacceptable emergence on general grounds, as being at odds with physics or, more generally, with an appropriately scientific or naturalist outlook; some aim to show that the conditions in the schema cannot be satisfied, since any new powers purportedly had by a Strongly emergent feature S are trivially inherited by, or otherwise ‘collapse’ so as to be had by, the lower-level physical feature P upon which S depends; some target the sufficiency claim, aiming to show that satisfaction of the conditions is compatible with higher-level feature S ’s being physically acceptable; others target the necessity claim, on grounds that other routes to physically unacceptable emergence are available for the cases in question. As I’ll argue, each of the objections admits of at least one general response that proponents of any of the aforementioned implementations of Strong emergence might accept. And here again, as with the case of Weak emergence, upon occasion an additional response to an objection is available which relies on a specific implementation of the schema—namely, a fundamental interaction-relative account of the sort I have previously endorsed.

Two points before getting started. First, I continue to assume that the base features are physically acceptable, and will speak of the ‘Strong emergentist’ as the proponent of the view that higher-level features conforming to the conditions in the schema for Strong emergence are sufficient for emergence of the physically unacceptable variety. Second, the schema for Strong emergence has been implemented as involving new fundamental features, laws, forces, interactions, and/or powers; different philosophers may come down differently on how to understand these phenomena and how each is related to the others. The schematic characterization of Strong emergence specifically in terms of powers reflects, as discussed in Chapter 2, the common core of these phenomena as pointing towards the having of a novel fundamental power, as per the Strong emergentist’s distinctive response to the problem of higher-level causation and the associated denial of *Physical Causal Closure*. That said, in assessing the viability of Strong emergence, we will sometimes have occasion to discuss certain of these other categories.

4.1 Incompatibility with physics, scientific practice, naturalism

Recall that Strong emergentists—prototypically, the British emergentists Mill (1843/1973), Bain (1870), Morgan (1923), and Broad (1925)—maintain that some higher-level entities and features have or bestow new powers, associated with higher-level laws, forces, or interactions that are as metaphysically and scientifically basic or fundamental as the fundamental physical powers, laws, forces and interactions. In terms of forces, for example: in cases of Strong emergence, the operative forces are a combination of physical and ‘configurational’ forces, where configurational forces (following McLaughlin 1992), are forces occurring only when certain lower-level configurations (or pluralities) are present; correspondingly, the higher-level entity or feature at issue has or bestows powers to produce forces going beyond the operative physical forces. These will generally include powers affecting the motion of lower-level physical entities; hence Strong emergence is usually thought to involve the nomological possibility of “downward causation”.

Several concerns with Strong emergence, so understood, have to do with its being in tension, one way or another, with the practice or content of scientific theories. One such concern is that Strong emergence is incompatible with contemporary physics. In response to this concern, McLaughlin (1992) convincingly argues that Strongly emergent configurational forces or interactions are compatible with the laws and conservation principles of physics (see also Horgan 1993 and Papineau 2001). For example, Newton’s second law of motion, $F = ma$, is neutral as regards which forces enter into the net force F ; hence compatible with these including a fundamental configurational force. Similarly, McLaughlin notes, for the contemporary descendant of Newton’s law, Schrödinger’s equation $H\psi = ih\frac{\partial\psi}{\partial t}$, into which is inserted the Hamiltonian H specifying the energies of the state (forces and energies being inter-translatable²):

It is not that British Emergentism is logically incompatible with non-relativistic quantum mechanics. It is not. Schrödinger’s equation

²See Wilson 2007 for discussion and some derivations.

could be the fundamental equation governing motion in a world with energies that are specific to types of structures. (54)

Kane (1993) confirms:

$F = ma$ is used to compute the motion of an object, given *any* force F on the object. And specific classical forces have been discovered, such as gravity with $F = \frac{G_N m M}{r^2}$ [...]. Hamilton's or Lagrange's equations are equivalent to $F = ma$ in a different formulation. In quantum theory there is an analogous structure. The Schrödinger equation [...] is like $F = ma$. It holds for any Hamiltonian. Specific forces lead to specific Hamiltonians. (2–3)

Configurational forces needn't contravene the physical forces; rather, physical and non-physical forces could operate in tandem, just as the diverse physical forces do. Nor are configurational forces incompatible with conservation laws, such as the relativistic principle of conservation of mass-energy. As McLaughlin observes:

[C]onfigurational forces need not involve any violation of this principle. [...] Configurational forces could involve various compensating shifts in mass and energy that maintained conformance to the principle of mass-energy. (74)

A second concern is that even if there is nothing in-principle problematic about adding fundamental configurational forces to the physical mix as entering into the operative equations of motion, such forces are compatible with scientific practice, or relatedly, with a ‘naturalist’ outlook, according to which metaphysical investigations should be consonant with such practice. As I discuss in Wilson 2002a, however, scientific theorizing itself provides a blueprint for how conservation laws might support the warranted posit of Strongly emergent configurational forces and associated powers, laws, and features. In the 1930's, the law of conservation of mass-energy appeared to be violated in nuclear β decay interactions. Rather than accept the apparent violation as genuine, physicists posited a new fundamental force, or interaction—the weak nuclear interaction—as carrying away the missing energy. Nuclei are composite entities; hence evidently scientists have no problem

with positing fundamental configurational forces. As it happens, the nuclear interactions are now understood as ultimately due to interactions between sub-nuclear entities, but the point remains: the posit of fundamental configurational forces or interactions—that is, forces or interactions that only come into play at a certain level of comparatively complex organization—and associated powers is compatible with scientific practice, and hence with a naturalistic approach to metaphysical theorizing.

A third concern is that, even granting that there is nothing in-principle problematic about configurational forces, in any case the present evidence of science is that there aren't any such forces. This appears to be McLaughlin's considered judgement, as well as Ladyman's and Ross's in their (2007). Strictly speaking, this isn't so much an objection to the viability of Strong emergence *per se* as it is to its actual applicability. Even so, it's worth noting that the claim that we presently don't have any reason to think that there is any Strong emergence is overstated. That certain historical candidates for Strong emergence now admit of lower-level physical explanations doesn't show that all such candidates have been so explained; and indeed, as we will see down the line, both consciousness and free will, not to mention any artifacts that are in part the products of these features, remain live candidates for Strong emergence, in ways that (as per the responses to previous concerns) are in keeping with a naturalistic methodology.

A final concern about the compatibility of Strong emergence and scientific theory and practice has to do with the usual understanding of such emergence as involving commitment to the nomological possibility of downward causation. Wouldn't the efficacy of Strongly emergent properties *vis-à-vis* physical or physically acceptable effects, if stemming from the power(s) at issue in satisfaction of the *New Power Condition*, violate *Physical Causal Closure*, according to which every lower-level physical effect has a purely lower-level physical cause? And isn't *Physical Causal Closure* widely accepted? Yes, and yes. However, *Physical Causal Closure* is not a principle of contemporary physics (though no doubt many physicists, and scientists more generally, believe it). The acceptance of this principle is rather a constraint on *physicalist* theorizing (which motivates, in par-

ticular, the physicalist approach to the problem of higher-order causation); hence that Strong emergentists deny it is not in itself a strike against their view.

4.2 Collapse

The second line of objection aims to show that the conditions in the schema for Strong emergence are never jointly satisfied, as per what Taylor (2015) evocatively calls the “collapse” objection, versions of which have been raised and/or addressed by a number of philosophers (including van Cleve 1990, Kim 1999, O’Connor 1994, Wilson 2002a, Francescotti 2007, Howell 2009, and Taylor 2015). The general concern is that Strong emergence makes no sense, since any purportedly Strongly emergent features or associated powers can be seen to ‘collapse’, one way or another, into the lower-level base features upon which they depend, undermining the supposed metaphysical and causal novelty or associated non-deductibility of the emergent features (Kim, van Cleve, Taylor).³

As discussed in Baysan and Wilson 2017, there are two main strategies for pressing the objection, to be discussed in the following sections. The first proceeds by arguing that on the most common understanding of when it is that a feature has a power, a lower-level physical base feature P automatically inherits any purportedly ‘new’ powers had by features satisfying the dependence condition *vis-á-vis* P , and that a natural revision of the account of power possession has the converse problem of rendering it too easy for a higher-level feature to be Strongly emergent. The second aims to show that any lower-level physical feature P will inherit the powers of purportedly Strongly emergent features in virtue of either P or some other lower-level feature(s) having propensities or other dispositional capacities to give rise to features having those powers. After presenting each ver-

³A related concern is that such collapse, combined with the supposed physical unacceptability of Strongly emergent features and powers, threatens the physical acceptability of the base features (Howell); since Howell offers this objection as attaching only to accounts on which Strongly emergent features are metaphysically necessitated by base features, in support of a supervenience-based approach to physically unacceptable emergence, I discuss this concern later in this chapter, in §4.3.2.

sion of the collapse objection, noting certain difficulties with previous responses along the way, I offer three ways via which a Strong emergentist can respond to the collapse objection(s), drawing on [Wilson 2002a](#) and [Baysan and Wilson 2017](#).

4.2.1 Collapse via power possession

The first route to the collapse objection is one according to which an intuitive way of assigning powers to features entails that any purportedly new power of a Strongly emergent feature S will in fact be inherited by its base feature P . The concern here underlies what [Kim \(1998\)](#) calls the “crucial” or “critical” question of emergence:

If an emergent, M , emerges from basal condition P , why can't P displace M as a cause of any putative effect of M ? Why can't P do all of the work in explaining why any alleged effect of M occurred? (32)

[Kim \(2006\)](#) expands on this concern, as follows:

M , as an emergent, must itself have an emergence base property, say P . Now we face a critical question: if an emergent, M , emerges from basal condition P , why cannot P displace M as a cause of any putative effect of M ? [...] If causation is understood as nomological (law-based) sufficiency, P , as M 's emergence base, is nomologically sufficient for it, and M , as P^* 's cause, is nomologically sufficient for P^* . It follows that P is nomologically sufficient for P^* and hence qualifies as its cause. (558)

Given that, as we are assuming, a feature's powers are a matter of what effects the having of that feature can contribute to causing, when in certain circumstances, the threat to the viability of Strong emergence is clear. For as above the Strong emergentist standardly supposes that the dependence of a Strongly emergent feature S involves, at a minimum, the base feature's being nomologically sufficient for S ; moreover, nomological sufficiency (in the circumstances, etc.) is transitive. Consider, then, any power of S to contribute to causing an effect E in circumstances K . If causation is a matter of nomological sufficiency in appropriate circumstances, if

P is nomologically sufficient for *S* in *K*, and if *S* is nomologically sufficient for *E* in *K*, then *P* will also be nomologically sufficient for *E* in *K*, and so also have the power to contribute to causing *E* in *K*, ruling *S*'s Strong emergence out of court. In other words, any supposedly novel powers of *S* will ‘collapse’ into those of *P*.

Insofar as the line of thought here depends on certain assumptions about causation, one might wonder whether the Strong emergentist can respond by rejecting a view on which nomological sufficiency in the circumstances is sufficient for causation, or by denying that causation is transitive, or by denying that causation can be (as with *P* and *S*) synchronic. Such responses are unsatisfactory, however. To start (following Hall 2004), accommodation of many intuitive cases of causation requires a notion of causation as “production”, involving nomological sufficiency in the circumstances; moreover, other accounts of causation seem likely to introduce similar or other difficulties.⁴

Most importantly, even if it is possible to block taking *P* to cause *S* in cases of Strong emergence, there would remain a case to be made that *P* inherits any powers of a feature *S* that at least nomologically depends on *P*. Here O'Connor's (1994) presentation of the following “strong objection” to a powers-based account, which he credits to Carl Ginet, is apropos:

If an emergent property is a necessary consequence of certain base-level properties (as is implied by the supervenience condition), then its instantiation is one of the potentialities of that set of properties. But then are not the further potentialities of this emergent property also a subset of the total set of potentialities of the base properties, in virtue of the necessary connection between the base properties and it? These

⁴In particular, if causation is counterfactual dependence (the other main category of causation that Hall identifies), such that a power is associated with a feature only if the associated effect is counterfactually dependent on the feature, two difficulties ensue. If the counterfactual dependence concerns the token instances of *S*, *P*, and *E*, then it might be reasonably thought that if *S*'s power to cause *E* reflects *S*'s being necessary in the token circumstances for *E*, then *P* was also so necessary for *E*, in being necessary in the token circumstances for *S*. So collapse remains. If the necessity rather attaches to the types at issue, then a different problem arises—namely, that any higher-level feature with multiple dependence bases will be deemed Strongly emergent, including multiply realized features that are intuitively physically acceptable. It shouldn't be that easy to falsify physicalism! And while it would be less costly to deny that causation must be transitive or to require that it be diachronic, these responses are both overly committing and ultimately ad hoc.

further potentialities are simply potentialities of the base properties at one remove. And now one is led to wonder why we might ever think to postulate an emergent property at all, since it provides no explanatory gain over an account which excises the mediating link by taking the “further” potentialities as directly tied to the base properties. This objection implies, in effect, that the features of supervenience and novel causal influence are incompatible. (98)

The deeper collapse objection raised here, as well as in Kim’s (1998) framing of the objection, does not hinge on the supposition that Strong emergence is a causal relation but rather hinges merely on the supposition that P synchronically materially necessitates S . Such necessitation alone suggests that anything that S can do in circumstances K is also something that P can do in circumstances K , such that there is no way for S to have a novel power, and hence no way for it to be Strongly emergent.

O’Connor’s own response to the deeper collapse concern is unsatisfactory. He suggests that if P is taken to inherit S ’s powers, then the lower-level physical laws would have a “very odd complexity, involving tacked-on disjuncts to cover the special cases” (1994: 98). Effectively, O’Connor’s suggestion is that collapse would entail that lower-level entities (e.g., atoms) would interact each other in a uniform way until entering into a complex aggregate, at which point they would start doing “quirky” things, and that such discontinuous behaviours are best explained by positing Strongly emergent features. This response problematically presupposes that complex behaviour, if it is to be physically acceptable, must be smoothly aggregative; but physicalists (reductive or non-reductive) are happy to allow that “quirky” behaviour can come about simply as a result of complexity (as with, e.g., chaotic non-linear systems).

4.2.2 Collapse via lower-level dispositions

A second version of the collapse objection focuses on the question of when a feature is placed at the presumed lower-level of physical goings-on. An early version of this objection is registered by van Cleve (1990), who after arguing that

physically unacceptable emergence represents the best option for making sense of dependent but irreducible higher-level mental features, says of Broad's "in-principle deducibility" account:

There is one more point about Broad's account that needs to be discussed. It could be objected to what has so far been said that there is simply no room for the concept of an emergent property, since for any property P of any whole w , there will always be properties of the parts from which P may be deduced. For example, is it not true of sodium that it comes with chlorine to form a whole having such-and-such properties, including its odor and anything else one might have claimed to be emergent? And from such properties of the parts, may not all properties of the whole be deduced? The answer, of course, is yes; but it is also clear that if properties of this sort are admitted in the "supervenience base", the doctrine of anti-emergence [...] becomes completely trivial. (223–4)

 Taylor (2015) develops this line of thought, observing that Broad took sodium chloride to be Strongly emergent (that is, to have fundamentally novel powers, etc.), on grounds that from complete knowledge of the properties of sodium and chlorine in isolation, or in compounds different from that associated with sodium chloride, one could not deduce that salt will dissolve in water. But, Taylor argues, it seems that dispositional properties are among the features that can be had by the components "in isolation", in which case the characteristic features and associated powers of sodium chloride will be deducible, after all:

This case of emergence "collapses" when [...] dispositional properties are included among the micro-level properties. [...] For example, one of the characteristic properties of sodium chloride is its solubility in water. Accordingly, sodium has the following dispositional property: to generate a compound that is soluble in water when combined with chlorine into sodium chloride. In Broad's terms, this property is a property of sodium "in isolation". [...] The emergent features of the whole $R(A, B, C)$ can obviously be deduced from complete knowledge of the features of the parts A , B , and C and the knowledge that they are arranged as a whole $R(A, B, C)$, so long as the features of the parts include these dispositional properties. (736)

Taylor sees a general problem here for accounts of Strong emergence:

[C]ases of emergence presuppose a distinction between micro-level and macro-level properties. For any purported case of emergence, there are properties that *prima facie* belong to the micro level, but if they are included in the micro level then the purported emergent fails to meet a necessary condition for emergent autonomy. I call these problematic properties collapse-inducing properties because when they are included in the micro level, the purported emergent effectively ‘collapses’, and yet it seems arbitrary to exclude them. [...] This is the problem of *collapsing emergence* (or, for short, *the collapse problem*). (732–733, emphases in the original)

Again, the problem such a “dispositional move” poses for the viability of Strong emergence is clear. Both van Cleve and Taylor focus on Broad’s “failure of deducibility criterion”, but as above, the intended import of this criterion is to track the fundamental novelty of a Strongly emergent feature, as reflected in such a feature’s having powers not had by the lower-level physical features upon which it depends (or, for that matter, any other lower-level physical feature). And notwithstanding that the dispositional features of the “isolated” lower-level entities at issue in van Cleve’s and Taylor’s discussions are, to use O’Connor’s (1994) terminology, at various “removes” from either *P* or *S* (understood as per usual), nonetheless such dispositions call into question the intended fundamental novelty of a Strongly emergent feature. Here again O’Connor’s (1994) discussion is useful in highlighting the deeper concern at issue, which he credits to Sydney Shoemaker (p.c.), according to which one can always insist that purportedly Strongly emergent features are in fact “further (hitherto undetected) micro-properties” which are manifested only in certain complex circumstances.⁵

Previous responses to this version of the collapse objection, due to O’Connor and van Cleve, are less than satisfactory.

⁵The possibility of lower-level dispositions to bring about purportedly Strongly emergent features also poses a further threat to the viability of Strong emergence, according to which such lower-level dispositions threaten the supposed physical acceptability of the base entities and features (see Howell 2009). As prefigured above, I reserve treatment of this issue to §4.2.

O'Connor maintains that it would be “implausible” to posit micro-properties that make their presence known only in highly complex systems, such that it would be ad hoc to do so: “the only motivation one could have for postulating [such a] micro-property is a very strong methodological principle to the effect that one is to avoid emergentist hypotheses at all costs” (98). But it seems clear that in general, dispositions “make their presence known” only when certain conditions are in place, and sometimes such conditions might well involve highly complex states of affairs (here again cases of complex non-linear but presumably physically acceptable behaviour are relevant); so the mere fact that micro-dispositions would manifest in complex circumstances is not enough to show that the collapse-inducing suggestion is either implausible or ad hoc.

Van Cleve suggests that restricting the base features to those that are manifested in non-emergence-engendering combinations might do the trick:

Clearly, some sort of anti-triviality stipulation is required. Perhaps the required work can be done by Broad’s phrase “taken separately and in other combinations,” for one could plausibly refuse to regard the property “forming a whole with such-and-such features when combined with chlorine” as a property of sodium taken separately. (223)

But it is not clear that Broad’s qualification provides a basis for plausibly refusing to regard the property “forming a whole with such-and-such features when combined with chlorine” as a property of sodium “taken separately”; for it is commonly assumed (see, e.g., [Martin 1996](#)) that dispositions can be had by individuals even when the dispositions aren’t being manifested: a vase can be fragile even if it is never broken. Taylor considers another response—namely, to restrict lower-level features to be non-dispositional. But as she correctly notes, this would be overly restrictive, since many uncontroversially lower-level physical features—e.g., having a mass of 5 g—are to some extent dispositional.

4.2.3 Three responses to the collapse objection

I will now present three more satisfactory ways in which a Strong emergentist might respond to the triviality and collapse concerns (again, see [Wilson 2002a](#)

and [Baysan and Wilson 2017](#) for further discussion).

Direct vs. indirect powers

Perhaps the simplest line of response is one that distinguishes between direct and indirect having of powers. Here the Strong emergentist grants that while in cases of Strong emergence there is a loose sense in which P or other lower-level physical features inherit S 's purportedly new power(s) (either in P 's being nomologically sufficient for, or in P or some other lower-level features being disposed to give rise to, S), in a stricter sense S 's novel power(s) are not had or manifested by lower-level features in the same direct or immediate way as they are had or manifested by S . Notwithstanding that P synchronically materially necessitates S , P has these powers only in that P is a (at least nomologically) necessary precondition, in the circumstances, for S , which is the more direct locus of the power. Similarly for lower-level dispositions of isolates which are even further removed, to use O'Connor's terminology, from S than is P : granting that there are such dispositions and that these in some sense refer to S and its novel powers, the notion of disposition here is again simply that of a precondition or precursor of a feature (namely, S) that more directly has the power in question.⁶

Such a strategy seems intuitively well-motivated, given the Strong emergentist understanding of lower-level physical goings-on as being precisely such (at least nomologically) necessary preconditions for Strongly emergent features. There are moreover two ways to substantiate the intuition and associated strategy.

First, the Strong emergentist can appeal to an analogy to temporally extended causal chains: even if each link in the chain is, in the circumstances, nomologically sufficient for the next link, one can nonetheless distinguish more and less direct causes of the end result, and the mere fact that, say, a person lights (or could light) the fuse that leads the fireworks to explode doesn't entail that the explosion

⁶Note the contrast here with Shoemaker's suggestion (discussed in Ch. 3.1.2) that in cases of physical realization (Weak emergence), certain of the powers of a realized feature may be had in a more 'direct' way than as had by the realizing feature. As previously discussed, given physicalist acceptance of *Physical Causal Closure* and, relatedly, the *Token Identity of Powers Condition*, there isn't any room for such a distinction between ways of having powers to get a grip.

isn't a novel phenomenon, or that there is any but an indirect sense in which that person has the power to produce such an explosion. Similarly, the Strong emergentist can maintain, in cases of Strong emergence, the base feature *P* is metaphysically, if not temporally, antecedent to *S* in the chain of feature instantiations potentially leading to the effect associated with *S*'s novel powers.

Second, the Strong emergentist can appeal to an analogy to sets and subsets to make the notion of the synchronic yet indirect having of a power more precise, in order to argue that there are in fact different circumstances associated with *S* and with *P* vis-à-vis the having of the power at issue. As is uncontroversial, powers are individuated, in part, by the circumstances in which they manifest and contribute to the production of a given effect; but just as we can distinguish between a set and its subsets at a time, there seems to be no in-principle reason why we cannot distinguish different sets of circumstances associated even with a single temporal interval (instantaneous or extended). In particular, the Strong emergentist can say that *P* has the power to contribute—nomologically, if not causally—to the production of *S*, in circumstances which do not include the presence of *S*. In virtue of having this power, *P* indirectly has the power to contribute to causing anything that *S* can cause. By way of contrast, *S* has at least one power—its novel power(s)—directly, which is manifest in circumstances *K* which, whatever else they might be or contain, do not include the absence of *S*.

Perhaps the main concern with the direct/indirect having strategy is that there may be some indeterminacy in what counts as direct (as opposed to indirect) having or manifestation of a power, just as there might be indeterminacy as regards what counts as the most temporally proximal (to a given effect) link in a causal chain. Here the Strong emergentist has two responses. First, they can maintain that, as per usual, the presence of indeterminacy or borderline cases needn't undermine the usefulness of a given distinction. Second and relatedly, they can avail themselves of one or other of two strategies for accommodating indeterminacy in properly metaphysical (as opposed to merely semantic or epistemic terms): first, the metaphysical supervaluationism endorsed by Akiba (2004), Barnes (2010), Barnes and Williams (2011), and others; second, the determinable-based account

endorsed by Wilson (2013a) (recently applied by Bokulich (2014), Wolff (2015), and Calosi and Wilson (forthcoming) to the case of quantum metaphysical indeterminacy).⁷

Powers relativized to fundamental interactions

A second response to the collapse problem appeals to an independent way of sorting powers, based in the notion of a fundamental interaction, making room for higher-level features to have powers that are in some sense new, as Strong emergence requires (see Wilson 2002a). It is a scientific truism that powers are metaphysically dependent on one or more fundamental forces or interactions. The power of being able to bond with an electron, in circumstances where one is in the vicinity of a free electron, is grounded in the electromagnetic (or electroweak) interaction, as opposed to the strong nuclear or gravitational interactions. The power of being able to fall when dropped, in circumstances where one is poised above Earth's surface, is grounded in the gravitational force, as opposed to the other fundamental forces in operation. The power of being able to bond with other atomic nuclei in a stable configuration is grounded in the strong nuclear interaction, as opposed to the electromagnetic, weak, or gravitational interactions. The power of being able to sit on a chair without falling through it is grounded (at least) in the gravitational and the electromagnetic interactions. And so on. In grounding the powers bestowed by properties, fundamental interactions explain, organize and unify vast ranges of natural phenomena. Hence Auyang (1999) says, in discussing the currently accepted fundamental interactions:

There are four fundamental interactions. Gravity holds our feet on earth and the earth in orbit; it is responsible for the large-scale properties of the universe [...] Electromagnetism binds electrons and nuclei into atoms and atoms into molecules; it is responsible for all physical and chemical properties of solids, liquids, and gasses. The strong interaction binds quarks into nucleons and nucleons into atomic nuclei.

⁷See Wilson (2016a) for a comparison of metaphysical supervaluationist and determinable-based views.

The weak interaction is responsible for the decay of certain nuclei.
 (46)

The metaphysics of fundamental interactions, treating both what these are and how they serve as a basis for powers, is an underdeveloped area of research, and a full exploration of these interesting questions would take us too far afield (though see *Wilson in progressb* for an initial survey of some plausible answers). Here I will limit myself to saying just enough about these issues to motivate the interaction-based strategy of response to the collapse objection.

To start, the notion of an interaction is a contemporary generalization of the notion of a force: whereas forces are pushes or pulls (or component contributions thereof), now commonly seen as ultimately involving particle exchanges, interactions may involve not just forces but other sorts of interactions, such as particle creations and annihilations. As in the case of (what used to be called) fundamental forces (e.g., gravity, electromagnetism), talk of a fundamental interaction is shorthand for talk of token interactions of a given type, which do or may occur in certain circumstances, that are at least partly constituted by the presence of properties (e.g., charge) that are lawfully associated with the interaction. As above, certain interactions are deemed fundamental, in the sense of providing a metaphysical basis for all other interactions and associated phenomena. What criteria are operative in deeming a given form of interaction fundamental is again a large question; for present purposes what is most crucial is that there is an operational test for the introduction of a novel fundamental interaction—namely, that the posit of the interaction is needed to ‘balance the books’, so to speak, as regards various quantities (e.g., energy) which are taken to be conserved. Hence it was, in particular, that the weak nuclear interaction was introduced in response to seeming violations of conservation laws associated with radioactive decay.

Though the operative test for positing a new fundamental interaction does not hinge, it seems, on any particular metaphysical account of such interactions or how these provide a basis for powers, it may nonetheless be worth registering certain options on this score. To start, it is common to take a given fundamental interaction to either be or be associated with a specific collection of fields.

If fields are understood as objects (or some other kind of entity), then it might be natural to see them as having features and associated powers of their own, in which case one part of the answer to the question ‘how are powers grounded in fundamental interactions?’ would be that certain powers—ones plausibly deemed fundamental—are grounded (to speak schematically) in fundamental interactions in virtue of being associated with features of fundamental fields. The question would remain of how exactly the powers of ordinary entities (objects, systems, etc.) were grounded (again, schematically) in fundamental interactions; and here one answer might be that these non-fundamental powers are second-order powers of fields: powers of fields to produce powers of ordinary entities or their ultimate non-field substantial components (e.g., protons and electrons). A somewhat more lightweight metaphysical variation on this theme would interpret fields not as objects (entities), but as collections of comparatively fundamental features and associated powers at spacetime points or regions; here again one might take powers of ordinary entities or their constituents to be second-order powers of powers of space-time points or regions. And of course, as per usual, there here remain the usual options for understanding talk of powers in more or less heavyweight terms. Independent of further metaphysical details, however, given that the claims that distinct fundamental interactions exist and serve as a foundational basis for the spectrum of diverse powers of ordinary objects are claims in unassailably good scientific standing, a Strong emergentist is within their rights to speak of a feature’s having (or not having) a power, relative to a given set of fundamental interactions.

Of course, physicalists of whatever stripe think that physical interactions are the only fundamental interactions there are, while (as McLaughlin 1992 emphasizes) the Strong emergentist thinks that, in addition, there are one or more non-physical ‘configurational’ fundamental interactions. Strong emergentists can thus grant that, taking both physical and non-physical interactions into account, P has every power S has; but coherently maintain that, when S is Strongly emergent, it will have powers that are ‘new’ relative to those powers of P grounded only in fundamental physical interactions. Such a conception clarifies the sense of nov-

elty at issue in the new power condition in Strong emergence, making explicit that this novelty—hence Strong emergence itself—is interaction-relative, along the following schematic lines:

 *Interaction-relative Strong emergence:* Seemingly higher-level feature S is Strongly emergent from feature P just in case, relative to the fundamental interactions in a set F , if (i) S synchronically materially depends on P , and (ii) S has at least one power that is not identical with any power of P that is grounded only in fundamental interactions in F .

Condition (i) again encodes satisfaction of the usual (synchronic material) dependence condition, understood as involving minimal nomological sufficiency in the circumstances, while condition (ii) refines the new power condition in Strong emergence, making explicit that the sense of ‘new’ at issue adverts in part to a fundamental interaction that is new relative to some specified set (typically, the physical fundamental interactions). Again, the use of ‘grounded in’ at issue here and elsewhere is intended as schematic for some or other specific metaphysical relation, to be further determined (see Wilson 2014). For purposes of characterizing Strong emergence in a way that appropriately contrasts with physicalism, one specifies that the interactions in F are the fundamental physical interactions.

An interaction-relative conception of Strong emergence is clearly in the spirit of the original British Emergentist suggestion that Strong emergence involves what “we may call ‘configurational forces’: fundamental forces that can be exerted only by certain types of configurations of particles” (McLaughlin 1992, 52). And it makes room for there to be Strong emergence in the face of the collapse objection(s). To start: even if, taking all fundamental interactions into account, features of the composing system in some sense inherit all the powers of any features they nomologically necessitate, it remains that higher-level features may be associated with powers that are new, in not being grounded only in the set of physical fundamental interactions. Properly relativized, the novel powers of Strongly emergent features do not collapse. Relatedly, relativizing powers to fundamental interactions provides a principled basis for distinguishing dispositions expressing

mere preconditions for the occurrence of Strongly emergent features from those that are more directly implicated in the having of the novel powers at issue.

One might be concerned that, as with conceptions of emergent features as “surprising”, or with Taylor’s alternative (2015) conception in terms of what is (perhaps only presently and contingently) scientifically unexplained, a relativized conception of physically unacceptable emergence would fail to track anything metaphysically interesting or ‘joint-carving’. Interaction-relative Strong emergence doesn’t have this problem, however: new fundamental interactions are interesting and joint-carving, if any natural phenomena are. A further concern might be that the conception requires realism about fundamental forces and/or interactions, but it is unclear at present what might be problematic about such notions.⁸ As previously, fundamental interactions are plausibly understood as second-order dispositions of fundamental fields (namely, dispositions to give rise to further dispositions of non-field entities), and dispositions are not just familiar but moreover admit of more or less metaphysically lightweight interpretations. In any case, and again independent of exactly how fundamental interactions should be understood, participants to the debates over physicalism or Strong emergence typically take a fallibilist realist stance towards the posits of science, including any fundamental interactions there might be.

Strongly emergent objects

The third available response to the collapse problem is motivated by the thesis that features (properties, etc.) have their powers derivatively on the powers of their bearers, as suggested by Baysan 2016 (see also Baysan and Wilson 2017). Drawing on this idea, the Strong emergentist might maintain that the novelty of powers at issue in Strong emergence can be understood as always involving the coming-into-existence of a new object (more generally, entity; I stick with ‘object’ for continuity with Baysan’s discussion), different from that which is the bearer

⁸See Wilson 2007 for a defence of the reality and irreducibility of Newtonian forces, at least some aspects of which would carry over to defence of the reality and irreducibility of fundamental forces and/or interactions.

of the lower-level base feature P , which object is suited to be the bearer of S understood as having novel powers. Call this ‘the new object strategy’. Indeed, though it is common to assume that what powers an entity has are a matter of what features (and associated powers) it has, one might rather maintain that the association of powers with features is derivative on the association of powers with objects. As [Baysan \(2016\)](#) puts it:

What do we mean when we attribute powers to properties? [...] Being knife-shaped has the power to cut bread—conditionally on being instantiated with certain other properties, of course. When we attribute this power to the property of being knife-shaped, do we really mean that the property itself has this power? Unless we want to identify properties with bundles of powers, I don’t think that we have any good reason to give an affirmative answer to this question. Properties don’t cut bread. Their bearers might. To generalize, properties don’t (literally or fundamentally) have powers; their bearers do. (386)

Accordingly, the Strong emergentist can maintain that the novel powers associated with a Strongly emergent feature are in the first instance powers of a novel object. Such a view provides the basis for a principled response to the collapse objection(s): in cases of the Strong emergence of a feature S , S ’s novel power presupposes the coming-into-existence of a new object, different from the bearer(s) of P . Since, on this approach, powers are derivative on the objects having the powers (and associated features), P would inherit S ’s power only if P were born by the same object as S ; but again, the Strong emergentist can reasonably maintain that this is incompatible with S ’s having the novel power at issue, since new powers require new objects. Effectively, the new object strategy turns the collapse objection on its head: given that Strong emergence requires a novel power, and given that powers of features are derivative on powers of their bearers, Strong emergence requires a novel object to have the power, which is then associated with the emergent feature S . Indeed, the Strong emergentist can implement the strategy even if they don’t agree with Baysan that the powers of features are always derivative on the powers of objects. They can simply maintain that new objects are required to be the bearers of any fundamentally novel powers or features there might be.

One might wonder whether the implied commitment, on the new object strategy, to the existence of distinct but spatiotemporally coincident objects is problematic. Here I think that the Strong emergentist is within their rights to shrug their shoulders. To start, the *prima facie* appearances of emergence encourage such a commitment: special science entities appear to be spatiotemporally coincident with lower-level physical aggregates, individual persons appear to be spatiotemporally coincident with their bodies, and so on. To be sure, there remains controversy over how best to treat seemingly coincident objects, a debate which is often characterized as over the nature of material constitution (see [Wasserman 2017](#) for an overview); but antecedent to the identification of some specific difficulty with Strongly emergent coincident objects, the Strong emergentist who endorses the new object strategy can maintain that their view constitutes one among the many options here.

Another concern with the new object strategy is that it might be seen as avoiding the collapse objection only by giving rise to an “explosion” objection—namely, by committing the Strong emergentist to a form of substance dualism, contra the traditional supposition that emergence of whatever variety is supposed to be compatible with substance (more specifically: physical) monism.

Here the Strong emergentist has two main lines of response. First, they may grant that a Strongly emergent object counts as a new (type of) substance, and moreover one that is in some sense non-physical, but maintain that notwithstanding the traditional characterization of emergence as a form of substance monism, what is most important is that viable forms of such emergence suitably contrast with views on which the additional substances or associated subjects of Strongly emergent features are immaterial or otherwise extremely different from physical substances. A commitment to Strongly emergent objects doesn’t entail anything of this sort. Indeed, the assumed synchronic material dependence of Strongly emergent entities and features on lower-level physical entities and features is typically offered as a basis for contrasting this view with serious (e.g., Cartesian) forms of substance dualism.

Second, the Strong emergentist can deny that from the mere positing of a new

object (entity) a new substance is thereby posited. To start, the claim that all and only objects are substances is controversial, and may be rejected. For example, Lowe (1998, 181) argues that some entities (e.g. surfaces, holes, heaps, events) are objects, in being countable and in having determinate identity conditions, but not substances, since they are not capable of independent existence. More generally, on a conception of substance as capable of independent existence, a Strongly emergent object does *not* count as a substance, since such an object requires the existence of whatever entity or entities are the proper bearers of the base feature. Nomologically as well as conceptually, an emergent entity requires the existence of something different from that entity from which it ‘emerges’. So, there is reason to deny that the new object strategy of response to the collapse objection(s) leads to substance dualism, much less a problematic form of such dualism.

Finally, it is worth noting that although the new object strategy supposes that fundamentally novel features and powers bring new objects in their wake (or *vis-à-vis*),⁹ it doesn’t thereby follow that having Strongly emergent features or novel powers is required for a new object to emerge. Indeed, Weak emergentists also commonly appeal (as per Bedau’s guiding assumption that an emergent entity is one having an emergent feature) that Weak emergence involves the coming-into-existence of new entities, associated with a distinctive subset of the powers of the base entities/features.

4.3 Objection: non-necessity

I turn now to considering and responding to objections to the claim that satisfaction of the conditions in Strong emergence is necessary for physically unacceptable emergence. As in the case of Weak emergence, the necessity claim is bold—but, as I’ll now argue, surprisingly robust.

⁹At least, modulo the availability of the other responses to the collapse objection(s); nothing prevents a Strong emergentist from taking a mixed approach to these objections) that fundamentally novel features and powers bring new objects in their wake (or vice versa)

4.3.1 Epiphenomenalism

Some attempts to characterize a form of physically unacceptable emergence have aimed to do so in ways that suggest that such emergent features might be epiphenomenal, and so fail not only to have a new power (as per the *New Power Condition*), but to have any powers at all. Hence Morris (2014) says that “we could attempt to characterize varieties of noncausal nonphysicalist novelty that are compatible with supervenience” (353), and Chalmers (1996) endorses a form of what he calls ‘naturalistic dualism’ which is naturally seen as involving “a limited form of epiphenomenalism (158–9). Indeed, Chalmers argues that the sort of ‘interactionist dualism’ endorsed by the British emergentists, and which the schema for Strong emergence aims to characterize, doesn’t represent any real advantage over epiphenomenalist naturalistic dualism, as regards conscious experience:

[O]n a close analysis, [interactionist dualism] leaves consciousness as superfluous as before. To see this, note that nothing in the story about emergent causation requires us to invoke *phenomenal* properties anywhere. The entire causal story can be told in terms of links between configurations of physical properties. There will still be a possible world that is physically identical but that lacks consciousness entirely. It follows that at best phenomenal properties *correlate* with causally efficacious configurations. (note 41, 378–9)

The best response to this line of thought, and the associated suggestion that physically unacceptable emergent features might be epiphenomenal rehearses the response I previously gave to the concern that satisfaction of the conditions Weak emergence is compatible with there being non-causal phenomenal aspects of higher-level features. There I argued that it is reasonable to maintain that any phenomenal or qualitative aspects of features would be fully encoded in powers of those features, as per the thesis of Phenomenal Incorporation. There I also argued that the Phenomenal Incorporation thesis is neutral as between a physicalist (Weak emergentist) and non-physicalist (Strong emergentist) conception of qualitative features. As such, neither the Weak nor Strong emergentist need agree that “nothing in the story about emergent causation requires us to invoke *phenomenal* properties

anywhere” such that “The entire causal story can be told in terms of links between configurations of physical properties”. On the contrary, they can reasonably maintain that phenomenal properties are crucially referenced as part of the causal story, and similarly for consciousness more generally (a point to which we will return in Ch. 7).

4.3.2 Supervenience

A common baseline assumption of accounts of emergence, whether of Strong or Weak varieties, is that emergent features “minimally nomological supervene” on base features, such that, at least in worlds with the same laws of nature as actually hold, the occurrence of an emergent feature requires the occurrence of some base feature, and any such base feature necessitates, at least in worlds with the same laws of nature, the emergent feature. A number of more specific conceptions of supervenience are on offer, aimed at precisifying glosses such as that ‘there can be no change in supervenient features without a change in base features’ or (equivalently) that ‘duplicating the base features duplicates the supervenient features’. For present purposes what is most relevant is, first, that supervenience is an abstract modal correlational notion or relation, typically understood as holding between lower-level and higher-level features, and second, that some have thought that physically acceptable and physically unacceptable higher-level features will differ in the strength of their supervenience-based correlations they stand in to lower-level physical goings-on. More specifically, the suggestion is that physically acceptable features will supervene with metaphysical necessity on lower-level physical features, whereas physically unacceptable/Strongly emergent features will supervene with only nomological necessity on lower-level physical features. For example, [van Cleve \(1990\)](#) characterizes Strong emergence as follows:

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

(Here by ‘logical necessity’, Van Cleve has in mind what is now usually called “metaphysical” necessity, along lines of “truth in all possible worlds.”) Chalmers (2006a) endorses a similar conception:

[W]e can say that Strong emergence requires that high-level truths are not conceptually or metaphysically necessitated by low-level truths.
(note 1)

The proposed distinction in modal strength here is sometimes taken to serve as a basis for formulating physicalism (understood as per usual as incompatible with Strong emergence) as the thesis that all broadly scientific goings-on supervene with metaphysical necessity on lower-level physical goings-on. In terms of a global supervenience thesis, for example, the suggestion is that physicalism is true just in case any world duplicating the physical goings-on (including physical laws) duplicates the rest of natural reality:

Physicalism is true of our world iff any world that is a physical duplicate of our world either is a duplicate of our world *simpliciter* or contains a duplicate of our world as a proper part.¹⁰ (Howell 2009, 85)

Such supervenience-based characterizations form the basis for an objection to the necessity of the schema for Strong emergence, according to which a distinction between strength of modal correlations, rather than a distinction between powers, is what is key to making sense of physically acceptable and physically unacceptable emergence.

The supposition that mere modal correlations suffice to distinguish physically unacceptable from physically acceptable dependent features is problematic, however.

To start, as has frequently been observed, Strongly emergent features might supervene with metaphysical necessity on physical goings-on and laws (see Horgan

¹⁰The latter condition is intended to accommodate the intuition that worlds that are just like our world except for the addition of ghosts or immaterial souls and the like would not falsify physicalism as holding in this world.

1993, Kim 1998, Levine 2001, Melnyk 2003, Tye 1995, and Wilson 2005)—that is, Strongly emergent features might accompany physical goings-on and laws in any world where the latter base elements exist. In my (2005), I offer several scenarios illustrating how this might be. One somewhat fanciful but still metaphysically coherent scenario is a version of Malbranchean occasionalism, in which a consistent Malbranchean God brings about certain higher-level features upon the occasion of certain lower-level features in every possible world. Another less fanciful scenario involves a view on which features are essentially individuated by (all) the laws of nature into which they directly or indirectly enter (as is endorsed in, e.g, Shoemaker 1980, Swoyer 1982, and Bird 2001, 2002, and 2007). On such a view, any Strongly emergent features there might be would be metaphysically necessitated by physically acceptable base features. Yet another scenario on which this would be the case is one on which Strongly emergent features involve a non-physical interaction, and all the fundamental interactions are unified.

Such scenarios indicate that neither the distinction between metaphysical and nomological necessity, nor the distinction between supervening or not supervening on physical goings-on in worlds where the actual physical laws are operative, can serve to distinguish synchronically dependent features which are physically unacceptable from those that are physically acceptable.

There are two main lines of resistance to this claim; I address each in turn.

The first proceeds by rejecting the seeming counterexamples on grounds that these violate what is sometimes called ‘Hume’s Dictum’, according to which there are no metaphysically necessary connections between wholly distinct existences. As Stoljar (2001) describes the strategy:

[One] suggestion points out that the problem is only genuine if the cases that generate it are coherent—and are they? One reason against supposing so is that both seem to violate Hume’s dictum that there are no necessary connections between distinct existences. According to [Strong] emergentism, for example, mental and physical properties are metaphysically distinct, and yet are necessarily connected.

Adherence to something like Hume’s Dictum is present, for example, in Noordhof’s (2010) endorsement of a supervenience-based characterization of the dis-

tinction between physically acceptable and physically unacceptable supervenient features:

The intuitive thought is that the reason why emergent dualists [should] reject appeal to metaphysical necessity is that they suppose that some of the target properties determined by narrowly physical property causes are wholly distinct from them, whereas non-reductive physicalists are committed to thinking that they are not. (71)

But appeal to Hume's Dictum is ultimately unsatisfactory by way of response. To start, as Stoljar notes, Hume's Dictum is controversial:

Hume's dictum is itself a matter of controversy, so it is unclear if the cases can be dismissed in this way (see [Jackson 1994](#), [Stalnaker 1996](#), [Stoljar 2010](#), and [Wilson 2005, 2010c](#))

Indeed, post-Humean reasons for believing Hume's Dictum are in short supply. To be sure, if one is a strict empiricist like Hume, who takes the content of our ideas and beliefs to ultimately be a matter of fairly superficial sense experiences, one might well be inclined to endorse Hume's Dictum: any such ideas might, at least in principle, either go together or come apart. But contemporary philosophers are not strict empiricists, and as [MacBride 1999](#) notes, “it is a curious fact that the proponents of the contemporary Humean programme [...] having abandoned the empiricist theory of thought that underwrites Hume's rejection of necessary connections provide precious little by way of motivation for the view” (127). Moreover, in a series of papers (see [Wilson 2010c](#), [Wilson 2010a](#), [Wilson 2015b](#), and [Wilson 2015a](#)) I have considered a number of potential reasons for accepting Hume's Dictum, and argued that none withstands scrutiny.

Moreover, even if Hume's Dictum is accepted, endorsement of this thesis won't sidestep many of the seeming counterexamples, for Strongly emergent features need not be ‘wholly distinct’ from lower-level physical goings-on (see [Wilson 2002a](#) and [Stoljar 2007](#)); for example, Strongly emergent features might share some (though not all) powers with lower-level physical base features. As such, there is no principled route to maintaining a supervenience-based approach to non-reductive realization that proceeds by way of Hume's Dictum.

A more principled response to defending a supervenience-based approach is suggested by Howell (2009), who argues against the seeming counterexamples to this approach that if a Strongly emergent feature S were to be metaphysically necessitated by a lower-level base feature P , then we would have good reason to take P to itself be physically unacceptable:

The basic argument is that if emergence laws are necessary, and the emergent properties are “new” enough to count as non-physical, then the supervenience base will be polluted and will no longer be purely physical. If this is the case, then [supervenience physicalism] will judge an emergence dualist world to be non-physical, because duplicating the purely physical properties will not duplicate the world *simpliciter*. (93)

The general suggestion here is that if feature S arises from lower-level feature P with metaphysical necessity, then one should take P to be individuated, in part, by the disposition to give rise to S ; and if P is so individuated—if part of what it is to be an instance of P is to be disposed to give rise to instances of S —then one should not regard P as a physical property. Howell illustrates his intended point as follows:

If it turns out that part of what makes electrons what they are is that they give rise to ‘unpredictable’ qualitative experiences when in a certain setting, then it seems that electrons are somewhat magical and are at least partly constituted by non-physical dispositions. [...] In such a world, a sort of quasi-panpsychism is true: at least some of the basic stuff in our world is not conscious, but it is infused with mentality in that it is individuated by the brute tendency to produce it. (93–94)

In that case, the possibility of metaphysically necessitated Strongly emergent features poses no threat to characterizing the distinction between Weak and Strong emergence in terms of stronger or weaker modal correlations, since any such features would ‘pollute’ the supervenience base features in such a way that the latter would no longer be properly considered (lower-level) physical, with the consequence that “duplicating the purely physical properties will not duplicate the world *simpliciter* (93).

Howell's argument can be expressed as follows:

1. Strongly emergent features and associated powers are physically unacceptable.
 2. If Strongly emergent features metaphysically supervene on (i.e., are metaphysically necessitated by) base features, then the base features will be disposed to bring about physically unacceptable features (powers), and so themselves be physically unacceptable.
- ∴ Metaphysically supervenient Strongly emergent features pose no threat to characterizing metaphysical emergence in terms of asymmetric supervenience.

The Strong emergentist has two available responses.

The first is due to Morris (2014), according to which “reflection on the base pollution maneuver reveals a new worry about supervenience physicalism”. Morris starts by noting that the proponent of a supervenience-based approach to physicalism or non-reductive realization will, like the Strong emergentist, distinguish the lower-level physical goings-on in the supervenience base from the higher-level features and entities that supervene on this base. In that case, one might naturally wonder whether such higher-level features would also end up ‘polluting’ the base:

The base pollution defense supposes [...] that the physicality of a property may be called into question by that property necessarily giving rise to another property that is not itself physical. But in this case, it may seem that supervenient properties will generally end up “polluting the base”. Why, in other words, does the problem of “base pollution” specifically concern [Strong] emergentism? Why doesn't it threaten the very idea of properties distinct from [lower-level] physical properties supervening on [lower-level] physical properties? (356)

Howell is aware of this issue, and addresses it by saying that if the supervenient properties are not “substantively new”, then there is no difficulty with maintaining that the base properties are (lower-level) physical (93, note 18). But as Morris

notes, the need to provide an account of when some supervenient features are or are not substantially new introduces a new challenge for supervenience physicalism:

Granting that the [Strong] emergentist's base cannot be regarded as physical, some account is needed of the conditions under which a supervenient property is not "substantially new" with respect to subvenient, putatively physical properties. Further, without such an account, a supervenience definition of physicalism will provide no guidance for distinguishing between a physical supervenience base and the kind of polluted base associated with [Strong] emergentism, and at least in this way would appear incomplete. The challenge for a supervenience physicalist is to account for the difference between the polluted base and a physical supervenience base without, in effect, rendering talk of supervenience superfluous [if] the only or best way to mark the requisite distinction appeals to the very resources at work in alternative formulations of physicalism or, likewise, alternative accounts of what it is to be a physicalist about some [higher-level] feature of reality. (357)

Morris then goes on to consider two candidate accounts of when a supervenient property is not 'new enough' to result in 'base pollution'—one appealing to a second-order functionalist account, and one appealing explicitly to satisfaction of the *Proper Subset of Powers Condition*—and to observe that while each account plausibly wards off base pollution, each at the same time delivers a non-supervenience-based account of non-reductive realization—that is, physically acceptable emergence—rendering the appeal to supervenience in characterizing such realization otiose.

I agree with Morris's assessment, and moreover add that (as previously argued) insofar as functionally realized features also satisfy the *Proper Subset of Powers Condition*, the more general moral of his assessment is that supervenience-based accounts of non-reductive realization ultimately do not provide an approach to such realization alternative to one implementing the conditions in Weak emergence.

A second response to Howell's argument is available, according to which the second premise (according to which base features will be rendered physically unacceptable, in being disposed to bring about metaphysically necessitated Strongly emergent features or powers) is false. To start, one need not interpret Howell's illustrative case of electrons giving rise to 'unpredictable' qualitative experiences as showing that if base features or entities metaphysically necessitate higher-level features, then 'part of what makes the [base features] what they are' is that they give rise to these higher-level features; nor need one see his case as showing that if the higher-level features are physically unacceptable, then in virtue of the base features' having dispositions to give rise to physically unacceptable features, the base features will thereby be physically unacceptable. Consider the case I previously offered, involving consistent Malbranchean occasionalism, whereby God brings about mental features on the occasion of certain physical features, and moreover, in virtue of being divinely consistent, does so with metaphysical necessity. In such a case it need not be any part of 'what it is to be' the occasioning physical features that God takes them to be such occasions. Relatedly, one might reasonably maintain that if lower-level physical features do have 'dispositions' to bring about Strongly emergent features, the sense of 'disposition' here is a weak one, reflecting just that the lower-level features are, for whatever reason, metaphysically necessary preconditions for the Strongly emergent features, which sense implies nothing about the natures of the occasioning features that would impugn their physical acceptability.

That said, on certain proposed counterexamples to a supervenience-based approach to metaphysical emergence, the sense in which a lower-level feature might be disposed to produce a Strongly emergent feature would plausibly inform some part of its nature—for example, if (as discussed in Wilson 2005) features are individuated by all of the laws into which they enter. It remains, even here, that the sense of 'disposition' at issue is weak, again reflecting that the lower-level features are metaphysically necessary preconditions for the Strongly emergent features. Still, for such cases it seems more is needed to block the threat of 'pollution'. And what more is needed seems to be some principled way of characterizing the

lower-level features as physical, in spite of their having ‘nature-involving’ dispositions to produce Strongly emergent features.

One strategy for doing this appeals to the operative characterization of physical entities and features operative, according to which these are the entities and features treated (approximately accurately) by present (and in the limit of inquiry, ideal) physics, and which are not fundamentally mental (see Wilson 2006a).

Here it is crucial to appreciate two qualifications of the intended characterization of the physical goings-on as ‘not fundamentally mental’: first, that the restriction on fundamental mentality is intended to apply to entities and features in comparatively non-complex combination, and second, that the restriction on fundamental mentality pertains not to any unmanifested dispositions there may be, but rather only to manifested features of entities that, again, are in comparatively non-complex combination. So long as individual lower-level entities and features do not have or bestow mentality, that they or associated configurations might have dispositions to give rise to Strongly emergent mentality poses no threat to their physical acceptability.

Interestingly, while Howell’s criterion of physicality imposes a restriction on mentality, his case for taking the restriction to be violated depends on ignoring both of these qualifications. Howell’s argument relies on the following necessary condition on something’s being physical:

ND: Something is physical only if it does not ineliminably involve mental features.

It is the supposed violation of ND by the dispositional features at issue in premise 2 of his argument that is supposed to establish that metaphysically necessitated Strongly emergent phenomenal features would undermine the physical acceptability of the dependence base features. But—and notwithstanding that Howell cites my characterization of the physical (2006a) as involving ND as a necessary condition—ND is too broad, in failing to specify that the necessary condition here applies only to physical entities and features in comparatively non-complex combinations, and moreover only to features manifested under such non-complex cir-

cumstances. Since the dispositional features under discussion are not so manifest, they pose no threat to the physical acceptability of lower-level features.

A second strategy for characterizing the lower-level features as physical, even given that they have ‘nature-involving’ dispositions to produce Strongly emergent features, adverts to fundamental interactions. Again, even if, taking all fundamental interactions into account, lower-level physically acceptable features are essentially disposed to bring about certain Strongly emergent features, it remains that the occurrence of these lower-level features is just a matter of physical fundamental interactions, reflecting the Strong emergentist understanding of lower-level physical features as something like synchronically necessary (nomologically or metaphysically, no matter) preconditions for any Strongly emergent features there might be. In short: the occurrence or instantiation of lower-level physical features ultimately relies just on the operation just of fundamental physical interactions, whereas the occurrence or instantiation of a Strongly emergent feature requires the operation of an additional, non-physical fundamental interaction. As in the case of an interaction-based response to the collapse objection(s), fundamental interactions provide a basis for distinguishing lower-level physical from Strongly emergent goings-on, even when these are deeply dispositionally connected.

4.3.3 Epistemic criteria

A final objection to the claim that satisfaction of the conditions in Strong emergence is necessary for physically unacceptable emergence in our target cases is that appeal to one or other epistemic criterion suffices for such characterization. Indeed, characterizations of ‘over and aboveness’ as involving one or other epistemic failure have been common. To start, one might again look to Broad’s (1925) account of emergence as involving an in-principle failure of deducibility:

Put in abstract terms the emergent theory asserts that there are certain wholes, composed (say) of constituents A , B , and C in a relation R to each other; that all wholes composed of constituents of the same kind as A , B and C in relations of the same kind as R have certain characteristic properties; that A , B and C are capable of occurring in

other kinds of complex where the relation is not of the same kind as R ; and that the characteristic properties of the whole $R(A, B, C)$ cannot, even in theory, be deduced from the most complete knowledge of the properties of A , B , and C in isolation or in other wholes which are not of the form $R(A, B, C)$. (61)

And recall Chalmers's (2006a) suggestion that

[W]e can say that Strong emergence requires that high-level truths are not conceptually or metaphysically necessitated by low-level truths.
(note 1)

Even if metaphysical necessitation or supervenience, understood as a purely modal notion, is too weak to characterize physically unacceptable emergence, Chalmers here suggests that a failure of conceptual entailment will do the trick. Finally, consider Horgan's (1993) epistemic account of "superdupervenience"—that is, supervenience of the sort guaranteeing that supervening properties are nothing over and above their physically acceptable base properties:

Horgan's Constraint: Any genuinely physicalist metaphysics should countenance ontological inter-level supervenience connections only if they are robustly explainable in a physicalistically acceptable way.

Conversely, one might suggest, it suffices for physically unacceptable emergence that a dependent higher-level feature fails to be 'robustly explainable in a physicalistically acceptable way'.

Though not uncommon, there are two good reasons to reject 'epistemic failure' characterizations of Strong emergence.

First, these epistemic characterizations are intended by their proponents to track a metaphysical distinction relevant to properly metaphysical emergence. For example, as previously discussed, Broad's epistemic characterization was intended to characterize metaphysical emergence as involving fundamentally novel laws—so-called 'trans-physical laws' and associated fundamentally novel powers. Similarly, the sort of case studies that Chalmers and Horgan offer as illustrating the sort of conceptual entailment or robust explanation that is supposed to be

lacking in cases of Strong emergence involve something like functional or causal realization, which are reasonably seen as satisfying the (metaphysical) conditions in the schema for Weak emergence.¹¹ Conversely, it would seem that, on these accounts, the intended sense in which a dependent feature might *fail* to be conceptually entailed or robustly explainable would, as on Broad's account, reflect the feature's having a new power, as per the schema for Strong emergence. As such, the Strong emergentist might maintain that even if Strongly emergent features were always and distinctively (as compared to Weakly emergent features, in particular) accompanied by certain epistemic failures, it would still be advisable to characterize physically unacceptable emergence in terms of what is metaphysically at issue—that is, in the terms encoded in the schema for Strong emergence.

Second, in any case epistemic failures are not distinctive of physically unacceptable emergence. As I will discuss in more detail in Chapter 5 ('The metaphysical emergence of complex systems'), while it made sense in Broad's day to suppose that in-principle failures of deducibility or predictability from lower-level physical goings-on (including laws) were indicative of new fundamental laws and associated powers, it has since become clear that many clearly physically acceptable dependent goings-on are not deducible, even 'in principle', from lower-level physical goings-on, for reasons having to do not with fundamental novelty but

¹¹Hence Chalmers (2006a) says of heat:

The concept of heat that we had *a priori*—before the phenomenon was explained—was roughly that of ‘the thing that plays this causal role in the actual world.’ Once we discover [*a posteriori*] how that causal role is played, we have an explanation of the phenomenon. (45)

And Horgan (1993) says of water:

Explaining why liquidity supervenes on certain microphysical properties is essentially a matter of explaining why any quantity of stuff with these microphysical properties will exhibit these macro-features [tendency to flow, to assume shape of vessel that contains it, etc.] [...] this suffices to explain the supervenience of liquidity because those macro-features are definitive of liquidity [and because] it seems explanatorily kosher to assume a “connecting principle” linking the macro-features to liquidity, precisely because those features are definitive; the connecting principle expresses a fact about what liquidity is. (579)

rather with sensitivity to initial conditions (a la the ‘butterfly effect’), rendering predictions about such goings-on impossible, even were the resources of the entire universe to be in hand. As such, the proponent of an epistemic characterization of physically unacceptable emergence will need to provide some means of distinguishing unexplained physically unacceptable from unexplained physically acceptable higher-level features.

Here the Strong emergentist can implement a variation on the sort of complaint made by Morris (2014) (see also Melnyk 1999) against supervenience-based accounts of metaphysical emergence, according to which making sense of the contrast between physically acceptable and physically unacceptable emergence requires appeal to specific metaphysical relations, which as I have previously argued plausibly aim to satisfy the conditions in either Weak emergence or in Strong emergence. In particular, it is unclear what distinction might be appealed to here besides one encoding the conditions on the powers in the schema: unexplainable features satisfying the conditions in Weak emergence are physically acceptable, whereas unexplainable features satisfying the conditions in Strong emergence are physically unacceptable. In that case, however, one can dispense with the appeal to lack of explanation (deducibility, conceptual entailment) and simply stick with the metaphysical powers-based conditions in the schemas.

4.4 Concluding remarks

In this chapter I have defended the claim that satisfaction of the conditions in Strong emergence is core and crucial to—and when sensibly filled in, necessary and sufficient for—a higher-level feature’s being metaphysically emergent from a lower-level feature, in such a way that, given the physical acceptability of the lower-level feature, the higher-level feature will be physically unacceptable. I have argued that each objection to the viability of Strong emergence can be answered. Again, each objection admits of at least one response relying only on general considerations, which could be offered by proponents of any of the diverse implementations of the schema, including accounts pitched in terms of fun-

damentally novel laws, powers, forces, or interactions. In a few cases, additional responses are available which rely on features specific to an interaction-relative account. In particular, an interaction-relative account provides the basis for responses to the collapse objection and the associated ‘base pollution’ objection. My sense is that if any of the considered objections requires tweaking the schema for Strong emergence, it is these. If so, the needed relativization to interactions would be straightforward, and hardly fatal to the overall approach; but again, this is a choice point, and how exactly a Strong emergentist chooses to respond to the various objections may depend on further of their commitments. In any case, these results collectively indicate that Strong emergence is not just a viable and indeed attractively robust means of accommodating physically unacceptable emergence, but moreover captures what is core and crucial to this notion.

Chapter 5

Complex systems

In previous chapters I have aimed to answer the first key question driving this book: What, more precisely, is metaphysical emergence of the sort motivated by our target cases, in which synchronic material dependence appears to be coupled with ontological and canonlinearusal autonomy? In Chapter 2, I argued that underlying the seeming diversity associated with the many accounts of such emergence there are two and only two schemas, expressing conditions which are core and crucial to (and when sensibly filled in, necessary and sufficient for) metaphysical emergence of physically acceptable and physically unacceptable varieties:

Weak emergence: Token apparently higher-level feature S is weakly metaphysically emergent from token lower-level feature P on a given occasion just in case, on that occasion, (i) S synchronically materially depends on P , and (ii) S has a (non-empty) proper subset of the token powers had by P .

Strong emergence: Token apparently higher-level feature S is strongly metaphysically emergent from token lower-level feature P on a given occasion just in case, on that occasion, (i) S synchronically materially depends on P , and (ii) S has at least one token power not identical with any token power of P .

In Chapters 3 and 4, I moreover defended each of these schemas for metaphysical emergence against a battery of objections, with the upshot being that each schema

is not just a viable and indeed attractively robust means of accommodating emergence of the physically acceptable or unacceptable variety at issue, but moreover captures what is core and crucial to this notion.

Correspondingly, we are now in position to investigate into the second key question driving this book: Is there actually any metaphysical emergence? In this and the following chapters, I will consider this question as applied to four kinds of phenomena that have frequently been taken to involve metaphysical emergence: complex systems, ordinary objects, consciousness, and free will.

I begin with complex systems, as perhaps the phenomena that have most often been offered as being emergent, by scientists as well as philosophers. Such systems take many forms, both natural (as in cases of turbulent water flows, phase transitions, and weather patterns) and artificial (as in the ‘Game of Life’, to be discussed in more detail in §5.2.1). While there is no agreed-upon definition of what it is to be a complex system, among the commonly highlighted features of such systems, especially of the ‘chaotic’ variety, are as follows:

- *Nonlinearity*: Complex systems are nonlinear, in that certain of their features (including associated powers and behaviors) cannot be seen as linear or other broadly additive combinations of features of the system’s composing entities, and relatedly, in that mathematical expressions describing the evolution of such systems contain nonlinear terms.
- *Extreme sensitivity to initial conditions*: For ‘chaotic’ complex systems, small differences in initial conditions can result in huge differences in the trajectory or behaviour of the system.
- *Unpredictability*: The precise behaviours of complex systems are unpredictable and relatedly, often surprising and seemingly novel.
- *Algorithmic incompressibility*: The dynamical equations of complex systems do not admit of analytic or “closed” solutions.
- *Universal behaviours*: Compositionally different systems may exhibit highly similar behaviour.

- *Self-organization:* Complex systems exhibit coherent patterns arising as a result of interactions among the parts, in a way suggesting that they are ‘self’-organizing.

Each of these features has been offered as supporting the claim that a given complex system is either Strongly or Weakly emergent. As I’ll argue in what follows, however, previous discussions of these features as indicative of metaphysical emergence do not succeed in establishing the existence of such emergence, and in particular, do not successfully rule out the possibility that complex systems are ontologically reducible to lower-level physical goings-on. That said, I will eventually argue that attention to another distinctive feature of many complex systems—being such as to have strictly fewer degrees of freedom than those associated with the systems of their composing entities—can serve as a principled basis for establishing that some complex systems are actually Weakly emergent.¹

5.1 Are complex systems Strongly emergent?

As above, complex systems are both nonlinear and unpredictable—features which the British emergentists took to be related and sufficient for Strong emergence. In this section, I provide a compressed historical discussion of why the British emergentists took these features to be sufficient for fundamental novelty—a view that, while reasonable at the time, was undermined by the discovery or creation of complex systems clearly not involving any new fundamental interactions or laws. This broadly historical discussion is useful for four reasons: first, for purposes of appreciating how nonlinearity moved from being a criterion of Strong emergence to being a criterion of Weak emergence; second, for purposes of identifying and putting aside certain non-starter conceptions of nonlinearity as a criterion of any form of emergence; third, for purposes of appreciating the typically proffered reasons for thinking that complex systems are at best Weakly emergent don’t withstand scrutiny; and fourth, for purposes of seeing how a recognizable descendant

¹In this chapter I draw on and extend Wilson 2010b and Wilson 2013b, in particular.

of nonlinearity as a criterion of Strong emergence is present in the aforementioned scientifically motivated criterion of a new fundamental interaction, in the form of a (seeming) violation of a conservation law. By lights of the latter criterion, I observe, there is little motivation for seeing non-mental complex systems as Strongly emergent—though the case is less clear for certain mental phenomena, a topic to which I return in Chapter 7.

All this said, the reader primarily interested in whether any complex systems are actually Weakly emergent can cut to the next section of this chapter without undue loss of continuity.

5.1.1 Nonlinearity and unpredictability in the British emergentist tradition

An early suggestion that nonlinearity was indicative of metaphysical emergence is found in Mill (1843/1973, Chapter X, ‘On the Composition of Causes’). Here Mill distinguishes two types of effects of joint or composite causes (i.e., causes operating together, rather than in isolation or in tandem). First are “homopathic” effects, which conform to “the principle of composition of causes” in being (in some sense) mere sums of the effects of the component causes when acting in relative isolation, as when the weight of two massy objects on a scale is the scalar sum of their individual weights, or when the joint operation of two forces conforms to vector addition in bringing an object to the same place it would have occupied had the forces operated sequentially. Second are “heteropathic” effects, which, by way of contrast, violate the principle in not being mere sums in any clear sense. As Mill sees it, this distinction is crucial, in that (he supposes) the advent of heteropathic—i.e., non-additive or more generally, nonlinear—effects is indicative of the operation of new laws:

This difference between the case in which the joint effect of causes is the sum of their separate effects, and the case in which it is heterogeneous to them; between laws which work together without alteration, and laws which, when called upon to work together, cease and give

place to others; is one of the fundamental distinctions in nature. (408–409)

And by way of illustration, Mill offers chemical compounds and living bodies as entities capable of producing heteropathic effects.

Now, Mill did not use the term ‘emergence’ (as noted, Lewes was evidently the first to do so, in his 1875), and his discussion appears to target a diachronic notion of emergence rather than the broadly synchronic emergence seemingly at issue in the target cases. But given the reciprocal connection between powers and effects (as discussed in Ch. 1, ‘Preliminaries’), it is straightforward to translate Mill’s talk of effects into talk of powers: to say that an effect of a feature of a composite entity is heteropathic (i.e., nonlinear), relative to effects of features of its composing parts acting separately, is just to say that the feature of the composite (the ‘higher-level’ feature) has a power not had by its lower-level base features when in linear combination. Mill himself moves seamlessly from talk of heteropathic effects to talk of new properties of and laws governing entities capable of causing such effects:

[W]here the principle of Composition of Causes [...] fails [...] the concurrence of causes is such as to determine a change in the properties of the body generally, and render it subject to new laws, more or less dissimilar to those to which it conformed in its previous state. (1843/1973, 435)

Both Mill’s reference to “new laws” and his taking such cases to contrast with “the extensive and important class of phenomena commonly called mechanical” indicate that Mill’s appeal to the nonlinearity of effects is aimed at identifying a criterion for a higher-level feature’s having a new fundamental power, enabling it (or its possessing “body”) to override the usual composition laws in the production of certain effects. As McLaughlin (1992) notes, “Mill holds that collocations of agents can possess fundamental force-giving properties” (65). Hence it is that Mill’s conception is appropriately seen as a variety of Strong emergence,² of the

²A minor complication here is that Mill’s treatment seems to presuppose a many-one approach

sort that, were it to exist, would falsify physicalism, or “mechanism”, as it was more commonly called in Mill’s day.³

Other British Emergentists followed Mill in characterizing Strong emergence in terms of nonlinearity, as involving violations of broadly additive composition laws, including Alexander (1920), who characterized emergent properties as having powers to produce heteropathic effects; Morgan (1923), who contrasted resultant with emergent features as being “additive and subtractive only”; and Broad (1925), who offered scalar and vector addition as paradigms of the compositional principles whose violation was characteristic of emergence. As in Mill’s case, these appeals to nonlinearity are best seen as attempts to provide substantive metaphysical criteria of the operation of new fundamental powers, forces, or laws that come into play only at certain complex levels of existence, as when Broad (1925) says, “[T]he law connecting the properties of silver-chloride with those of silver and of chlorine and with the structure of the compound is, so far as we know, an unique and ultimate law” (64–5). As such, McLaughlin accurately characterizes British Emergentism as

[...] the doctrine that there are fundamental powers to influence motion associated with types of structures of particles [...] In a framework of forces, the view implies that there are what we may call “configurational forces”: fundamental forces that can be exerted only by certain types of configurations of particles [...] . (1992, 52)

Before continuing, it is worth noting that the operative notion of nonlinearity in the British emergentist tradition was *not* one according to which it sufficed

to emergence rather than the one-one approach in the schema for Strong emergence. For reasons not unlike those discussed in Ch. 1 in re Gillett’s (2002a) approach, nothing very deep turns on this issue. The key issue, for Mill, concerns whether complex phenomena ever bring anything fundamentally causally novel in their wake; and this issue can be framed, *mutatis mutandis*, in one-one terms comparing powers of lower-level configurations, understood as building in operative linear or other lower-level combinations or relations, to powers of higher-level ‘composites’.

³See Crane and Mellor 1990 and Wilson 2006a for discussion of the historical transition from ‘materialism’, a view on which the characteristic features of the foundational entities (as, e.g., extended and governed by deterministic laws) were specified largely *a priori*, to ‘physicalism’, on which the foundational goings-on were rather to be specified, *a posteriori*, by physics (albeit, I argue, reasonably subject to the constraint that the fundamental (comparatively non-complex) physical entities and features cannot individually have or bestow mentality).

for nonlinearity (hence, on their view, fundamental novelty) that a feature of a composite entity fail to be a linear combination of *intrinsic* features (powers, etc.) of its composing entities—i.e., a linear combination of features of its composing entities when “in relative isolation”. Such a contrast would render the view immediately implausible. Consider, for example, the shape of a molecule composed of some atoms. The molecule’s shape (and associated powers, etc.) is presumably not the product of any new fundamental interactions; but on the other hand, this shape is clearly not a function (linear or otherwise) just of the shapes of the atoms when in relative isolation—also involved are the bonding relations between the atoms holding these at some distance from each other.

In light of such cases, the most sophisticated and careful of the British emergentists—namely, Broad—included pairwise and other relatively non-complex relations between the lower-level entities (or states of affairs consisting of lower-level entities standing in relatively non-complex lower-level relations) as among the physically acceptable “summands” apt to be combined in broadly additive fashion, and against which a given claim of emergent nonlinearity was to be assessed. Hence it was that Broad characterized “pure mechanism” as involving broadly additive deducibility of all higher-level features from features of lower-level entities “either individually or in pairwise combination”, and couched his official formulation of emergence in terms of failures of in-principle deducibility of higher-level features from features of lower-level entities both “in isolation” and “in other wholes”. The notion of linearity in these construals clearly adverts to lower-level relational as well as intrinsic features of lower-level entities. Such an appropriately broad understanding of linearity can accommodate, e.g., the shape of a molecule, as a (vector) additive function of shapes associated with atoms in pairwise or other relatively non-complex combination.

Closely related to these appeals to nonlinearity as indicative of fundamental novelty is the British emergentist supposition that unpredictability, understood to be in some sense ‘in-principle’, was also indication of such novelty. Indeed, as discussed in Ch. 2, notwithstanding that Broad clearly aimed to characterize a Strong (anti-mechanistic or anti-physicalist) conception of metaphysical emergence, his

‘official’ characterization of emergence was couched in epistemic terms involving in-principle unpredictability:

The emergent theory asserts that there are certain wholes, composed (say) of constituents A , B , and C in a relation R to each other and that the characteristic properties of the whole $R(A, B, C)$ cannot, even in theory, be deduced from the most complete knowledge of the properties of A , B , and C in isolation or in other wholes which are not of the form $R(A, B, C)$. (1925, 64)

Though this formulation is in epistemological terms, the discussion preceding the formulation makes clear that Broad’s appeal to failure of deducibility aims to characterize a metaphysical notion of emergent autonomy, and moreover one tracking a form of emergence incompatible with complete lower-level determination. Similar remarks go for Alexander’s remarks that emergent phenomena must be accepted with ‘natural piety’—i.e., as failing to be deducible, in either a theoretical or metaphysical sense, from lower-level goings-on.

5.1.2 The fall of nonlinearity and unpredictability as guides to fundamental novelty

Given an appropriately sophisticated understanding of the notion of linearity at issue, it was quite reasonable for the British Emergentists to take nonlinearity to be a mark of Strong emergence. To start, various paradigm cases of non-emergent features of composite entities (which might be lower-level configurations, or Weakly emergent higher-level entities) are simple scalar sums of features of their composing entities, as when the mass of a composite entity is the sum of the masses of its composing entities. More generally and more importantly, at the time it was common to suppose that effects ultimately involve the exertion of various fundamental forces, including gravity and electromagnetism, operating either singly or together. Moreover, as Mill’s discussion makes especially clear, the combination of fundamental forces was taken to proceed in accord with linear composition laws—that is, by means of vector addition. As such, and again recalling the corre-

spondence between powers and effects, linearity looked to provide a general handle on when the features of composite entities did not involve or invoke any new fundamental powers, forces, or laws. Conversely, failures of features or behaviors of composite systems to be subject to linear analysis would thus have been reasonably interpreted as indicating that some additional fundamental force—a force not operative at lower, less complex levels of natural reality—was now on the scene.

Though reasonable at the time, the British Emergentist supposition that nonlinearity is sufficient indication of the sort of fundamental novelty at issue in Strong emergence is no longer plausible. For one thing, the picture of causal relations as constituted by the operation of additive combinations of fundamental push-pull forces is now seen as largely heuristic, or in any case not generalizable; it is fundamental interactions, involving particle exchanges, or yet more abstract accounts of the existence and evolution of natural phenomena, that provide the ultimate story as regards the “go” of events, and to the extent that broadly Newtonian forces can be seen as real (as they arguably can be; see Wilson 2006) they are now assumed to be non-fundamental (as, e.g., constituted by fundamental interactions). For another, in the course of the 20th century, investigations into a wide range of complex systems revealed not just that nonlinearity was rampant, but that in many of these cases the nonlinearity was generated in ways that clearly did not involve the positing of any novel fundamental interactions.

The recognition of many complex composite systems as genuinely nonlinear proceeded along several fronts. Even as early as the late 1880s, there were difficulties in seeing chaotic complex systems of the sort associated with turbulence in fluids and gasses, and with phase transitions, as linear. Attempts were made to explain away failures of linear prediction in these cases as due to noise or imprecision in measurement; but in a nice recapitulation of the move from a Ptolemaic to a Copernican system of astronomy, the anomalies and epicycles associated with the supposition of linearity eventually gave way to an understanding of complex systems as being genuinely nonlinear. This is not to say, of course, that failures in prediction were thereby (always) overcome; rather, such failures were given an alternative explanation as reflecting, most saliently, the typical highly sensitive

dependence of the associated nonlinear functions on initial conditions (a.k.a. “the butterfly effect”).

That the predictive anomalies in some complex systems could not generally be put down just to noise or imprecision was confirmed by attention to natural and artificial nonlinear systems for which the relevant initial conditions could be specified with complete accuracy. Population growth, for example, is straightforwardly modelled by the nonlinear logistic map:

$$x_{n+1} = ax_n - ax_n^2$$

Here a is a parameter representing birth and death rates, and is different for different systems. The behavior of a given system is heavily dependent on a . For most values of a , the system evolves to a fixed point; as a approaches 4, the system’s behavior becomes periodic, and subject to increasingly rapid bifurcation; for $a = 4$, the system’s behavior becomes chaotic, with very small differences in initial conditions x_i , associated with distant decimal places, eventually leading to wildly different trajectories. The discovery of natural nonlinear systems encouraged attention to nonlinear systems in general, and with the advent of computers in the latter half of the 20th century, much attention focused on artificial complex systems such as cellular automata, where, as in Conway’s ‘Game of Life’, the stipulated dynamics are nonlinear.

The recognition of such genuinely nonlinear systems fatally undermined the British Emergentist supposition that nonlinearity is sufficient for Strong emergence, in that at least some cases of nonlinear complex systems—e.g., those associated with population growth or cellular automata—clearly do not involve any additional fundamental forces or interactions, or associated novel powers (in the case of automata, as a matter of explicit stipulation).

Similar remarks apply to the British Emergentist supposition that unpredictability is a sufficient indication of the fundamental novelty at issue in Strong emergence. It was reasonable enough at the time to assume that an insuperable failure of predictability would be an epistemic marker of such novelty: after all, if even a Laplacian demon couldn’t deduce the higher-level phenomena from the lower-level goings-on, then what else, besides something fundamentally new, could ex-

plain the occurrence of the higher-level phenomena? But again, increased awareness of and appreciation of the distinctive features of complex systems undermined the supposition. In particular, given the extreme sensitivity to initial conditions associated with chaotic complex systems, it turns out to be in-principle impossible to deduce the features of such systems, at least if it counts as a failure of ‘in-principle’ deducibility that these features could not be deduced even given the resources of the entire universe.

5.1.3 Are all nonlinear phenomena at best Weakly emergent?

In light of the previous considerations, it is often taken for granted that Strong emergence is not at issue in cases of nonlinear complex systems. Hence [Bedau \(1997\)](#) says:

An innocent form of emergence—what I call “weak emergence”—is now a commonplace in a thriving interdisciplinary nexus of scientific activity [...] that includes connectionist modeling, nonlinear dynamics (popularly known as “chaos” theory), and artificial life. (375)

As I’ll discuss in the next section, there is indeed a good case to be made that many complex systems are, at best, Weakly emergent. Still, it is worth noting that stated reasons for thinking that complex systems are *always* physically acceptable aren’t compelling.

[Newman \(1996\)](#) cites the fact that such systems are “strictly deterministic” in support of this claim; but nothing prevents Strongly emergent features from entering, both as regards their emergence and their subsequent evolution, into a deterministic nomological net. Nor does the fact that features of complex systems are “derivable” from nonlinear equations and initial (or boundary) conditions establish physical acceptability, since—as the British Emergentist tradition makes explicit—unlike linear combinations, that nonlinear combinations of physically acceptable features are themselves physically acceptable is not obvious. Bedau (1997) claims that features of complex systems are physically acceptable because “structural”—that is: are features of a relational system consisting in composing

entities standing in lower-level relations; but given that the features of such a relational entity do not consist solely in additive combinations of features of the parts, that such features are merely structural (as with, e.g., the shape of a molecule), in a sense that would entail physical acceptability, is again not obvious. One might aim to support the general claim via an argument by analogy, maintaining that insofar as various surprising features of complex systems (period doubling, extreme sensitivity to initial conditions) can be modeled in comparatively simple and artificial systems for which it is uncontroversial that no new fundamental novelty is at issue, there is no reason to suppose that more complex natural complex systems involve fundamental novelty, either. But this argument by analogy fails, for precisely what is at issue is whether, in the more complex natural cases, the nonlinear behaviours at issue have a physically acceptable source.⁴

Indeed, there is in-principle room for maintaining that Strong emergence is at issue in at least some cases of complex systems. Consider, for example, cases where the nonlinear phenomena involves feedback between the micro-entities constituting the base, associated with strange attractors and other dynamic phenomena. As [Silberstein and McGeever \(1999\)](#) note, the nonlinearity at issue in complex systems might be taken to involve a kind of system-level holism:

What is the causal story behind the dynamics of strange attractors, or behind dynamical autonomy? The answer, it seems to us, must be the nonlinearity found in chaotic systems. [...] But why is non-linearity so central? [...] Non-linear relations may be an example of what Teller calls ‘relational holism’ [...].⁵ (197)

Silberstein and McGeever go on to suggest that relational holism of this sort might reflect emergent features’ possessing fundamentally new powers (“irreducible causal capacities”), in line with Strong emergence.

⁴The thought here is not so different from that undercutting the supposition that Strong emergence has been generally discredited by scientific advances; from the fact that some previous candidates for Strong emergence are no longer such candidates, it doesn’t follow that all relevant phenomena will admit of treatment in physically acceptable terms.

⁵In re relational holism, see also the provisional definition of emergence offered by [Thompson and Varela \(2001\)](#).

It has also been suggested (or interpreted as being suggested) that the singularities standardly associated with thermodynamic phase transitions are indicative of Strong emergence. Hence Menon and Callender (2013) interpret Batterman's claim that "thermodynamics is correct to characterize phase transitions as real physical discontinuities and it is correct to represent them mathematically as singularities" (Batterman 2005, 234) as signaling Batterman's commitment to phase transitions' being emergent along British Emergentist lines. Whether Teller and Batterman would agree that relational holism or thermodynamic singularities should be understood as involving new fundamental powers/interactions/laws is disputable (as I'll shortly discuss in re Batterman's view). Still, for present purposes the crucial point is that one *could* coherently take Strong emergence to underlie some features associated with some complex natural nonlinear systems.

Let's sum up the results so far.

First, whether or not all complex systems are physically acceptable, in any case there's no doubt that some are. The general moral to be drawn from the identification of straightforwardly mechanistic and artificial complex systems is that, contra Mill, Broad, and the other British Emergentists, neither nonlinearity nor (even 'in-principle') unpredictability are sufficient indication of fundamental higher-level powers/interactions/laws.

Second, though many complex systems clearly do not involve Strong emergence, stated reasons for the general claim that all complex systems are at best Weakly emergent are unconvincing; as it stands, the general claim is something of an article of faith. It would be nice if, given that nonlinearity and in-principle unpredictability can no longer be seen as criterial of Strong emergence, there were an alternative criterion which could distinguish physically acceptable from physically unacceptable cases of complex systems (assuming any of the latter exist). I turn now to identifying such a criterion, as it enters into my preferred account of Strong emergence.

5.1.4 Nonlinearity's descendant

As discussed in the last chapter, in [Wilson 2002a](#) I offer an account of Strong emergence along British Emergentist lines, which fills in the sense in which the power of a Strongly emergent feature is new in terms which explicitly (as opposed to implicitly, as on Broad's characterization in terms of 'in-principle failure of deducibility') involve the coming into play of a new fundamental force/interaction. As above, at the heart of the Strong emergentist position is that some composite systems have features associated with new powers, grounded in new fundamental forces or interactions, suited to enter into producing effects that the system of lower-level composing entities can't (directly) enter into producing. [McLaughlin \(1992\)](#) claims that those suspicious of forces (and presumably also of interactions) can dispense with this aspect of the view, retaining only the appeal to new powers. But as we saw in the previous chapter, there is some motivation for thinking that an appeal to fundamental forces or interactions provides perhaps the best response to (among other objections) the "collapse" objection(s), according to which the supposedly novel powers associated with Strongly emergent entities or features are always inherited by the lower-level dependence base entities or features. Recall that on this strategy of response, the collapse objection is avoided by understanding Strong emergence in interactive-relative terms:

Interaction-relative Strong Emergence: Feature S is strongly emergent from feature P just in case, relative to the fundamental interactions in F , if (i) S synchronically materially depends on P , and (ii) S is associated with at least one power that is not identical with any power of Q that is grounded only in fundamental interactions in F .

Again, *Interaction-relative Strong Emergence* makes room for there to be Strong emergence: even if, taking all fundamental interactions into account, features of the lower-level object or system inherit all the powers of any features they nomologically necessitate, it remains that higher-level features may be associated with powers that are relevantly "new", in not being grounded *only* in the set of physical fundamental interactions.

Interaction-relative Strong Emergence also has three features relevant to investigating the bearing of the sort of nonlinearity characteristic of complex systems on Strong emergence.

First, the account nicely accommodates the supposition that features of a composite entity that can be analyzed as broadly additive combinations of physically acceptable features of the composing entities will not be Strongly emergent. For whatever the precise account of how powers are grounded in fundamental interactions, in any case it is clear—to go back to Mill’s original discussion—that every power of a feature that is a broadly additive combination of physically acceptable features will be grounded only in fundamental physical interactions, hence fail to be Strongly emergent.

Second, the account suggests an alternative criterion for Strong emergence, which criterion not only survives the advent of physically acceptable nonlinear systems, but moreover provides the means of distinguishing, at least in principle, between cases of nonlinearity that do and cases that don’t involve Strong emergence. Here I have in mind the aforementioned criterion of a new fundamental interaction (operative, for example, in the posit of the Weak nuclear interaction), adverting to apparent violations in conservation laws. A similar strategy makes in-principle empirical room for testing whether or not the unusual features associated with complex natural nonlinear systems are or are not due to configurational fundamental interactions, by comparing the values of relevant conserved quantities predicted by fundamental physical theory as attaching to composite entities, with the actually observed values of these quantities. Again: if there’s less (or more) energy coming out than going in, for example, we might well be inclined to conclude, following accepted scientific procedure and as per the Strong emergentist thesis, that a new configurational force/interaction is in operation.

Third, in the appeal to apparent violations of conservation laws as a sufficient criterion of Strong emergence we have, it seems to me, a recognizable descendant of the British Emergentist appeal to apparent violations of linearity (or predictability) as such a criterion. For an apparent violation of a conservation law serves, as an apparent violation of a linear composition law was reasonably but incorrectly

taken to do, to flag that the whole is more than the mere sum of its parts, such that some fundamentally novel powers (forces, etc.) and laws must be posited, if the sum—of forces, of conserved quantities—is to come out right.

All this said, I take it that there is not much motivation for thinking that any complex natural nonlinear systems involve new fundamental interactions, with the possible exception of those systems associated with qualitative consciousness and free choice (to which we will return in Chapters 7 and 8). Still, it is useful to observe, first, that there is a criterion for Strong emergence upon which (and unlike bare appeals to nonlinearity, relational holism, or representational mismatch) all parties are likely to agree, and which could, in principle, be tested for, and second, that the Strong emergence of any and all complex systems is as yet not empirically discredited, and second, by lights of this criterion the Strong emergence of at least some complex systems has not been empirically discredited.

5.2 Are complex systems Weakly emergent?

I now want to turn to the roles that nonlinearity, unpredictability, and other features of complex systems have played in a representative range of accounts of Weak emergence taking such systems as their starting point. As above, accounts of Weak emergence aim to combine the dependence of a higher-level system and its features on lower-level goings-on with the higher-level system's being ontologically and causally autonomous, in a way compatible with physicalism. But as I'll argue, with the exception of the degree of freedom-based (DOF-based) account discussed previously, accounts of the sort of physically acceptable emergence at issue in complex systems have thus far been compatible with the ontological (hence causal) reducibility of the complex systems at issue, with the upshot being that such accounts at best motivate taking complex systems to be epistemologically or representationally emergent.

I'll establish this for three prominent such accounts, targeting three different kinds of complex system: first, Bedau's account of Weak emergence as involving algorithmic or explanatory incompressibility, applied to properties such as being

a glider gun in the Game of Life; second, Mitchell’s account of emergence as involving self-organization, applied to group behaviours such as the flocking of birds; and third, Batterman’s account of emergence as involving asymptotic universality, as applied to thermodynamic systems undergoing phase transitions.⁶ I’ll then present my (2010b) account of Weak emergence as involving an elimination in degrees of freedom (DOF) associated with the complex system, as compared to the unconstrained system of its composing base entities, and argue that at least some nonlinear systems are Weakly emergent, by lights of the DOF-based account.

5.2.1 Bedau’s appeal to incompressibility

Bedau’s focus is on a feature of nonlinear systems shared by both chaotic and nonchaotic nonlinear systems; namely, that such systems typically fail to admit of analytic or “closed” solutions. The absence of analytic or otherwise “compressible” means of predicting the evolution of such systems means that the only way to find out what this behavior will be is by “going through the motions”: set up the system, let it roll, and see what happens. It is this feature—namely, algorithmic incompressibility, understood as characterizing a kind of unpredictability—that serves as the basis for Bedau’s (1997) account of Weak emergence, as follows:

Where system S is composed of micro-level entities having associated micro-states, and where microdynamic D governs the time evolution of S ’s microstates: Macrostate P of S with microdynamic D is weakly emergent iff P can be derived from D and S ’s external conditions but only by simulation. (378)

⁶I do not discuss Newman’s (1996)’s account of emergence, as directed at the feature *being in the basin of a strange attractor*, since his account (according to which such features are identical to lower-level features, albeit in such a way that it is ‘epistemically impossible’ for us to discover) is explicitly ontologically reductive. Newman’s account is representative of approaches to the emergence of nonlinear systems according to which such emergence is solely epistemological (see also Popper and Eccles 1977, Klee 1984, and Rueger 2001). However illuminating such accounts may be as regards why we find the features and behavior of such systems interesting, novel, or unpredictable, they do not provide any basis for taking complex systems to be even Weakly metaphysically emergent. The accounts I will consider in the text aim, or can be seen as aiming, to provide such a basis.

Derivation of a system's macrostate "by simulation" involves iterating the system's micro-dynamic, taking initial and any relevant external conditions as input. The broadly equivalent conception in Bedau's (2002) takes physically acceptable emergence to involve "explanatory incompressibility", where there is no "short-cut" explanation of certain features of a composite system. In being derivable by simulation from a micro-physical dynamic, associated macrostates are understood to be physically acceptable; as Bedau says, such systems indicate "that emergence is consistent with reasonable forms of materialism" (1997, 376).

By way of illustration, Bedau focuses on Conway's Game of Life, an example of a non-chaotic nonlinear map. The game consists in a set of simple rules, applied simultaneously and repeatedly to every cell in a lattice of "live" and "dead" cells. Here there is no problem of sensitivity to initial conditions, since these conditions consist just in the discrete "seeding" of the lattice. Still, Bedau argues that the property of being a glider gun in the Game of Life is Weakly emergent, in his sense. That this property does not involve any Strong emergence is clear, since for cellular automata the long-term behavior of the system is completely metaphysically determined by ("derived from") the lower-level "rules" applying to cells in the grid. But that a given system will evolve in such a way as to generate a glider gun can typically not be predicted from knowledge of initial conditions (seeding) and these rules.

On the face of it, Bedau's account, like Newman's (see note 4), does not characterize a genuinely metaphysical account of physically acceptable emergence, an impression seemingly confirmed when Bedau says that "weakly emergent phenomena are autonomous in the sense that they can be derived only in a certain non-trivial way" (2002, 6). Indeed, Bedau is explicit that he takes emergent features of composite systems to be both ontologically and causally reducible to features of their composing systems:

[W]eakly emergent phenomena are ontologically dependent on and reducible to micro phenomena. (2002, 6)

[T]he macro is ontologically and causally reducible to the micro in principle. (2008, 445)

Notwithstanding these reductive assumptions, Bedau maintains that the autonomy of Weakly emergent entities, on his account, is not just epistemological, but is also properly metaphysical. He offers two reasons for thinking this, but as I'll now argue, neither establishes the point.

The first is that the incompressibility of an algorithm or explanation is an objective metaphysical (if broadly formal) fact:

The modal terms in this definition are metaphysical, not epistemological. For P to be weakly emergent, what matters is that there is a derivation of P from D and S's external conditions and any such derivation is a simulation. [...] Underivability without simulation is a purely formal notion concerning the existence and nonexistence of certain kinds of derivations of macrostates from a system's underlying dynamic. (1997, 379)

But such facts about explanatory incompressibility, though objective and hence in some broad sense “metaphysical”, are not suited to ground the metaphysical autonomy of emergent entities. What is needed for such autonomy is not just some or other metaphysical distinction between macro- and micro- goings-on, but moreover one which plausibly serves as a basis for rendering weakly emergent features ontologically autonomous from—that is, distinct from—the lower-level features upon which they depend.

The second reason Bedau gives is somewhat more promising; namely, that Weakly emergent features typically enter into macro-level patterns and laws. As Bedau says:

[T]here is a clear sense in which the behaviors of weak emergent phenomena are autonomous with respect to the underlying processes. The sciences of complexity are discovering simple, general macro-level patterns and laws involving weak emergent phenomena. [...] In general, we can formulate and investigate the basic principles of weak emergent phenomena only by empirically observing them at the macro-level. In this sense, then, weakly emergent phenomena have an autonomous life at the macro-level. (1997, 395)

As such, Bedau maintains, “weak emergence is not just in the mind; it is real and objective in nature” (2008, 444). Attention to macro-level patterns sounds like a move in the right direction towards autonomy; but, two points. First, I don’t see how Bedau can maintain that some Weakly emergent goings-on are not “merely epistemological” (2008, 451) and rather reflect an “autonomous and irreducible macro-level ontology”, while also maintaining that all Weakly emergent goings-on are ontologically and causally reducible to the micro-level goings-on. Either the metaphysical (ontological) autonomy or the metaphysical (ontological) reducibility has to go.

Second, for purposes of blocking the potential reducibility of higher-level to lower-level features, it isn’t enough merely to point to the fact that the higher-level feature enters into nomological patterns that are in some sense more general than those into which the lower-level features enter. To see this, it is useful to recall a related dialectic in the physicalism debates. There, would-be non-reductive physicalists point to the fact that mental features are associated with functional roles (i.e., “macro-level patterns”) that can be multiply implemented, or realized, by lower-level physical features; but the reductionist’s standard response is that the existence of functional or other comparatively general patterns can be accommodated, on their terms, by identifying a given mental feature with (to cite the usual candidate) a disjunction of its first-order physical realizers. The dialectic sometimes continues, with the non-reductive physicalist rejecting the existence of disjunctive features on grounds that they are gerry-mandered, unprojectible, or problematically infinitary; but such considerations seem, from a reductionist perspective, either unprincipled or uncompelling. Similarly, in the case of the feature *being a glider gun*, while something seems right about attending to the comparative generality of the patterns into which this feature enters, more needs to be said if reductionist accommodation of this sort of macro-level pattern is to be blocked in a principled and compelling fashion. Bedau does not say more along these lines, however, and the upshot is that this strategy, like the previous, fails to establish that any complex systems are even Weakly metaphysically emergent.

5.2.2 Mitchell's appeal to self-organization

I next turn to Mitchell's (2012) account of the emergence of chaotic complex systems exhibiting what I will call 'dynamic self-organization'. In general, Mitchell follows Wimsatt (1994, 1996, 2007), Bedau (1997, Bedau 2008), Thompson and Varela (2001, Kauffman (1993, 1995), Camazine et. al (2001), Thompson (2007), and others in taking emergence to involve "certain types of non-aggregative compositional structures" (179). In the case of chaotic complex systems, the non-aggregativity is dynamic, she maintains, both in arising in a process-like fashion from interactions of the constituents, and in involving feedback loops of the sort characteristic of self-organized systems:

Self-organized systems are ones in which feedback interactions among simple behaviors of individual components of a system produce what appears to be an organized group-level effect. (183)

The flocking of birds is a case in point:

Simple additive relations and simple linear equations [...] will fail to make sense out of much of the complexity that we find in nature even though patterns and structures emerge from the simple interactions of the constituents. The vee pattern that emerges in a flock of geese or the more complex patterns of flocking starlings are not predictable by an aggregation of behaviors of individuals in solo flight, but only from the non-aggregative interaction or self-organizing that derives from the local rules of motion plus feedback among the individuals in group flight (see Couzin 2007; see Rosen 2007 for photos of the starling patterns). Ontologically, there are just physical birds; there is no new substance, no director at a higher level choreographing the artistic patterns the flocks make. Nevertheless, this type of behavior is emergent. (179)

Mitchell is sensitive to the threat of ontological and causal reduction faced by certain accounts of physically acceptable emergence, but maintains that emergent features arising from dynamic self-organization are not subject to such reduction, on grounds, first, that "interaction among the parts generates properties which

none of the individual components possess” and second, that “these higher-order properties in turn can have causal efficacy, i.e., novelty” (179). I address each of these claims, in turn.

First, as previously discussed, a conception of emergence according to which it suffices for the operative notion of ‘non-aggregativity’ that a composed whole have “properties which none of the individual components possess” is too weak to serve as a basis for an interesting conception of emergence. Again, the reductionist will happily allow that some relational aggregates (configurations) or pluralities have properties that are ‘novel’ in the weak sense of not being had, either by way of type or token, by individual components. Hence it is that if there is to be any interesting distinction between reductionists and emergentists of whatever stripe, the reductionist must be granted resources to accommodate relational aggregates or pluralities, and associated features.

Moreover, these additional resources need not be restricted to linear combinations, as Mitchell seems to suggest in observing that “simple additive relations and simple linear equations” do not suffice to accommodate flocking behaviour. For—not least because fundamental physics is itself nonlinear—the reductionist is well within their rights to take any lower-level interactions (linear or nonlinear, as the lawful case may be) between components into account as input into their theorizing—at least when those interactions are manifest when in relatively non-complex combinations. But in that case, it is unclear why the reductionist cannot agree with Mitchell’s characterization of scientific emergence as involving “concrete accounts of how and why rules of interaction among components produce difficult-to-predict emergent behaviors”. The bare fact that, e.g., certain properties of flocks of birds are not had by individual birds, or that these group-level properties are nonlinear functions of properties of individuals, itself goes no distance towards blocking ontological reduction.

Nor does Mitchell’s discussion of the causal novelty associated with group-level behaviours block causal reduction. Mitchell notes a number of ways in which emergent dynamics is associated with interesting causal interactions. For example, self-organization results in stable properties (pertaining, e.g., to flocks

having density concentrated on the interior) which can be “the target of natural selection [and hence] exhibit causal saliency (see [Carere et al. 2009](#))” (180). For another example, behaviours giving rise to group-level properties can also, if heritable, explain certain facts about natural selection. And Page and Mitchell ([1990a](#), [1990b](#)) have run experiments on ‘computer bees’ which suggest that division of labour is a kind of new behaviour of groups, even when the individuals are assigned very ‘minimal’ behaviours. [Mitchell \(2012\)](#) goes on to suggest that these forms of causal interaction are distinctive in incorporating historical and contextual features:

It is still the case that the properties and behavior of the component parts cause the ensuing behavior of the system, but there is a shift of emphasis to the features of the history and context that the system experiences to understand why one outcome occurred rather than another. (182)

But just as the novelty associated with nonlinear non-aggregativity is no barrier to ontological reductionism, nor is the novelty associated with nonlinear complex systems any barrier to causal reductionism. Here again, the reductionist will presumably maintain that, while these forms of causal interaction are massively complex, there is nothing—at least for all Mitchell’s cases establish—to prevent their being given a reductionist treatment. In particular: nothing requires that causal interactions on a reductionist account be spatiotemporally local or somehow free of history or context; on the contrary, the reductionist can allow that spatiotemporally ‘wide’ circumstances crucially enter into what causes what. It is moreover worth noting that the reductionist will point to Mitchell’s reference to ‘a shift in emphasis’ from local to broader features as suggestive of, and in any case as consonant with, a merely pragmatic reading of the higher-level efficacy at issue.

So far, then, Mitchell’s considerations do not establish that self-organizing behaviours are anything more than highly complex lower-level processes.

5.2.3 Batterman's appeal to asymptotic singularities

Batterman has written a great deal concerning the status as emergent or reducible of special science entities (see his 2002, 2005, and elsewhere), though there has remained unclarity as regards whether he takes emergence to be a metaphysical rather than merely epistemic phenomenon, and if so, of what strength. As I'll discuss, there is good reason to think that Batterman does not have even a Weak account of metaphysical emergence in mind, much less the Strong account some attribute to him. Still, Batterman's work on emergence in asymptotic regions of the sort associated with phase transitions is relevant to the present discussion, both because chaotic nonlinear systems are associated with such transitions, and because, whatever Batterman's view of the matter, at least one of the features of systems in asymptotic regions that he mentions can, as I'll later discuss, support taking complex systems to be Weakly metaphysically emergent.

Let's start with the basics of Batterman's account of asymptotic emergence. An asymptote in mathematics is a limiting value of a function that is approached indefinitely closely, but never reached. So, for example, as $x \rightarrow 0$ the function $1/x$ goes to infinity; in this case (though not all, of course) the asymptote is associated with a discontinuity. Interestingly, many “near-neighbor” scientific theories involve asymptotes: special relativity asymptotically approaches Newtonian mechanics in the limit as $v/c \rightarrow 0$, wave optics asymptotically approaches ray optics as $1/\lambda \rightarrow 0$, quantum mechanics asymptotically approaches Newtonian mechanics as Planck's constant approaches 0, statistical mechanics asymptotically approaches thermodynamics in the “thermodynamic limit”, where particle number N and volume $V \rightarrow \infty$. Now, in some of these cases—in particular, the latter three—the asymptotic limits at issue are associated with discontinuities in the regions near the asymptote. In such cases of “singular” asymptotic limits, Batterman suggests, we have reason to take various objects or features associated with the asymptotic region (or associated theory) to be emergent. In particular, to cut to the case which concerns us, Batterman suggests that various features of systems undergoing phase transitions, including those associated with certain critical exponents, are emergent features of such systems. Why so, and in what sense?

Batterman's most explicit stated considerations concern broadly explanatory factors. Again focusing on the case which concerns us, one explanatory concern reflects a kind of theoretical mismatch between the near-neighbor theories at issue, insofar as the discontinuities associated with taking the thermodynamic limit, and which are commonly supposed to be needed to accommodate the associated asymptotic phenomena, find no representational mirror in the analytic functions of statistical-mechanics. Even if there were no problem with deriving specific instances of asymptotic features from the micro-theory, however, Batterman's second explanatory concern would remain; namely, that the characteristic universality of asymptotic phenomena cannot be properly explained by reference just to lower-level "causal-mechanical" explanations. As above, the behavior of systems undergoing phase transitions is characterized by a small set of dimensionless numbers called "critical exponents". As [Batterman \(1998\)](#) says,

What is truly remarkable about these numbers is their universality [...] the critical behavior of systems whose components and interactions are radically different is virtually identical. Hence, such behavior must be largely independent of the details of the microstructures of the various systems. This is known in the literature as the "universality of critical phenomena". Surely one would like to account for this universality. (198)

Lower-level causal-mechanical explanations, even if available, cannot account for universality, for, as [Hooker \(2004\)](#) puts it, these "will be infinitely various in detail and this will block any reconstruction of what is universal about them" (442).

By way of contrast, Batterman argues, various methods for modeling asymptotic phenomena—most notably, the Renormalization Group (RG) method—do provide an explanation of the universal features of systems undergoing phase transitions. The RG method takes a system's governing laws (e.g., its Hamiltonian) and iteratively transforms these into laws having a similar form but (reflecting moves to increasingly 'larger' scales) fewer parameters. In the limit of applications of the method, the resulting Hamiltonian describes the behavior of a single 'block', corresponding to the macroscopic system. The method is suitably applied to systems undergoing phase transitions reflects that near critical points, for

such systems cease to have any characteristic length scale, and are “self-similar” in that the laws governing the systems take the same form at all length scales. As Batterman (1998) puts it:

One introduces a transformation on this space that maps an initial physical Hamiltonian describing a real system to another Hamiltonian in the space. The transformation preserves, to some extent, the form of the original physical Hamiltonian so that when the interaction terms are properly adjusted (renormalized), the new renormalized Hamiltonian describes a system exhibiting the same or similar thermodynamical behavior. Most importantly, however, the transformation effects a reduction in the number of coupled components or degrees of freedom within the correlation length. Thus, the new renormalized Hamiltonian describes a system which presents a more tractable problem. It is to be hoped that by repeated application of this renormalization group transformation the problem becomes more and more tractable until one can solve the problem by relatively simple methods. In effect, the renormalization group transformation eliminates those degrees of freedom (those microscopic details) which are inessential or irrelevant for characterizing the system’s dominant behavior at criticality. (200)

(I quote this and the next passage at length, in order to make explicit, for future purposes, Batterman’s appeal to certain features potentially relevant to emergence.) In particular, application of the RG method to the cases at hand enables calculation of the critical exponents associated with phase transitions, and hence provides an explanation of the universal behavior of systems near critical points:

[I]f the initial Hamiltonian describes a system at criticality, then each renormalized Hamiltonian must also be at criticality. The sequence of Hamiltonians thus generated defines a trajectory in the abstract space that, in the limit as the number of transformations goes to infinity, ends at a fixed point. The behavior of trajectories in the neighborhood of the fixed point can be determined by an analysis of the stability properties of the fixed point. This analysis also allows for the calculation of the critical exponents characterizing the critical behavior of the system. It turns out that different physical Hamiltonians

can flow to the same fixed point. Thus, their critical behaviors are characterized by the same critical exponents. This is the essence of the explanation for the universality of critical behavior: Hamiltonians describing different physical systems fall into the basin of attraction of the same renormalization group fixed point. This means that if one were to alter, even quite considerably, some of the basic features of a system (say from those of a fluid F to a fluid F' composed of a different kind of molecule and a different interaction potential), the resulting system (F') will exhibit the same critical behavior. This stability under perturbation demonstrates that certain facts about the micro-constituents of the systems are individually largely irrelevant for the systems' behaviors at criticality. (201)

Batterman's account of asymptotic emergence cites three features potentially relevant to the emergence of many complex systems:

- 
1. elimination of micro-level degrees of freedom (DOF)
 2. universality of certain features or behavior
 3. stability of certain behavior under perturbation.

As Hooker (2004) notes, these features are characteristic of chaotic nonlinear systems:

In every case of so-called ‘critical phenomena’, e.g. near the ‘critical point’ beyond which there is no vapour phase between liquid and gas, the asymptotic domain shows a universally self-similar spectrum of fluctuations. [...] This is indicative of chaos and occurs when behaviours are super-complexly, but still systematically, interrelated [...]. (440)

Indeed, the core similarities between critical phenomena in statistical mechanics and chaotic nonlinear phenomena, including period doubling and intermittency routes to chaos of the sort displayed by the logistic map, have led to an active area of investigation in which “the logistic map is [understood as] a prototypical

system [...] for the assessment of the validity and understanding of the reasons for applicability of the nonextensive generalization of [...] Boltzmann-Gibbs statistical mechanics” (Mayoral and Robledo 2006, 339). If the features entering into Batterman’s account can be seen as supporting metaphysical emergence, this result would apply to a wide range of complex systems.

Now, as it happens (and as I lay out in [Wilson 2013b](#)) there is good reason to believe that Batterman does not intend his discussion to be interpreted as offering either an account of or a case of metaphysical emergence. As above, his discussion is squarely focused on the question of what is required if the critical behaviors of the systems in question are to be explained, with the general idea being that, in the case of such systems, neither theoretical derivations nor causal-mechanical considerations can do the trick. Hence [Morrison \(2012\)](#) reads Batterman as offering an “explanatory” account of emergence:

I characterize Batterman’s account [of emergence] as explanatory insofar as the main argument centers on how asymptotic methods (via the RG) allow us to explain features of universal phenomena that are not explainable using either intertheoretic reduction or traditional causal mechanical accounts [...]. (143)

One might wonder if Morrison’s reading is undermined by Batterman’s seeming to suggest, especially in some passages in his ([2002](#)), that he sees metaphysical emergence as following from explanatory emergence, as when he says, “It seems reasonable to consider these asymptotically emergent structures [as constituting] the ontology of an explanatory ‘theory’” (96). For another example, in discussing the asymptotic domain between wave and ray optics, Batterman first argues that rainbows cannot be explained without referring to “caustics”—ray tangent curves associated with the higher-level, but not lower-level theory, then suggests that we

are ontologically committed to such optical objects:

[I]f I'm right and there is a genuine, distinct, third theory (catastrophe optics) of the asymptotic borderland between wave and ray theories—a theory that of necessity makes reference to both ray theoretic and wave theoretic structures in characterising its “ontology”—then, since it is this ontology that we take to be emergent, those phenomena are not predictable from the wave theory. They are “contained” in the wave theory but aren't predictable from it. (119)

These suggestive passages are misleading, however, since Batterman has elsewhere explicitly denied that the emergence at issue here carries ontological weight:

I do not believe that there is any new ontology in the asymptotic catastrophe optics. The wave theory has replaced the ray theory and there simply are no caustics (as characterized by the ray theory). Asymptotic analysis of the wave equation yields terms (think syntax here) which require for their interpretation (semantics) that we make reference to ray theoretic structures. In effect, it is the understanding/interpretation of these terms in the asymptotic expansions that requires ‘appeal’ or reference to structures that ‘exist’ only in the ray theory. (p.c. with Hooker, discussed in [Hooker 2004](#), 448)

Hooker concludes that “the status claimed for caustics based on essential reference is not after all that of ontological commitment but one of ineliminable semantic role, without ontological commitment” (448). Similar remarks go for the “emergent” phenomena in phase transitions or chaotic nonlinear systems, more generally. Strictly speaking, Batterman's account of asymptotic emergence is intended to be an account of (merely) theoretical/representational or epistemological, not metaphysical, emergence.

That said, nothing prevents us from considering whether any of the aforementioned features—elimination in micro-level DOF, universality, and/or stability under perturbation—might serve as sufficient indicators of metaphysical emergence.

With an eye to sticking somewhat closely to Batterman's work, we might look especially to universality and stability under perturbation, since it is these features which he has most consistently highlighted as motivating the 'emergence' not just of systems near critical points, but also of special science entities to which the RG approach and its associated strategy for eliminating DOF do not apply.

Universality and stability under perturbation are really two sides of the same coin; as Batterman says, "most broadly construed, universality concerns similarities in the behavior of diverse systems" (2000, 120). The suggestion that physically acceptable emergence might be a matter of universality or of stability under micro-level perturbations is common enough; indeed, we have already seen a version of this suggestion in Bedau's suggestion that the fact that certain features of nonlinear automata (e.g., glider guns) enter into "macro-level patterns" might support such features' being metaphysically autonomous. It is unsurprising, then, that the same concerns with Bedau's suggestion, which again echo the debate over the import of multiple realizability in the metaphysics of mind, also attach to attempts to locate metaphysical emergence in universal or stable features of composite entities; namely, that reductionists have various strategies for accommodating such features, and that the standard anti-reductivist responses (rejecting disjunctive properties, denying that these track genuine natural kinds, and the like) are not compelling. What is needed, if these features are to be seen as tracking the ontological autonomy (distinctness) of composite entities, is a better response to the usual reductivist strategies; but Batterman does not provide such a response—arguably because his concern is ultimately with whether an appropriately explanatory account of universal or multiply realized features is available, and not with whether such features are (or are not) ontologically reducible.

5.2.4 Eliminations in degrees of freedom

Let's sum up the results so far. I've considered a representative survey of accounts of physically acceptable emergence intended to apply to complex systems of one or other variety, but in each case the occurrence of the feature highlighted in the account (epistemic indiscernibility, algorithmic or explanatory incompressibility, presence in a macro-pattern, self-organization, universality or stability under perturbation) has turned out to be, at least for all that proponents of these accounts have established, compatible with the ontological and causal reduction of the complex systems at issue. So far, then, it has not been established that any complex systems are even Weakly metaphysically emergent.

I turn now to arguing that an account of physically acceptable emergence understood as involving an elimination in degrees of freedom (DOF), as endorsed in [Wilson 2010b](#), can do better by way of establishing the Weak emergence of many complex systems. The most straightforward cases of such emergence involve complex systems undergoing phase transitions, where the applicability of the renormalization group entails the elimination of DOF that is criterial for Weak emergence on the DOF-based account. There are also plausible cases to be made that certain other complex systems, including glider guns in the Game of Life and flocking birds, can be seen as satisfying the criterion, thus providing a principled basis for resisting reductionist treatments of the macro-patterns and self-organization manifested by such complex systems.

A DOF-based account of Weak emergence

I start by providing a more detailed presentation of a DOF-based account of Weak emergence. I follow my previous presentation in focusing on the DOF-based

emergence of entities rather than features; but as previously, nothing deep hangs on this variation on the Weak emergent theme.

Call states upon which the law-governed (i.e., nomologically possible) properties and behaviour of an entity E functionally depend the ‘characteristic states’ of E . A DOF is then, roughly, a parameter in a minimal set needed to describe an entity as being in a characteristic state. Given an entity and characteristic state, the associated DOF are relativized to choice of coordinates, reflecting that different sets of parameters may be used to describe an entity as being in the state. More precisely, the operative notion of DOF is as follows:

Degrees of Freedom (DOF): For an entity E , characteristic state S , and set of coordinates C , the associated DOF are parameters in a minimal set, expressed in coordinates C , needed to characterize E as being in S .

I’ll sometimes speak for short of “characterizing an entity”, with state and coordinates assumed.

Some common characteristic states, and DOF needed to characterize certain systems as being in those states, are as follows:

- *The configuration state*: tracks position. Specifying this state for a free point particle requires 3 parameters (e.g., x , y , and z ; or r , ρ , and θ); hence a free point particle has 3 configuration DOF, and a system of N free point particles has $3N$ configuration DOF.
- *The kinematic state*: tracks velocities (or momenta). Specifying this state for a free point particle requires 6 parameters: one for each configuration coordinate, and one for the velocity along that coordinate; hence a free point

particle has 6 kinematic DOF, and a system of N free point particles has $6N$ kinematic DOF.

- *The dynamic state*: tracks energies determining the motion. Specifying this state typically requires at least one dynamic DOF per configuration coordinate, tracking the kinetic energy associated with each position coordinate; other dynamic DOF may track internal/external contributions to the potential energy.

Why attend to DOF in hopes of illuminating Weak emergence? To start, as above, different systems, treated by different sciences, may be functionally dependent on the same characteristic state (e.g., the configuration state). Moreover, as above, the DOF needed to characterize intuitively different systems as being in these states typically vary. Following these observations, I take the main cash value of attention to DOF to lie in the fact that DOF track the details of a system's functional dependence on its characteristic states, in a more fine-grained way than the mere fact that the system is in the state does. Driving my account is the idea that the fine-grained details concerning functional dependence that are encoded in the DOF needed to characterize a given entity or system serve as a plausible ontological basis for the individuation of broadly scientific entities or systems.

I start by observing an important tripartite distinction (again, see Wilson 2010*b*) relevant to such functional dependence, reflecting that the DOF needed to characterize an entity may be reduced, restricted, or eliminated in certain circumstances (typically associated with the imposition of certain constraints or more generally, the presence of certain energetic interactions), compared to those needed to characterize a (possibly distinct) entity, when such circumstances are not in place. To prefigure: eliminations in DOF, in particular, enter into the upcoming account of

Weak emergence. Let's get acquainted with these different relations and note an example of each.

First, constraints may reduce the DOF needed to characterize an entity as being in a given state. So, for example, a point particle constrained to move in a plane has 2 configuration DOF, rather than the 3 configuration DOF needed to characterize a free point particle. In cases where a given DOF is given a fixed value, the laws governing an entity so constrained are still functionally dependent on the (now constant) value of the DOF; hence such constraints do not eliminate the DOF, but rather reduce it to a constant value. By way of example: rigid bodies treated in classical mechanics have such a reduced set of DOF relative to the unconstrained system of their composing entities.

Second, constraints may also restrict DOF needed to characterize an entity. A point particle may be constrained, not to the plane, but to some region including and above the plane. Characterizing such a particle still requires 3 configuration DOF, but the values of one of these DOF will be restricted to only certain of the values needed to characterize the unconstrained particle. Cases of restriction in DOF are more like cases of reduction than elimination of DOF, in that, again, the entity's governing laws remain functionally dependent on specific values of the DOF. By way of example: molecules, whose bonds are like springs, have a restricted set of DOF relative to the unconstrained system of their composing entities.

Third, sometimes the imposition of constraints eliminates DOF. So, for example, N free point particles, having $3N$ configuration DOF, might come to compose an entity whose properties and behavior can be characterized using fewer configuration DOF, not because certain of the DOF needed to characterize the uncon-

strained system are given a fixed value, but because the properties and behavior of the composed entity are functionally independent of these DOF. By way of example: a spherical conductor of the sort treated in electrostatics has DOF that are eliminated relative to the system of its composing entities, since while the *E*-field due to free particles depends on all charged particles, the *E*-field due to the spherical conductor depends only on the charges of particles on its surface. Certain quantum DOF are also eliminated in the classical (macroscopic) limit. So, for example, spin is a DOF of quantum entities; entities of the sort treated by classical mechanics are ultimately composed of quantum entities; but the characteristic states of composed classical-mechanical entities do not functionally depend on the spins of their quantum components.

Two features of the illustrative special science case studies are important for what follows. First is that the holding of the constraints relevant to reducing, restricting, or eliminating DOF occurs as a matter entirely of physical or physically acceptable processes. Such processes suffice to explain why sufficiently proximate atoms form certain atomic bonds, why atoms or molecules engage in the energetic interactions associated with SM ensembles, and so on. More generally, for each of the aforementioned special science entities *E*, the constraints on the e_i associated with their composing *E* are explicable using resources of the theory treating the e_i (or resources of some more fundamental theory, treating the constituents of the e_i). Call such constraints “ e_i -level constraints”. A second important feature of these special science entities *E* is that all of their properties and behavior are completely determined by the properties and behavior of their composing e_i , when these stand in the relations relevant to their composing *E*. People disagree about the metaphysical ground for this determination, but all parties agree that the

determination is in place.

These preliminaries in hand, the DOF-based account of Weak emergence is as follows:

DOF-based Weak emergence: An entity E is Weakly emergent from some entities e_i if

1. E is composed by the e_i , as a result of imposing some constraint(s) on the e_i .
2. For some characteristic state S of E : at least one of the DOF required to characterize the system of unconstrained e_i as being in S is eliminated from the DOF required to characterize E as being in S .
3. For every characteristic state S of E : every reduction, restriction, or elimination in the DOF needed to characterize E as being in S is associated with e_i -level constraints;
4. The law-governed features of E are completely determined by the law-governed features of the e_i , when the e_i stand in the relations relevant to their composing E .

Systems that are emergent by lights of the above account are physically acceptable, given that the unconstrained system of composing entities is physically acceptable. In my (2010b), I argue for this in some detail, but here I'll just observe that this result is to be expected, given that both the constraints relevant to composing E , as well as all of E 's law-governed features, are physically acceptable.

The ontological irreducibility of DOF-eliminated entities

Now, as above, the key concern with other accounts of the physically acceptable emergence of complex systems, as appealing to unpredictability (or variations thereof), macro-patterns, self-organization, universalizability, or stability under perturbation, is that there are available reductionist treatments of the having of

each of these features, contra the purported metaphysical emergence of the complex systems at issue. By way of contrast, a DOF-based account provides a principled means of blocking the ontological and causal reduction of at least some complex systems.

There are two strategies for showing this, depending on whether a ‘lightweight combination’ or rather a ‘law consequential’ account of levels is assumed (see Ch. 1, ‘Preliminaries’). As will become clear, both ultimately turn on the fact that goings-on with eliminated DOF cannot be identified with any lower-level goings-on, since lower-level laws must take as input only goings-on with the full complement of DOF—else the lower-level laws will not know, so to speak, what to do with such input.

If a lightweight combination account of levels is assumed, the strategy proceeds by way of an argument by cases, reflecting the various ways in which a composed entity E might be reduced to some lower-level goings-on; this was the approach I took in my 2010. As previously discussed, the candidate lower-level reducing entities include all the lower-level entities, properties and relations, plus any ontologically “lightweight” constructions—lower-level relational, Boolean (conjunctive or disjunctive), or mereological—out of these. Hence if E is to be ontologically reducible to the e_i , then E must be identical to either:

- (i) a system consisting of the jointly existing e_i
- (ii) a relational entity consisting of the e_i standing in e_i -level relations, or
- (iii) a relational entity consisting in a boolean or mereological combination of the entities at issue in (i) and (ii).

Since E satisfies the conditions in *DOF-based Weak emergence*, there is some

state S such that characterizing E as being in this state requires fewer DOF than are required to characterize the unconstrained system of E 's composing e_i as being in S . Hence a necessary condition on E 's being identical with an entity of the type of (i)–(iii) is that the DOF required to characterize the candidate reducing entity as being in S are similarly eliminated, relative to the unconstrained system. I'll now argue that this condition isn't met for any of the entities of type (i)–(iii).

First, consider the e_i understood as (merely) jointly existing, as per (i). Such a system of e_i is not subject to any constraints; hence for any state, characterizing this system will require the same DOF as are required to characterize the system of unconstrained e_i . So E is not identical to the system consisting of (merely) jointly existing e_i .

Second, consider the relational entity e_r consisting of the e_i standing in certain e_i -level relations, as per (ii). Though e_r realizes a constrained entity (namely, E), e_r is not itself appropriately seen as constrained—even throwing the holding of the constraints into the mix of e_i -level relations at issue in e_r . Why not? Because the laws governing entities (such as e_r) consisting of the e_i standing in e_i -level relations are, unlike the laws governing E , compatible with the constraints not being in place. Hence characterizing e_r as entering into these laws, hence as capable of evolving into states (of, perhaps other e_i level entities) where the constraints are not in place, requires all the DOF associated with the unconstrained system of e_i . So E is not identical to e_r .

Third, consider a Boolean (disjunctive or conjunctive) or mereological combination of entities of the previous varieties (or the closure of any such entity). To start, consider a relational entity consisting in a disjunctive combination of entities of the sort at issue in (i) or (ii). The occurrence of a disjunctive entity consists in

one of its disjunct entity's occurring. Hence for any state, characterizing a disjunctive entity as being in that state will require all the DOF required to characterize any of its disjunct entities as being in that state. Such disjunct entities, being of type (i) or (ii), will require all the DOF required to characterize the system of unconstrained e_i , for any state. So characterizing a disjunctive relational entity will require all the DOF required to characterize the system of unconstrained e_i , for any state. The same goes for conjunctive entities. So E isn't identical to a disjunctive or conjunctive relational entity.

Finally, consider a relational entity consisting in a mereological combination of entities of the sort at issue in (i) or (ii). Mereological wholes are identified with the mere joint holding of their parts; hence characterizing the whole will require all the DOF required to characterize each of the parts. Such parts, being of type (i) or (ii); will require all the DOF required to characterize the system of unconstrained e_i , for any state. So characterizing a mereological relational entity will require all the DOF required to characterize the system of unconstrained e_i , for any state. So E isn't identical to a mereological relational entity.

That exhausts the available candidates to which purportedly higher-level entities might be ontologically reduced. In each case, the entity at issue has fewer DOF than the candidate reducing entity, so by Leibniz's law the former is not identical to the latter. Attention to the metaphysical implications of eliminations of DOF thus indicates that theoretical deducibility is compatible with ontological irreducibility. Hence physical acceptability is compatible with ontological autonomy, as DOF-based Weak emergence requires.

Next, consider a law-consequence account of levels. Recall that on this approach, laws are understood in metaphysical terms as expressing the “rules” gov-

erning certain entities and features (though as per usual, discussion of laws typically focuses on associated scientific theories as stand-ins). The suggestion is then that the laws governing entities characteristic of a given level can serve as a basis for appropriately expanding the domain of entities and features at that level, in a way which makes room for complex pluralities or relational aggregates of entities, and associated features, at a level. The laws of fundamental physics, for example, are capable of taking as input or initial conditions various complex configurations of characteristic physical entities and features; hence the laws/theories themselves have resources to expand beyond the explicit commitments of the laws/theory treating small numbers of physical individuals. More generally, the law-consequence approach takes levels to be individuated not just by the individuals and associated features that are characteristic of L (e.g., atoms in atomic physics, cells in cellular biology) but also as including any entities and features whose (potential) existence is deemed a metaphysical consequence—not to be confused with either mere necessitation or representational entailment—of the L -level laws.

As noted, a law-consequence approach to the individuation of levels has certain *prima facie* advantages over a lightweight combination approach, one of which is especially relevant to the question of whether any complex systems are metaphysically emergent—namely, that unlike an ontologically lightweight approach to levels, a law-consequence approach need not antecedently specify whether nonlinear entities or features are or are not to be placed at a given level L . Rather, whether or not this is so will follow from the laws governing the entities characteristic of L . More generally, the law-consequence approach can allow that entities and features which are causal consequences just of the L -level laws may

also be placed at L .

In presenting a law-consequence account of levels, I noted that one might be reasonably concerned that such an account will not make in-principle room for the possibility of physically acceptable emergence. For non-reductive physicalists grant that the higher-level entities and features that they take to be genuine, as well as the higher-level laws governing these entities and features, are *in some sense* metaphysical consequences of the fundamental physical laws, even if these higher-level goings-on are (on their view) different from any lower-level goings-on (and moreover, some non-reductive physicalists think, are at least sometimes epistemically beyond our ken). And indeed, this concern is just the one with which we are concerned—namely, the concern that if the physicalist grants that purportedly higher-level goings-on are theoretically deducible to lower-level goings-on (perhaps by an ideal calculator, and given appropriate idealizations), then they must also grant that the former are ontologically reducible to the latter.

In Ch. 1, ‘Preliminaries’, I sketched a way of making sense of the (the possibility of) Weak emergence on a law-consequence approach, based in the notion of a DOF. We are now in position to revisit this strategy, and more specifically to see how it provides a basis for taking certain higher-level goings-on—in particular, certain complex systems—to be Weakly emergent.

Let us suppose that the e_i are treated by the lower-level physical laws L_p , and that an entity E satisfies the conditions in DOF-based Weak emergence *vis-à-vis* the e_i , such that (among other things) the DOF needed to characterize E as being in some characteristic state is strictly fewer than those needed to characterize the system of lower-level physical e_i as being in that state. Suppose, to fix ideas and as is anyway typically the case, among the DOF that are eliminated in characterizing

E is quantum spin. In this case, E will not be appropriately identified with any lower-level physical goings-on, whatever they might be. Why not? Because the goings-on appropriately placed at L_p will be those whose specification includes all the degrees of freedom needed for the physical laws to operate—including, e.g., spin. Since E 's characterization, by assumption, fails to include any information about spin, the lower-level physical laws will not be able to take an entity such as E as input. E , and more generally any goings-on whose DOF are eliminated *vis-á-vis* their lower-level physical realizers, cannot be identified with any lower-level goings-on, since their specifications fails to include all the DOF needed for the physical laws to operate.

This much shows that an entity E 's satisfaction of the conditions in DOF-based Weak emergence suffices to ensure E 's ontological irreducibility. We can also see how satisfaction of these conditions provides a basis for E 's being casually autonomous, for reasons associated, more generally, with the schema for Weak emergence. What powers an entity has are plausibly a matter of what it can do; and the sciences are plausibly in the business of expressing what the entities they treat can do. It follows that, plausibly, what powers an entity has are expressed by the laws in the science treating it. The powers of E are thus those expressed by the laws in the theory treating (constrained) entity E , while the powers of e_i are those expressed by the laws in the more fundamental theory treating the (relatively unconstrained) lower-level constituents of E —that is, the constituents of E as existing both inside and outside the constraints associated with E . Consequently, the laws of the theory treating E express what happens when certain lower-level entities stand in relations associated with certain lower-level constraints, and the laws treating e_i express what happens when certain lower-level entities stand both

in these relations and in other relations not associated with the constraints. Hence the system of composing entities e_i has more powers than E , and the proper subset relation between powers at issue in the schema for Weak emergence is thus in place. And as such, for reasons discussed previously, E will be distinctively efficacious *vis-á-vis* the system of its composing e_i .

DOF-based Weak emergence and complex systems

Let's return now to cases of complex systems, and consider whether any of these are plausibly seen as Weakly emergent. As I'll now argue, there are cases to be made that many complex systems, including those having the features of discussed by Bedau, Mitchell, and Batterman, are Weakly emergent by lights of a DOF-based account.

I start with Batterman's account of complex systems undergoing phase transitions, having the features of universalizability and stability under perturbation. Given Batterman's observation (discussed above) that the Renormalization Group (RG) method applies to such complex systems, and given that eliminations in DOF (along with certain other suppositions which are here in place, concerning e_i constraints and e_i determination) are sufficient for Weak emergence, here we can be comparatively brief.

As previously noted, the RG method applies to systems undergoing phase transitions, which are relevantly similar to and indeed can be understood as chaotic nonlinear systems; and that the RG applies to a given system provides as good indication as we are likely to get that the system has DOF that are not just reduced or restricted, but eliminated as compared to the unconstrained system of its composing entities (that is, to the system of composing entities when not energetically

interacting in the way associated with phase transitions, or more generally, with chaotic behavior). We can thus argue as follows:

1. Systems that can be modeled by the RG have eliminated DOF ([Batterman 1998](#) and elsewhere)
 2. Chaotic nonlinear systems are modeled by the RG
 3. Therefore, chaotic nonlinear systems have eliminated DOF
 4. Systems with eliminated DOF are Weakly emergent ([Wilson 2010b](#))
- ∴ Chaotic nonlinear systems are Weakly emergent.

As confirmation of the fact that many chaotic systems have eliminated DOF, it is worth noting that one of the puzzles that Batterman raises for thermodynamic systems carries over to chaotic complex systems, and is answered in just the same way. The puzzle he raises concerns how thermodynamic systems can be a viable object of study. Such systems—e.g., an isolated gas E —are composed of massively large numbers of particles or molecules e_i . Since the composite entity E in this case is (boundary restrictions aside) effectively unstructured, shouldn't it have the same DOF as the system of unconstrained e_i ? Supposing so, however, the success of statical mechanics (SM) is mysterious, since obviously we can't track on the order of 10^26 DOF. As [Batterman \(1998\)](#) puts it:

One wants to know why the method of equilibrium SM—the Gibbs' phase averaging method—is so broadly applicable; why, that is, do systems governed by completely different forces and composed of completely different types of molecules succumb to the same method for the calculation of their equilibrium properties? (185)

The answer reflects that while the e_i are not bonded, they are interacting via exchanges of energy, and such interactions may not only restrict or reduce, but

eliminate DOF, which fact is indicated by the fact that the renormalization group method is appropriately applied to such systems. This, then, is the answer to the puzzle: such systems are tractable since the modes of interaction of their composing entities results in their having DOF that are massively eliminated compared to the unconstrained system of composing entities. Again:

[T]he renormalization group transformation eliminates those degrees of freedom (those microscopic details) which are inessential or irrelevant for characterizing the system's dominant behavior at criticality.
(200)

A similar puzzle applies to chaotic complex systems. Recall that chaotic complex systems are characterized by their extreme sensitivity to initial conditions. If nonlinear systems are so sensitive and their resulting trajectories so “chaotic”, how is it that they can be, as they are, a viable object of scientific study? The answer, I take it, is effectively the same: the composing entities, though not bonded, are energetically interacting, in ways that, as application of the RG method reveals, massively eliminates DOF needed to characterize the composite system. Here we have a solution to the puzzle, and more to the present point, a decisive, empirically supported case for taking the important class of chaotic nonlinear systems to be Weakly metaphysically emergent.

Though applicability of the RG method to a given complex system provides a quick route to taking the system to be Weakly emergent, such applicability is not necessary in order to establish that a given complex system satisfies the conditions of the DOF-based account. Indeed, while (as above) bare appeals to macro-patterns (*à la* Bedau) or to self-organization (*à la* Mitchell) do not alone suffice to block ontological or causal reducibility, there are cases to be made that the target systems also satisfy the conditions of this account, and so are appropriately taken

to be Weakly emergent, after all.

First, recall Bedau’s suggestion that certain phenomena, such as gliders and glider guns, in Conway’s Game of Life manifest a sort of autonomy characteristic of ‘innocent’—that is, physically acceptable—emergence, as involving behaviours that are “autonomous with respect to the underlying processes [involving] simple, general macro-level patterns and laws” (1997, 395). In this artificial case, we do not have laws of nature on hand to look to as a guide to the DOF of the entities at issue. Nonetheless, the rules of the Game of Life will do for these purposes.

In particular, we can consider what DOF are required to specify the location of the ‘live’ cells composing a ‘macro-entity’ in the Game of Life—e.g., a glider, or a glider gun, as compared to the DOF required to specify the location of the macro-entity itself. To fix ideas on a simple case, consider a glider (the composed entity E) composed of five ‘live’ cells, which starts at the origin in the following configuration:

Insert figure of initial configuration of glider here

The glider ‘moves’ in a 4 step sequence, at the end of which it returns to its original configuration, one grid diagonally to the NE, so to speak.

Now, let us suppose that the initial seeding of the grid contains only the initial state of a glider, so that we can ignore any potential collisions of the glider with other ‘entities’. Given that the Game of Life takes place on a (potentially infinite) two-dimensional grid, the location of each of the cells composing the glider gun E will require two DOF, corresponding to values (relative to some arbitrarily chosen origin point) along the x and y axes; hence determining the position of each of the five ‘live’ cells composing a glider, in ways that can then be input into the

‘lower-level’ rules of the Game of Life, will require ten DOF. By way of contrast, the position of a glider at n time-steps from t_0 requires only two DOF, specifying the x, y coordinates of the relevant diagonal cell, which at a time-step $n/4$ will be associated by the ‘glider laws’ with the relevant stage of the glider. As such, a glider E in the Game of Life has DOF that are eliminated relative to the system of its composing e_i .

Moreover, a glider E clearly satisfies the other conditions at issue in DOF-based Weak emergence. It is composed by the e_i , as a result of imposing some constraint(s) on the e_i ; here the ‘constraints’ simply reflect how the rules of the Game of Life introduce a stable configuration. There is a characteristic state S —namely, position—the specification of which for E requires strictly fewer DOF than those required to characterize the system of e_i . Every reduction, restriction, or elimination in the DOF needed to characterize E as being in a characteristic state is associated with e_i level constraints, since any such constraints are a matter only of the operation of the rules of the Game of Life. And more generally, the law-governed features of E are completely governed by the law-governed features of the e_i .

Correspondingly, the considerations previously brought to bear which show that there is no prospect of ontologically reducing an entity E satisfying the conditions of DOF-based Weak emergence to any lower-level goings-on are here appropriately brought to bear in support of taking a glider in the Game of Life to be Weakly emergent.

Next, recall Mitchell’s suggestion that certain complex phenomena, such as flocking birds, manifest dynamic self-organization. Here the composing entities are in the first instance the individual birds making up the flock. Specifying the po-

sition of each bird requires at least 3 DOF (actually, it would require many more, but for present purposes this won't matter); hence specifying the position of 100 birds would require 300 position DOF. When the birds compose a flock, however, they mutually constrain each others' positions, moving (to idealize somewhat) as an ensemble. Such constraints, as above, can introduce restrictions, reductions, or eliminations in the DOF needed to specify the position of the birds. Now, since a flock isn't a rigid body, and since knowing the position of the flock at a time requires knowing its shape at that time, the DOF needed to specify the position are all those needed to specify this shape. For this it suffices to know the position of every bird on the external boundary of the flock; the positions of birds in the interior of the flock are irrelevant.⁷ Since some birds in a flock will be in the interior, it follows that the DOF needed to specify, at a time, the position of a flock, will be strictly fewer than—will be eliminated as compared to—the DOF needed to specify, at a time, the positions of all the individual birds constituting the flock.

That said, the case of flocking birds, unlike the case of the other complex systems we have considered, introduces a new consideration as entering into the system at issue—namely, consciousness, at least of a perceptual variety. Presumably it is some part of birds' flocking the way they do that they perceive and are otherwise aware of other birds in the flock and what they are doing, and adjust their own behaviour accordingly. Of course individual birds are also perceptually conscious, so this fact doesn't in itself suggest that flocks fail to satisfy the conditions in DOF-based Weak emergence. But if the entities composing the flock are further decomposed into lower-level physical entities, then whether flocks are

⁷How exactly to determine this boundary is a good question. Perhaps some sort of density measure would come into play. In any case, the point in the main text doesn't hinge on this epistemological issue.

Weakly emergent from lower-level physical entities will hinge on the status as either Strongly or Weakly emergent of such conscious mental features—a topic to which we will return in the final chapter.⁸

5.3 Concluding remarks

Complex non-living systems have frequently been offered up as cases of actual metaphysical emergence, but as I've argued in this chapter, previous accounts of such systems as either Strongly or Weakly emergent do not in fact establish this result.

This is commonly acknowledged as regards British emergentist claims that unpredictability and nonlinearity are sufficient marks of fundamental novelty of the sort associated with Strong emergence—though as I'll discuss in Ch. 8, the empirical jury is still out over whether complex systems associated with mentality manifest apparent violations in conservation laws, which criterion is, I've here suggested, a recognizable descendant of nonlinearity as a marker of fundamental novelty.

By way of contrast, it has frequently been assumed that certain characteristic features of complex systems, including, in addition to unpredictability and nonlinearity, the manifesting of macro-patterns, self-organization, universality, and stability under perturbation, themselves serve as a sufficient basis for taking the complex systems at issue to be Weakly emergent. As I've argued here, this is also

⁸This point highlights another way (beyond the relativization to fundamental interactions associated with my preferred account of Strong emergence) in which attributions of emergence may be relative. For all that has been established (or ruled out) thus far, it might be that flocks are Weakly emergent from birds, but birds (*qua* conscious entities) are Strongly emergent from lower-level physical phenomena.

incorrect, at least to the extent that there are available reductionist strategies for accommodating these features that need to be blocked if the claim of emergence is to be defended.

Luckily, there is another feature of complex systems—namely, having DOF that are eliminated as compared to the DOF needed to characterize one’s composing entities—which provides a principled basis for decisively blocking the threat of ontological and causal reduction, in a way compatible with physicalism (given the physical acceptability of the entities composing the complex system). This feature is most clearly present in cases of complex systems undergoing phase transitions, to which the RG method (which effectively works by means of eliminating DOF) applies. But I’ve also argued that there are cases to be made that certain macro-patterns in cellular automata (e.g., gliders), and certain biological complex systems (e.g., flocks of birds) also have eliminated DOF, and more generally satisfy the conditions in DOF-based Weak emergence.

The first upshot of these results is that while the usually cited features of complex systems may be indications of emergence, closing the deal ultimately requires connecting these features with the sort of eliminations in DOF that, again, provide a principled basis for blocking reductionist treatments of these features. The second upshot is that, modulo the status of any conscious experience there may be, we now have confirmation of the common but previously unsubstantiated belief that many complex systems are emergent in a way compatible with physicalism—that is, are Weakly emergent.

Chapter 6

Ordinary objects

In this chapter, I turn to the question of whether ordinary objects are either Strongly or Weakly metaphysically emergent. By ‘ordinary’ objects I have in mind objects which are uncontroversially inanimate (as Thomasson puts it) or non-living (as Merricks puts it), and of the sort with which creatures like us are or may be perceptually acquainted.¹ Such objects might be either natural (rocks, mountains) or artifactual (tables, baseballs, statues).² As per usual, we will approach this question indirectly, by attention to whether such objects have characteristic features which are appropriately seen as either Weakly or Strongly emergent. And while there are several competing metaphysical accounts of the nature of objects—e.g., as bundles of properties, either tropes or universals (as per [Campbell 1990](#) or [Paul 2002](#)), as combinations of substrate and property (as per [Locke 1690](#) and [Simons](#)

¹Hence the purview of the present chapter is more restricted than that at issue in [Korman 2011](#): “Very roughly, ordinary objects are objects belonging to kinds that we are naturally inclined to regard as having instances on the basis of our perceptual experiences: dog, tree, table, and so forth”.

²Some (e.g., [Grandy, 2007](#)) suggest that there is no deep difference between artifactual and natural objects; here I’ll assume that there is a distinction, though as we’ll see, much of my discussion of the one type of object applies, *mutatis mutandis*, to the other.

1994), as hylomorphic compounds (as per Fine 2003 and Koslicki 2008), and so on—our usual focus on the emergence of features of an emergent entity will allow us to investigate into the question of whether any ordinary objects are emergent in a way that is broadly neutral on which metaphysical account of objects is correct, so long as this account does not rule out of court the possibility that ordinary objects are metaphysically emergent.

I will argue that there are three different cases for thinking that many ordinary objects are either Weakly emergent or (as I will sometimes put it) are ‘at least’ Weakly emergent (in having at least one feature satisfying the conditions in the schema for Weak emergence). An object’s being ‘at least’ Weakly emergent is sufficient for its not being reducible to lower-level goings-on, but leaves open the possibility of some other feature’s satisfying the conditions in the schema for Strong emergence, in which case the object would be deemed Strongly, not Weakly, emergent.

First, I argue that ordinary objects of the sort appropriately treated by classical (or ‘Newtonian’) mechanics are Weakly emergent by lights of a DOF-based account; second, I argue that a common conception of artifacts as associated with distinctive functional roles supports thinking of these as being at least Weakly emergent by lights of a functional realization account; third, I argue that ordinary objects typically have metaphysically indeterminate boundaries, which when coupled with an attractive determinable-based account of such indeterminacy, indicates that any such ordinary objects are at least Weakly emergent, by lights of a determinable-based account. As for whether any ordinary objects are Strongly emergent: here I argue that the best case for this stems from artifactual ordinary objects whose functional or other characterization reflects or encodes certain so-

cial practices involving normative or aesthetic goings-on; the ultimate status of such objects as Strongly or rather just Weakly emergent hinges, like the status of certain complex systems involving mentality, on the status as Weakly or Strongly emergent of the associated mental features of persons, of the sort to be discussed in the next chapters.

I close by observing that the results of this chapter have consequences not just for the status of ordinary objects as existing and at least Weakly emergent, but also for the proper assessment of Amie Thomasson's metaontological view, as discussed in her (2010) and elsewhere, that investigations into the ontological status of ordinary objects should proceed differently from investigations into the ontological status of special science entities.

6.1 Are ordinary objects Weakly emergent?

6.1.1 Classical objects

I start by considering the status as Weakly emergent of so-called ‘classical’ objects: objects of the sort whose static and dynamic behaviours are appropriately treated by classical or Newtonian mechanics, understood as comprising, roughly, Newton’s three laws of motion and the gravitational and electromagnetic force laws. Such classical objects include feathers, rocks, islands, planets, and other comparatively structurally stable objects.

In what follows, I consider two cases, each of which suggests that certain classical objects are Weakly emergent, by lights of a DOF-based account, according to which the schema for Weak emergence is satisfied as a consequence of an elimination in the degrees of freedom needed to characterize the law-governed properties

(including behaviours) of a given composed entity. The first case is one involving an electrostatic sphere; the second more generally involves objects in the classical limit, in which certain quantum DOF are eliminated.

Classical mechanics as a special science

Before considering these cases, it's worth saying a few words in support of taking classical mechanics to be a special science. It is sometimes suggested that classical mechanics is at best a false but useful approximation to more fundamental theories. Hence *Cohen-Tannoudji et al.* (1977) say

Classical laws cease to be valid for material bodies [...] on an atomic or subatomic scale (quantum domain). However, it is important to note that classical physics [...] can be seen as an approximation of the new theories, an approximation which is valid for most phenomena on an everyday scale. (9)

Though natural, this line of thought can and should be resisted, at least if it is taken to suggest that classical mechanics is irrelevant to understanding the metaphysical structure of natural reality.

To start, it is characteristic of special sciences that they have restricted application; hence thermodynamics, chemistry, cell biology, geology, botany, neurobiology, psychology, and so on, traversing the ladder of theories of constitutional and biological complexity, generally fail to apply in quantum contexts, and moreover require, for their appropriate application, that certain constraints or boundary conditions be in place (as requisite for, e.g., the occurrence of thermodynamic systems, complex inorganic and organic structures, specific genetic or environmental factors, and so on). Hence while the restricted application of classical mechanics spells its failure as a fundamental theory, it remains a candidate for being a

non-fundamental theory, with something important to tell us about the nature of non-fundamental reality. Indeed, like other special sciences, classical mechanics tracks an important and distinctive level of ontological grain, and associated laws of nature, salient in circumstances wherein lower-level details are irrelevant to the dynamics of (relatively) large-scale goings-on. This is just how Feynman (1965) characterizes classical mechanics:

Newton's laws are the "tail end" of the atomic laws, extrapolated to a very large size. The actual laws of motion of particles on a fine scale are very peculiar, but if we take large numbers of them and compound them, they approximate, but *only* approximate, Newton's laws. Newton's laws then permit us to go on to a higher and higher scale, and it still seems to be the same law. In fact, it becomes more and more accurate as the scale gets larger and larger. [...] As we apply quantum mechanics to larger and larger things, the laws about the behavior of many atoms together do *not* reproduce themselves, but produce *new laws*, which are Newton's laws, which then continue to reproduce themselves from, say, micro-microgram size, which still is billions and billions of atoms, on up to the size of the earth, and above. (§19-2)

Hence it is that classical mechanics, qua special science, provides the basis for most scientific and engineering-based investigations into and treatments of ordinary objects.

Nor does the transition from force-based to energy-based mechanics—in particular, as was advisable for purposes of characterizing quantum phenomena—pose any in-principle difficulty for taking classical mechanics to be a special science in good standing. In particular, as I argue in detail in Wilson 2007, forces and energies are interdefinable, and even if fundamental quantum processes involve (e.g.) particle exchanges rather than Newtonian forces (which is not entirely clear, insofar as analogical descriptions of such exchanges tend to involve forces), it re-

mains that Newtonian forces can be seen as non-fundamental goings-on which, along with ordinary objects, are properly seen as part of the subject matter of classical mechanics.³

To be sure, there remain further questions about the status of Newtonian forces, which like other special science goings-on have been the target of anti-realist concerns. In [Wilson 2007](#), I address and respond to these concerns; to enter into these details at this point would take us too far afield, however, so I invite the interested reader to attend to that other discussion. Here I will follow Feynmann and many other scientists in seeing classical mechanics as a special science in good standing, associated with distinctive laws of nature. Seeing classical mechanics in this light provides our first route to the Weak emergence of some actual ordinary objects—namely, classical ordinary objects—by lights of a DOF-based account.

Spherical conductors

Recall that the characteristic states of an entity are those upon which its law-governed properties and behavior functionally depend; in the case of special science entities, the laws at issue are the associated special science laws. Now, ordinary objects are structurally composed of lower-level objects; and such composition involves constraints that can reduce, restrict, or eliminate the DOF needed to specify the configuration state (and/or states dependent on that state) of a composed entity E , relative to the DOF needed to characterize the system of its unconstrained composing entities e_i . And as previously discussed, only eliminations in DOF are clearly such as to block ontological and causal reduction in a way

³At least this is so for non-gravitational Newtonian forces. The prospects for seeing gravitational forces as real non-fundamental goings-on are less promising, given the present General Relativistic understanding of gravitational influence in terms of inertial motion.

supporting Weak emergence.

Consider, for example, structured entities of the sort treated by electrostatics. The electric field generated by a system of n free charged point particles will functionally depend on all $3n$ of the system's configuration DOF. If these particles come to compose a spherical conductor, however, the associated electric field will functionally depend only on the configuration DOF of particles on the boundary of the sphere. Spherical conductors are thus illustrative of *eliminations* in DOF. Consequently, a spherical conductor cannot be identified with any lower-level physical goings-on, since the elimination of configuration DOF as relevant to characterizing its electromagnetic influence removes the information needed for the lower-level laws to operate. As previously discussed, such an elimination in DOF suffices to ensure that the *Proper Subset of Powers* condition in the schema for Weak emergence is satisfied. Moreover, a spherical conductor is synchronically materially dependent on lower-level configurations, and so satisfies the other conditions in the schema for Weak emergence. In particular, the constraints associated with the existence and features of a spherical conductor are e_i -level (i.e., lower-level physical) constraints, ensuring the physical acceptability (in combination with such synchronic material dependence) of such conductors. The upshot is that spherical conductors represent a variety of classical object that is at least Weakly emergent.

The classical limit

A second and arguably much broader class of classical objects is associated with the so-called ‘classical limit’, in which (as in Feynman’s remarks above) certain quantum features cease to be relevant to the properties and behaviours of macro-

objects, in ways that are plausibly interpreted as involving an elimination (and not just a reduction or restriction) of quantum DOF.

For example, consider quantum DOF associated with spin. Classical objects are ultimately composed of quantum entities, whatever exactly these entities may be; but the characteristic states of classical objects do not functionally depend on the spins of the quantum components of these entities. Hence notwithstanding that the values of quantum parameters may in some cases lead to macroscopic differences—readings on a measurement apparatus, and the like, as in the case of Shrödinger’s cat—it remains the case that quantum DOF such as spin are eliminated (and not just reduced or restricted) from those needed to characterize entities of the sort appropriately treated by classical mechanics.

Another potential and very general source of elimination in quantum DOF in the classical limit reflects that the probabilistic values of quantum mechanical observables average out to their mean values in this limit, as per the main strategy for understanding how classical mechanics “emerges” from quantum mechanics (see, e.g., Messiah (1970, p. 215)). Forster and Kryukov (2003) provide a useful explanation by analogy of how this occurs:

It may be surprising that deterministic laws can be deduced from a probabilistic theory such as quantum mechanics. Here, curve-fitting examples provide a useful analogy. Suppose that one is interested in predicting the value of some variable y , which is a deterministic function of x , represented by some curve in the $x - y$ plane. The problem is that the observed values of y fluctuate randomly above and below the curve according to a Gaussian (bell-shaped) distribution. Then for any fixed value of x , the value of y on the curve is well estimated by the mean value of the observed y values, and in the large sample limit, the curve emerges out of the noise by plotting the mean values of y as a function of x .

To apply the analogy, consider x and y to be position and momentum,

respectively, and the deterministic relation between them to be Newton's laws of motion. Then it may be surprising to learn that Newton's laws of motion emerge from quantum mechanics as relations between the mean values of quantum mechanical position and quantum mechanical momentum. These deterministic relations are known as Ehrenfest's equations. In contrast to curve fitting, the Heisenberg uncertainty relations tell us that the quantum mechanical variances of position and momentum are not controllable and reducible without limit. Nevertheless, it is possible for both variances to become negligibly small relative to the background noise. This is the standard textbook account of how Newton's laws of motion emerge from quantum mechanics in the macroscopic limit. (1040)

It is natural to read references to probabilistic variances becoming ‘negligibly small’ as indicating that the DOF encoding such variances are eliminated in the classical limit, and so as providing support for classical objects’ being Weakly emergent—at least, on the assumption that the probabilistic aspects of quantum theory correspond to objective features of quantum goings-on. This is true for many interpretations of quantum mechanics, including versions of a Copenhagen (or ‘orthodox’) interpretation on which quantum probabilities are grounded in objective properties of particles to be, e.g., in a certain location if measured, and versions of a spontaneous collapse theory, à la [Ghirardi et al. 1986](#), on which quantum probabilities are DOF of the wavefunction, but where these DOF are understood as tracking properties of the associated quantum entities;⁴ It may also be possible to see probabilistic quantum DOF as eliminated even if the probabilities are ‘epistemic’, if what this comes to is something like statistically random behavior, and the properties and/or behavior of (some) quantum entities is functionally dependent on such randomness.

⁴Hence [Frigg \(2009\)](#) notes, “A crucial assumption of the theory is that hits occur at the level of the micro-constituents of a system [e.g.] at the level of the atoms that make up the marble”. See also Monton’s ([2004](#)) ‘object’ interpretation of a spontaneous collapse theory.

To sum up: various ordinary objects of the sort appropriately treated by classical mechanics have DOF that are eliminated as compared to the systems of lower-level entities upon which they synchronically materially depend, and hence are at least Weakly emergent by lights of a DOF-based account.⁵ Now, as above, for an object to be Weakly emergent it is required not just that at least one of its features satisfy the conditions in the schema for Weak emergence, but that the remaining features and behaviours of the object are such as to either also satisfy the conditions in the schema, or else be appropriately identified with (i.e., reducible to) lower-level goings-on. On the assumption that a given ordinary object is in fact appropriately treated by classical mechanics, this further constraint is in place, since a prerequisite for an object's being so appropriately treated is that its features and behaviours are completely determined by lower-level physical goings-on. Hence we may say, conditionally, that any object that is in fact appropriately treated by classical mechanics is Weakly emergent, *simpliciter*. And insofar as many special science objects, including rocks, planets, and the like, are reasonably supposed to be appropriately treated by classical mechanics, all such entities are reasonably supposed to be actually Weakly emergent.

6.1.2 Sortal features and functional realization

It is common to characterize artifacts as being associated with functional roles that to some significant extent serve to characterize what it is to be an object of the type in question. Hence [Mitchell \(2012\)](#) says:

⁵Note that the eliminations in DOF are relative to the system of quantum entities, not just when these are unconstrained, but when these entities stand in the relations relevant to their composing the macroscopic entities at issue, whether the system is understood as a configuration or a plurality. See [Wilson 2010b](#) for further discussion.

A chair is describable as an artifact designed to function as something suitable for humans to sit on. It is also the case that chairs are always made out of something material: wood, iron, plastic, etc. So a particular chair could also be described in terms of its material components. Chairs in general are described by their capacity to function as something humans use for sitting. Chair-function is realized by an individual entity's physical components and structure. (174)

And Searle (2001) says:

Many of the most common concepts that we use in dealing with the world, for example, concepts like ‘cars’ and ‘bathtubs’ and ‘tables’ and ‘chairs’ and ‘houses’, involve the assignment of function. (8)

(While Searle speaks of concepts, he clearly takes the notion of function as applying to the referents of the concepts.)

The view that artifacts are functionally characterized no doubt reflects that artifacts are typically understood to have been created with some purpose in mind. Here I want to develop a second route to seeing artifacts as functionally characterized, that proceeds by attention to what are sometimes called ‘sortal features’ (or ‘primary kinds’, as in Baker 1993) of such objects. As Lowe (2007) observes, “The idea that objects are sortally individuated has a long and distinguished pedigree” (see, e.g., Wiggins 2001, 150–1 and Lowe 1989, 11–13). Candidate sortal features for ordinary objects of the sort of interest here would be features expressing membership in the category at issue, such as ‘being a table’ or ‘being a statue (for artifacts).

Why is attention to sortal features useful for purposes of assessing whether a given object is metaphysically emergent? To start, note that while there are a number of different ways of characterizing sortal features (see Grandy 2016 for discussion of a number of definitions and applications), the common thread is

that a sortal feature encodes various conditions characterizing the object, both at the level of types and the level of tokens. At the level of types, an object's sortal feature specifies its ‘individuation conditions’, encoding what distinguishes objects of the type in question from other objects or types of entity. At the level of tokens, an object's sortal feature specifies its ‘identity conditions’, encoding what it takes for a particular object of the type to be identical to another particular object of the type, either at a time or over time. Identity conditions applying at a time are useful for purposes of counting how many objects of the type associated with the sortal are in a given collection (how many chairs are in the room right now?). Identity conditions applying over time—sometimes called ‘re-identity’ or ‘persistence’ conditions—are useful for purposes of determining whether an object of a given type has persisted through time, and more importantly, change.

It is the persistence conditions encoded in a given sortal feature that I want to highlight as relevant for present purposes. To start, on a standard understanding of the persistence conditions of artifacts such as tables and statues, these are compositionally flexible, in that the objects at issue can survive some (and sometimes quite considerable) changes to the lower-level configurations or pluralities upon which they depend. Relatedly, it is frequently supposed that artifacts have compositionally flexible material origins, such that a given table (‘Woody’) might have originated from a block of wood somewhat different from that from which it actually originated (see, e.g., Kripke 1972/80 and Salmon 1989); such ‘modal’ compositional flexibility might also be encoded in the sortal features of artifacts, as reflecting the conditions under which such objects⁶ are identical across worlds, as well as times.

⁶Or their counterparts, if an object cannot exist at more than one world, as per Lewis 1986.

Now, given that the sortal features of artifacts typically encode that such objects are, both temporally and modally, compositionally flexible, the question immediately arises: why so? What explains why artifacts are typically compositionally flexible? And here a plausible answer, supporting the sort of characterizations of artifacts mentioned above, is that artifacts are functionally characterized, in that what it is to be an object of the sortal type at issue is to be an object that plays or is capable of playing a specific functional role. For if artifacts are associated with functional roles, and these functional roles are sufficiently abstract that they can be implemented by multiple lower-level configurations or pluralities, then the compositional flexibility of artifacts would follow as a matter of course—effectively, as an objectual variation on the theme of multiple realizability, more commonly applied to features. So both the intuitive descriptions of artifacts, and a natural metaphysical basis for the compositionally flexible persistence conditions of artifacts, supports taking artifacts to be functionally characterized (again, such that what it is to be an object of the sortal type at issue is to be an object that plays a specified functional role); and I will henceforth assume that this account of artifacts is correct.

Now, the fact that a given artifact is irreducibly functionally characterized does not in itself guarantee that the artifact is Weakly as opposed to Strongly emergent, for reasons having to do with the bearing of mentality on the intentional specification and social implementation of the functions at issue. I'll expand on this further issue down the line. For now, it remains that such a characterization pushes towards thinking that the object is at least Weakly emergent.

To see this, first note that while numbers or other abstracta may be functionally characterized in ways that cannot (or at least cannot obviously) be cashed in causal

terms, for artifacts, the functional roles at issue are causal roles (or *vice versa*).

Hence Rosenberg (1994) says:

Causal descriptions are often called ‘functional’ role descriptions in the philosophy of science, and I will use this terminology hereafter, understanding ‘functional’ to mean simply “role in a network of causes and effects”. Thus, the functional description of an ax identifies it by its functional role, the causes and effects into which it typically enters. An ax is a tool [...] for cutting wood [that is, its effect when applied with sufficient momentum will be to sever wood into smaller pieces]. (23–24)

Given that the functional roles associated with artifacts (e.g., tables) are more specifically causal roles, it follows that what it is to be an artifact of a certain sortal type is to be an object characterized by a certain causal role, and in particular, to be an object with certain powers. These powers are encoded in the sortal feature of the artifact which, as Lowe put it, individuates the artifact. Hence at this point we can return to our usual focus on the features of a given object as the potential locus of the metaphysical emergence of the object itself.

Now, even putting aside the possibility that some of the powers associated with the sortal feature of a given artifact are fundamentally novel, in a way indicative of Strong emergence, it seems likely that many of the powers of a sortal feature will be token identical with whatever lower-level configuration or plurality serves as the dependence base for the artifact on a given occasion. Indeed, it is commonly assumed that all of the powers of artifacts are ultimately powers of their dependence base goings-on. Hence Rosenberg (1994) continues:

Now consider the class of objects the meet the purely functional characterization of an ax. They all have to be material objects and have some structural composition or other. But nothing in the functional definition requires that their handles be made of wood or that their

heads be made of steel. [...] An ax is thus “nothing but” the material out of which it is composed, even though two different axes may share no fact of structural composition in common (above the atomic level). (24)

Granting that an ax is “nothing but” its material base on a given occasion, every power of the ax (or associated sortal feature) must be identical with a power of that base (or associated sortal feature) on that occasion. [Merricks \(2003\)](#) is even more explicit, saying “a baseball and its constituent atoms cannot do any more than those atoms all by themselves (59), and later, “everything that a baseball causes is caused by its parts at some level of composition” (62). Even some of those who explicitly take into account the role of human intentions in the creation of or social practices involving artifacts seem to assume that an artifact’s lower-level dependence base is “capable of fulfilling the functions” of artifacts, as in these remarks of [Thomasson \(2007\)](#):

[I]f you think that the rules and practices that make up the game of baseball, and intentions of those who rearrange atoms into appropriate spherical shapes, are necessary for there to be baseballs, consider that for atoms to be arranged baseballwise requires atoms tightly bonded in a spherical shape of such-and-such diameter, such that they are jointly capable of fulfilling the functions of baseballs, [i.e., being such that they are] bonded by people with intentions to make baseballs that meet major league regulations, are usable and to-be-used in such games, and so on. (17)

Supposing that the powers of a given artifact do not go beyond its lower-level dependence base aggregate or plurality, as the physicalist assumes, we are a short step from seeing the artifact as Weakly emergent. For as above, and as per the compositional flexibility of artifacts, these powers are associated with an irreducible comparatively abstract functional or causal role, which can be imple-

mented by diverse lower-level base goings-on. Consider a baseball, for example. It cannot be among the powers of a given baseball to produce a certain precise weight reading on a scale, since given its compositional flexibility, the baseball might exist in the same extrinsic circumstances, and yet, if its dependence base is different, not produce that reading. In that case, the baseball (or its individuating sortal feature) has, on a given occasion, *fewer* powers than its dependence base entity on that occasion, and so satisfies the conditions in the schema for Weak emergence. And the same is true, more generally, for any compositionally flexible artifact. Since artifacts are typically (always?) compositionally flexible, artifacts are typically (always?) at least Weakly emergent.

It may be that the previous considerations also support, *mutatis mutandis*, taking natural ordinary objects to be at least Weakly emergent. After all, special science objects, ranging from cells to mountains, tectonic plates, and planets, are also commonly functionally characterized, and fall under sortal types encoding their compositional flexibility (c.f. Lowe's 2007, 519, remarks that "sortal persistence conditions [are] conditions determining what changes an object of any given sort may or may not undergo, as a matter of natural law, while remaining an object of that same sort"). Indeed, according to Rosenberg (1994), "[t]he complexity of nature above the level of the molecule is the result of selection for function and its blindness to structure" (55). In the case of natural objects, the functional specification occurs as a result of natural processes or laws as opposed to intentions or social practices, sidestepping the primary basis for thinking that functionally characterized objects might be Strongly emergent.

Be all this as it may, since the primary examples of functionally specified special science objects are drawn from the biological sciences, and we are here

focused on non-living ordinary objects, I won't here develop this line of thought, but will rather move on to a different reason for thinking that both artifacts and natural ordinary objects are at least Weakly emergent, which has not yet been discussed in the literature.

6.1.3 Metaphysically indeterminate boundaries

Ordinary objects, of either natural or artifactual varieties, typically appear to have indeterminate—that is, imprecise—boundaries. As Tye (1990) observes, “common sense has it that the world contains countries, mountains, deserts, and islands [...] and these items certainly do not appear to be perfectly precise” (396). And as has been frequently observed, even the seeming distinctness of spatial boundaries of ordinary objects such as rocks, tables, and statues, dissolves upon closer examination into an array of lower-level constituents (e.g., atoms) of decreasing but non-zero concentration in the region of a given macro-object boundary—a phenomenon which again suggests that the boundary is in some sense indeterminate.

How should such seeming boundary indeterminacy be understood? There are three main strategies. The first takes indeterminacy to have its source in how we represent the world (this is ‘semantic’ indeterminacy); the second takes indeterminacy to reflect the limits of our knowledge of the world (this is ‘epistemic’ indeterminacy); the third takes indeterminacy to have its source, somehow or other, in the world itself (this is ‘metaphysical’ indeterminacy). While some sorts of indeterminacy may be best treated in (merely) semantic or epistemic terms, object boundary indeterminacy is most naturally treated in metaphysical terms. It is unclear how indeterminacy in ordinary object boundaries might be a semantic matter, reflecting that we haven't gotten around to drawing certain precise lines:

we are not inclined to draw such lines—most importantly, because attribution of any particular precise boundary to an ordinary object would be arbitrary. Such arbitrariness renders it similarly implausible to take seeming boundary indeterminacy to reflect our inability to discern which perfectly precise boundary is in fact possessed by the object at issue, even granting (as is often assumed but which might be questioned, in light of quantum value indeterminacy, among other lower-level phenomena) that the boundaries of the lower-level aggregates or pluralities upon which ordinary objects depend are perfectly precise.

Though boundary indeterminacy is most naturally treated in metaphysical terms, many have found metaphysical indeterminacy problematic, as incoherent (Evans 1983), “not properly intelligible” (Dummett 1975), or fatally unclear (Lewis 1986). But lately, two accounts of metaphysical indeterminacy have been advanced, each of which has a claim to be both intelligible and coherent. In what follows, I’ll first argue that one of these—the account initially presented and defended in Wilson 2013a—is better than the other, so far as treating the indeterminacy of ordinary object boundaries. I’ll then show that the specific preferred treatment entails that ordinary objects are at least Weakly emergent, by lights of a determinable-based account of such emergence.

‘Meta-level’ and ‘object-level’ accounts of metaphysical indeterminacy

Here I very briefly present the two main approaches to metaphysical indeterminacy, highlighting certain of their differences along the way; see Wilson 2016a for a more detailed comparison.

At a general level, the two accounts differ structurally as regards where worldly indeterminacy is supposed to be located. On the account endorsed in Akiba 2004,

Barnes 2006, Barnes and Williams 2011, and elsewhere, what it is for a given state of affairs—i.e., an object’s having a property—to be metaphysically indeterminate is for it to be metaphysically indeterminate which precise state of affairs, out of some range of candidate precise states of affairs, is actually the case. Here the operating assumption is that all states of affairs are perfectly precise, and worldly indeterminacy consists in the world’s being unsettled about which precise state of affairs is in fact the case; hence indeterminacy is structurally located at the ‘meta-level’, so to speak, as indeterminacy among precise options.

What is it for the world to be ‘unsettled’ between various precise options? This is taken to be primitive. However, proponents of a meta-level account aim to render primitive indeterminacy comprehensible and illuminating by offering a logic and semantics of such indeterminacy along lines of the logic and semantics offered on a semantic ‘supervaluationist’ approach to indeterminacy. On the latter approach, indeterminacy reflects our language’s failing to determine a precise extension for certain expressions (e.g., ‘bald’), rendering the truth of certain propositions (e.g., ‘Bob is bald’) indeterminate; this failure is then modeled as entailing that precisifications of our language, each compatible with existing (determinate) use, disagree on the extension (such that, e.g., one precisification might deem ‘Bob is bald’ true, while another might deem this false). Analogously, on a metaphysical supervaluationist account, indeterminacy reflects our world’s failing to determine whether a given state of affairs obtains, rendering the truth of certain propositions (e.g., ‘Mount Everest is *there*’, for some precise location) indeterminate; this failure is then modeled as entailing that ‘precisificationally possible worlds’—worlds which are each compatible with existing (determinate) facts, but which disagree on whether the state of affairs obtains (such that, e.g., one pre-

cisificationally possible world might be one where Mount Everest has the precise boundary in question, while another might be one where it doesn't—and the corresponding proposition at each world would be true or false, respectively). As above, on both semantic and metaphysical supervaluationist accounts, the form of indeterminacy at issue gives rise to propositional indeterminacy; this last is registered in both cases by the introduction of an indeterminacy operator into the semantics. Finally, while on both semantic and metaphysical supervaluationist approaches, many propositions turn out to be indeterminately true, it is a stated advantage of supervaluationist approaches that if logical truth is understood as 'super-truth'—truth on all precisifications—then the theorems of classical logic are accommodated, since no matter how the language is extended or how the worldly indeterminacy gets settled, these theorems end up true. For example, it will be super-true that 'Either Mount Everest is *there* (for some precise location) or Mount Everest is not *there*'.

On the contrasting account presented and endorsed in Wilson 2013a, what it is for a given state of affairs to be metaphysically indeterminate is for the constitutive object (or objects; I won't carry this qualification through) of the state of affairs to have an irreducibly indeterminate property. More specifically, on this account, metaphysical indeterminacy in a state of affairs consists in the constitutive object's having of a determinable property, but no *unique* determinate of that property. Such a failure of unique determination can happen in two ways: first, if many candidate determinates of the determinable are instantiated, in such a way that no one determinate is non-arbitrarily taken to be the 'unique' determinate of the determinable instance (this is 'glutty' indeterminacy); second, if no candidate determinate of the determinate is instantiated (this is 'gappy' indeterminacy).

Here the operating assumption is that some states of affairs are irreducibly imprecise, and worldly indeterminacy consists in a certain pattern of determinables and determinates (namely, an object's having a determinable but no unique determinate of that determinable); hence indeterminacy is structurally located at the 'object-level', so to speak, in being located in states of affairs themselves.

At this point, one might naturally ask: how can it ever be that a determinable property is not uniquely determined? It has standardly been assumed (e.g., by [Funkhouser 2006](#), among many others) that it is a requisite feature of determinables that, for every level of specification L at which they may be determined, a determinable instance at a time has one and only one determinate at that time (see [Wilson 2017](#) for more historical discussion). However, as I argue in [Wilson 2013a](#) and ([2016a](#)), the standard supposition that determinables must be uniquely determined reflects an overly quick generalization from certain paradigm cases of determinable and determinate instances, and should be rejected as generally characterizing determinables and determinates.

Indeed, there is a compelling case to be made that even instances of colour, the feature most often highlighted as illustrative of a determinable that can be further determined, are not always uniquely determined. The case of an iridescent feather, the colour of which is relative to perspective (the feather can look red from one perspective, blue from another), in particular, is reasonably interpreted as one where the feather is coloured but is not any unique (one and only one) colour, primarily because any attribution of a specific shade, e.g., red rather than blue, as the purported unique colour would be unacceptably arbitrary. At best, one can say that the feather is red relative to one perspective (more generally, circumstance), and is blue relative to another. This and other cases of what I call

‘multiple relativized determination’ provide good reason for thinking that the conditions in *Determinable-based metaphysical indeterminacy* can be satisfied—in the case of a feather, in ‘glutty’ fashion (again, adverting to there being too many candidate determinate instances on the scene). Moreover, certain interpretations of certain quantum phenomena provide motivation for taking the conditions in *Determinable-based MI* to be satisfied in ‘gappy’ fashion, whereby no determinates are available to serve as the unique determinate of the determinable. Indeed, some have seen a determinable-based approach as offering a promising treatment of these phenomena as genuinely metaphysically indeterminate (see [Bokulich 2014](#), [Wolff 2015](#), and [Calosi and Wilson forthcoming](#) for discussion).

To return to our general comparison: besides the structural difference in where indeterminacy is located, an object-level determinable-based approach differs from a meta-level supervalueationist in two other important respects. First, on the determinable-based account, what it is for a state of affairs to be metaphysically indeterminate is not primitive; rather, this is *reducible* to a certain pattern of determinable and determinate properties or features, of the sort with which we are already familiar. Second, on a determinable-based account, metaphysical indeterminacy does not induce propositional indeterminacy. I argue for this in detail elsewhere (see especially [Wilson 2016a](#)), and will say a bit more to substantiate the claim shortly, after presenting the specific determinable-based treatment of indeterminacy in ordinary object boundaries. The underlying idea is straightforward, however, and consists in observing that it will follow from the holding of the pattern of determinables and determinates at issue in a given case that every associated relevant claim about the object and its properties will be straightforwardly true, or straightforwardly false. So, for example, it will be true that the object has the determinable

property in question, false that it has (in unrelativized fashion—about which more anon) any given determinate of the determinable, true that it has an indeterminate boundary, false that it is indeterminate which precise boundary it has, and so on.

Application to the case of ordinary object boundaries

I now want to apply the two accounts to a candidate case of indeterminacy in an ordinary object boundary, and argue that a determinable-based account better treats the case.

On a meta-level supervaluationist view, such indeterminacy would be treated as follows:

Supervaluationist MI (ordinary object boundaries): What it is for an ordinary object O to have an indeterminate boundary at a time t is for it to be primitively metaphysically indeterminate which precise boundary O has (i.e., for it to be primitively metaphysically indeterminate which precisificationally possible world is actual).

On an object-level determinable-based view, such indeterminacy would be treated as follows:

Determinable-based MI (ordinary object boundaries): What it is for an ordinary object O to have an indeterminate boundary at a time t is for O (i) to have a determinable boundary property P but (ii) for some level L of determination of P , not to have a unique level- L determinate of P at t .

For example, on a determinable-based account, what it is for Mount Everest to have an indeterminate boundary is for Mount Everest to have a determinable boundary property, but not to have a unique determinate of that boundary property. What it is for a table (call it ‘Woody’) to have an indeterminate boundary is

for Woody to have a determinable boundary property, but no unique determinate of that property. And so on for other natural and artifactual ordinary objects.

Why think that mountains, tables, and other ordinary objects have a determinable boundary property, but no unique determinate boundary property, and given that they satisfy the conditions in *Determinable-based MI*, is the indeterminacy at issue glutty or gappy? The general idea is that the indeterminacy of ordinary object boundaries is plausibly seen as reflecting that there are multiple candidate determinate boundary properties, associated with different and typically overlapping lower-level aggregates in the vicinity of the ordinary object's boundary, such that it would be inappropriately arbitrary to pick out just one of these determinate boundary properties as being the unique determinate such boundary had by the ordinary object. This sort of case, involving multiply relativized determination, involves glutty, not gappy, indeterminacy.

More specifically, consider Mount Everest. This mountain has the determinable property of having a boundary *around here* (gesturing at the general vicinity of the mountain). Now, does it make sense to take Mount Everest to have some unique maximal determinate of this determinable boundary property? It's hard to see how. To start, we are at present leaving open the possibility that some of the features of ordinary objects might be Strongly emergent; however, it seems reasonable to think that the boundary of an ordinary object is not a live candidate for being Strongly emergent. Rather, such a feature will be either reducible to or Weakly emergent from a lower-level base feature. Now, if Mount Everest's boundary is to be reducible to—that is, identical to—some lower-level feature, then on any given occasion, there must be some one microconfiguration (or plurality, as the case may be) whose boundary is identical with Mount Ever-

est's boundary. But which one? At any given time, there are multiple lower-level micro-aggregates in the vicinity of Mount Everest; these overlap as regards the clear interior regions of Mount Everest, but differ in their spatial extent: some have more dirt, some less, some are bigger, some smaller. Which microconfiguration is supposed to be 'the' microconfiguration upon which Mount Everest depends, on a given occasion, such that the boundary of Mount Everest might be thereby taken to be identical with the (assumed maximally precise) boundary of *that* microconfiguration?

There is not, it seems, any good answer to this question except one according to which the question is ill-formed. For none of these microconfigurations (or pluralities) has any more metaphysical claim to be considered 'the' dependence base configuration than any other. Correspondingly, no one boundary of any such microconfiguration has more of a metaphysical claim to be 'the' boundary of Mount Everest than any other. This observation is along lines of what Unger (1980) evocatively calls 'the problem of the many', according to which attempts to identify any given macro-object (in his preferred case: Tibbles the cat) with some lower-level configuration runs into the problem that there are many equally good candidates, at a given time, for being the lower-level dependence base entity, such that the identification of the macro-object with any one of these candidates would be unacceptably arbitrary. The problem of the many can also be seen as an interesting variation on cases of multiple realizability, where a given feature that is in fact realized by a given micro-feature might possibly be alternatively realized by a different micro-feature. In the case of Mount Everest, the multiplicity of realization of the determinable boundary property is actual, not merely possible; and as in Unger's problem, any attempt to pick out one of the microconfigurations

in the vicinity of Mount Everest and attribute its maximally precise boundary to Mount Everest as the unique determinate of the mountain's determinable boundary property would be unacceptably arbitrary.

The previous considerations indicate that the boundary of Mount Everest is not reducible to any specific determinate boundary property; nor is there any hope of identifying the mountain's boundary with some complex (e.g., disjunctive or conjunctive) combination of the boundaries of its micro-aggregates—not just because such disjunctive or conjunctive attributions do not make clear metaphysical sense, but because, first, the posit of any such complex would fail to resolve the problem of there being too many candidate lower-level boundaries at a given time, and second, higher-order indeterminacy is likely to render it unclear exactly which microconfigurations and associated boundaries are candidates for realizing Mount Everest and its boundary.

Given that the boundary of Mount Everest is not Strongly emergent (as is plausible), then if the mountain were to have a unique maximally precise boundary, it would have to be appropriate to identify the boundary of the mountain (at a time) with the precise boundary of a given microconfigurations (at that time); but this is not appropriate, since as above the presence of multiple micro-aggregates in the vicinity of Mount Everest entails that any such identification would be inappropriately arbitrary. So Mount Everest does not have a unique maximally precise boundary property. Second, as previously, Mount Everest does have a determinable boundary property. It follows that the state of affairs of Mount Everest's having a boundary satisfies the conditions in *Determinable-based MI*, for the property ‘being a boundary’.

Moreover, the satisfaction of the conditions here is indicative of ‘glutty’ rather

than ‘gappy’ indeterminacy. Like the iridescent feather, the determinable boundary of Mount Everest admits, at a time, of multiple determinations. As with the case of the feather, some may see it as appropriate to attribute the precise boundaries to the mountain, albeit in relativized fashion; others may maintain that even relativized attributions of precise boundaries are inappropriate unto characterizing ordinary objects. This is a choice point; but either way, in cases of such multiple relativized determination, it is not an option to attribute any precise boundary to the mountain (more generally, ordinary object) in unrelativized fashion.

Which approach to metaphysical indeterminacy better treats the case of boundary indeterminacy? I’ll now argue that a determinable-based approach is clearly better, along pretty much every dimension; in the next section I’ll connect this result to the question of whether any ordinary objects are actually Weakly emergent.

I start with some general advantages of a determinable-based account, which apply also to this case. To start, as above, on a supervaluationist account, metaphysical indeterminacy is taken to be primitive, whereas on a determinable-based account such indeterminacy is reduced to a pattern of determinable and determinate property instantiations. This difference expands into three advantages for a determinable-based account. First is that a determinable-based account is metaphysically more parsimonious than a supervaluationist account, in characterizing this phenomenon using off-the-shelf resources as opposed to introducing a new ‘indeterminacy’ primitive. Second, a determinable-based account has a claim to be metaphysically more intelligible, in characterizing indeterminacy in terms—namely, determinable and determinate features—with which we are already both experientially and theoretically familiar (for expansion on the sources and content of this understanding, see Wilson 2017). Third, the reductive strategy underly-

ing a determinable-based account does not introduce propositional indeterminacy, and hence abnegates the need to introduce an indeterminacy operator into the semantics. Relatedly, on a determinable-based account of indeterminacy, classical propositional logic and semantics can stay just as they classically are.⁷

For present purposes, however, the most important reason to prefer a determinable-based account of boundary indeterminacy is that a supervaluationist account incorrectly presupposes that it would be appropriate, in principle, to assign a perfectly precise boundary to the object in unrelativized fashion. A supervaluationist approach thus imports the same underlying problematic supposition as do semantic or epistemic approaches to boundary indeterminacy. The metaphysical indeterminacy in ordinary object boundaries cannot be a matter of the world's being unsettled about which precise boundary Mount Everest has, because it is part of what it is to be Mount Everest is to be an object that does *not* have a precise boundary. A determinable-based account accommodates this central feature of ordinary objects—a feature that also provides a natural basis for the compositional flexibility of ordinary objects (unlike a supervaluationist account, which does not accommodate this central feature). Coupled with the general reasons to prefer a determinable based account, this provides good reason to prefer a determinable-based treatment of ordinary object boundary indeterminacy.

⁷Consider our test case, involving Mount Everest's indeterminate boundary. On this treatment, it will be true that Mount Everest has a boundary, and false that Mount Everest has a perfectly precise boundary. It will be false, for every unrelativized attribution of such a precise boundary, that Mount Everest has that boundary. It may or may not be true, depending on the aforementioned choice point and on the specific facts at hand, that Mount Everest has such a precise boundary in relativized fashion. And so on. Similarly, all the usual theorems of classical logic and semantics hold.

The connection to Weak emergence

The previous result provides a basis for taking the vast store of ordinary objects (which typically have indeterminate boundaries, among other properties) to be at least Weakly emergent. There are two routes to this conclusion. One quick route proceeds as something of an argument by cases: given that ordinary objects are dependent on lower-level aggregates in a way supportive of substance monism, and given that the boundaries of ordinary objects are neither Strongly emergent from nor reducible to the boundaries of micro-aggregates, the remaining option is that the boundaries of ordinary objects are non-reductively realized—that is, Weakly emergent, in which case ordinary objects are at least Weakly emergent.

A somewhat longer but more substantive route proceeds by looking more closely at the analogy between multiple determination and multiple realization, and considering its implications. As above, the determinable boundary properties of ordinary objects are multiply determined at a time (albeit in relativized fashion), in a way that echoes the phenomenon of multiple realizability, with the main difference being that in cases of multiple determination the multiplicity is actual, rather than merely possible.⁸ Now, as previously discussed, while there are reductive strategies for accommodating cases of multiple realizability, attention to the relations between the powers associated with a multiply realized feature and the powers associated with any one of its realizing features suggests that reductive strategies must fail, since multiply realized types have fewer powers than each of their realizing types (reflecting effects that the latter can cause that depend on differences between the realizing types), and this proper subset relation will hold

⁸Even this difference is superficial, in that it has been observed that certain features might be actually multiply realized at a time.

between tokens—hence satisfying the conditions in the schema for Weak emergence.

Here I want to observe that in cases where a given determinable admits of multiple (relativized) determination, the determinable type will also be reasonably taken to have fewer powers than each of the more determinate types (again reflecting effects that the latter types can cause that depend on differences between determinate types, at a given level of specification). Again, for already-rehearsed reasons, this proper subset relation will also hold between tokens of determinable and determinate types—hence satisfying the conditions in the schema for Weak emergence. That the multiple determinates are all actually instantiated (albeit in relativized fashion) in the case of the determinable properties of ordinary objects, as opposed to being merely possibly instantiated (in non-relativized fashion) in typical cases of multiple realization, doesn't affect the relation between powers between the higher-level feature and lower-level feature at issue. Hence given that the determinable properties of ordinary objects satisfy the conditions in *Determinable-based MI* in gluttony fashion (again, involving multiply relativized determination), it will follow that the determinable property satisfies the conditions in the schema for Weak emergence, such that the ordinary object at issue will be at least Weakly emergent.

Adding further support to this result is that the connection between realization and determination has in past decades been recognized, and indeed the relation between determinables and determinates (in particular, one allowing that determinable and associated determinate instances may be possessed by different objects) has been appealed to (by MacDonald and MacDonald 1986, Yablo 1992, Wilson 1999 and 2009, Shoemaker 2000/2001, and others) as providing the basis

for a general account of realization.⁹ Considerations of how best to treat the metaphysical indeterminacy of ordinary object boundaries thus provide independent support for thinking that ordinary objects are at least Weakly emergent, by lights of a determinable-based account of Weak emergence.

6.2 Are ordinary objects Strongly emergent?

The considerations of the previous section suggest that many ordinary objects, of both natural and artifactual varieties, are at least Weakly emergent, in having at least one feature satisfying the conditions in the schema for Weak emergence. Might any of these objects be, moreover, Strongly emergent? While the possibility of there being natural Strongly emergent natural remains an empirical possibility, the best case for there being Strongly emergent ordinary objects attaches to the case of artifacts. In the remainder of this section I'll briefly lay out this case. I will not (in this chapter) assess this case, since as we will see, whether it goes through ultimately hinges on the status as Strongly emergent, or not, of conscious mentality—a topic to be discussed in the next chapter.

6.2.1 Two routes to the Strong emergence of artifacts

Recall that for an object to be Strongly emergent, it must have at least one feature associated with a fundamentally novel power—a power not associated with whatever serves as the dependence base feature on a given occasion. As above,

⁹Most objections to determinable-based accounts of realization have focused on whether such realization would be suited for accommodating mental-physical realization, as in Ehring 1996, Worley 1997, Funkhouser 2006, Walter 2006; see Wilson 2009 and Wilson 2017 for responses to these concerns.

philosophers commonly assume that every power of (every feature of) an artifact, on a given occasion, is uncontroversially a power of (some feature of) the lower-level aggregate or plurality upon which it depends, on that occasion.

One might think that this is a reasonable assumption. After all, while it might remain at least an open question whether mental features and associated entities have fundamentally novel powers, associated (as on the British Emergentist view, as well as my own interaction-relative version of Strong emergence) with non-physical fundamental interactions, it's considerably less plausible that there are fundamental interactions governing the behaviours of baseballs, tables, and the like.

This is too quick, however. For artifacts are associated with functional roles in which mentality is deeply implicated, in ways that, were the associated mental features to be Strongly emergent, might open the door to artifacts' having fundamentally novel powers.

Mentality potentially enters into the functional roles of artifacts in two ways. First, mentality typically enters into the identification of a given functional role with a given artifact. Second, mentality is, one way or another, typically constitutive of the role in question. Hence to be a baseball is to be an object that enters into a highly complex set of human practices, ranging from the playing of the game itself and all that mentally entails, to the emotional and even ethical appreciation that is a concomitant of the game, its players, and its values. And to be a statue is to be an object that enters into a highly complex set of human practices, ranging from the creation (or appropriation) of a given object of aesthetic value, which may be perceptually and cognitively appreciated (or problematized, or dismissed), which may have economic as well as other value, and so on. Mental

features are thus typically implicated in the characterization of artifacts, both via the intentional specification of the functional roles associated with artifacts, and more generally, via the social practices which typically are to some large extent constitutive of the roles in question. As such, we can ask the question: supposing that the mental features in question were Strongly emergent, would it reasonably follow that the associated artifacts were Strongly emergent?

Now, even on the assumption that the mental features in question are Strongly emergent, the bare involvement of a given mental state in either of these aspects might not be enough to render the associated artifact Strongly emergent. Were someone with a Strongly emergent mental state to stipulate that a given rock is to be deemed a ‘Faller’—an artifact whose role is to fall when dropped in accord with the laws of nature, their doing so would probably not be seen as sufficient unto bestowing a fundamentally novel power on the Faller.

Still, for other functional roles, and contingent on the status of normative, intentional, and perceptual features, it seems plausible to think that an intentional act of stipulation could be such as to reasonably be seen as bestowing upon the artifact in question a fundamentally novel power.

In any case, at present we can say this much: insofar as various forms of mentality typically enter into both constructing and constituting the functional roles associated with artifacts, and insofar as artifacts are plausibly taken to be essentially characterized by these sorts of mentally informed features, it follows that the status of artifacts as potentially Strongly emergent ultimately hinges on the status of persons or other bearers of mental features.

It is worth connecting this result to Merricks’s conclusions in his (2003). There, recall, Merricks argues that while ordinary objects, such as baseballs, do

not exist, persons do exist—precisely since he thinks persons have fundamentally novel powers as compared to their dependence base goings-on. In other words, for Merricks, persons are Strongly emergent. The present discussion shows that there is a tension in Merricks’s view, at least for the artifacts that are the focus of his discussion; for supposing that the existence or features of some artifacts crucially depends on the existence or features of persons and their mentality, the Strong emergence of persons and associated mental features might ‘infect’, so to speak, any artifacts that they create. And in that case, Merricks might be committed to the existence of baseballs and other ordinary objects, after all.

6.3 Concluding remarks

Let’s sum up the results of this chapter. I’ve argued that there are three different cases to be made for the claim that some ordinary objects are Weakly emergent, or at least Weakly emergent, First, classical objects are arguably Weakly emergent, by lights of a DOF-based account, thanks to the elimination of quantum DOF in the classical limit, Second, artifacts are arguably at least Weakly emergent, by lights of a functional realization account, thanks to the support for such a treatment that can be extracted from the compositionally flexible persistence conditions typically encoded in the sortal features of artifacts. Third, both natural and artifactual ordinary objects are arguably at least Weakly emergent, by lights of a determinable-based account, thanks to the metaphysical indeterminacy of the boundaries of ordinary objects, which indeterminacy is best treated in determinable-based terms.

Moreover, the possibility remains, especially for artifacts, that these are ul-

timately not just Weakly but Strongly emergent, due to the role mentality plays both in the specification and the constitution of the functional roles which are typically associated with such ordinary objects. As such, the status of any artifacts as Strongly emergent ultimately depends on the status of conscious mentality, to be explored in the next chapter. Irrespective of the answer, these considerations suggest that those who are committed to the Strong emergence of persons and their conscious mental features might well find themselves with considerably more entities in their ontology than they are on record as accepting.

Before moving on to the next chapter, I pause to note one other implication of the results of this chapter. In her (2010), Thomasson discusses a number of metaphysical concerns against ordinary objects, and argues that these arise from inappropriately applying broadly scientific methodology in service of answering philosophical questions:

Although the arguments against the existence of ordinary objects do not rely on any particular piece of scientific knowledge, many of them do rely on a certain scientific approach to metaphysics: the view that metaphysics is of a piece with (and indeed part of the same total theoretical enterprise as) natural science. [...] Lying behind many of the arguments against ordinary objects is the assumption that metaphysics is engaged in explanatory theory construction following the same principles as those governing the natural sciences—or (perhaps better) engaged in one and the same enterprise of constructing the best ‘total theory’. (596)

Thomasson takes such an assimilation of the metaphysical enterprise to broadly scientific desideratum to be problematic. So, for example, in addition to expressing concerns about metaphysics aiming, like (or with) the natural sciences for a ‘total theory’, she notes,

Lying behind many of the arguments against ordinary objects is the assumption that meta-physics is engaged in explanatory theory construction following the same principles as those governing the natural sciences—or (perhaps better) engaged in one and the same enterprise of constructing the best ‘total theory’. Straightforward causal redundancy arguments rely on accepting the Eleatic Principle: that we should only accept the existence of those entities that ‘make a causal difference’. This may be a reasonable principle for deciding whether or not we should accept the existence of neutrons as well as protons and electrons—theoretical particles posited to explain experimental data. But whether or not it carries over to the question of whether we should ‘posit’ tables ‘as well as’ the particles that make them up is less clear. (596)

One might think that investigations in the metaphysics of science are not in fact “scientific”, at least in that the methodology of metaphysicians of science is not empirical *per se*. However, a related question, encouraged by the fact that the debates over the status of special science entities and ordinary objects have proceeded in some independence of each other, might be that the issues in the two debates are ultimately quite different, and it would be in some sense a mistake to try to apply strategies in the one debate to problems in the other.

One thing to say here is that it would be in some sense surprising if the metaphysical treatment of special science entities were really different from that of ordinary inanimate objects in general, since after all, many of these objects are special science entities. Indeed, from the perspective of classical mechanics, every ordinary object is in some sense a special science entity.

In any case, the results of this chapter show that there is no in-principle reason to think that investigations into the existential status of ordinary objects should proceed in a way much different from such investigations into the corresponding status of special scientific entities. Thomasson’s suggestion to the contrary seems

to have been primarily motivated by thinking, first, that the usually stated concerns with ordinary objects (e.g., Kim's causal overdetermination concern) arise from trying to give scientific and ordinary objects (including artifacts) a unified treatment, and second, taking the concerns as attaching to scientific goings-on not to admit of any good answers. But as I have argued, there are good responses to the concerns at issue, whether natural or artifactual ordinary objects are at issue. Nothing stands in the way, at least in principle, and modulo the import of mentality to be next considered, of providing a systematic treatment of both natural and artifactual ordinary objects, as either Weakly or Strongly emergent.

Chapter 7

Consciousness

I turn now to considering whether consciousness of the sort that we and other creatures enjoy is either Weakly or Strongly emergent. There are, to be sure, many forms or species of consciousness, including perceptual awareness of the external world, conscious awareness of internal states (e.g., pain), and self-consciousness—consciousness of ourselves as conscious beings. As Chalmers (1996) notes,

Conscious experiences range from vivid color sensations to experiences of the faintest background aromas; from hard-edged pains to the elusive experience of thoughts on the tip of one's tongue; from mundane sounds and smells to the encompassing grandeur of musical experience; from the triviality of a nagging itch to the weight of a deep existential angst; from the specificity of the taste of peppermint to the generality of one's experience of selfhood. All these have a distinct experienced quality. All are prominent parts of the inner life of the mind. (4)

Notwithstanding this diversity, little in this chapter hinges on differences between these forms of consciousness; so unless some specific variety is under discussion, I will speak generically of consciousness or conscious awareness (or associated

mental features), which may have as its seeming object the external world, one's internal states, or (as a special case of the latter) consciousness itself.

Interestingly, and in some contrast to the cases of complex systems and ordinary objects considered in previous chapters, the cases proffered for the emergence of consciousness are most frequently aimed at showing that consciousness is Strongly rather than merely Weakly emergent. Hence Chalmers (2006a) says, "there is exactly one clear case of a strongly emergent phenomenon, and that is the phenomenon of consciousness" (3). In turn, it is often suggested that the main motivation for taking consciousness to be Strongly emergent reflects a commonly acknowledged failure for consciousness to be predictable from or explainable in terms of lower-level physical phenomena. Hence Broad (1925) says:

[An archangel] would know exactly what the microscopic structure of ammonia must be; but he would be totally unable to predict that a substance with this structure must smell as ammonia does when it gets into the human nose. The utmost that he could predict on this subject would be that certain changes would take place in the mucous membrane, the olfactory nerves and so on. But he could not possibly know that these changes would be accompanied by the appearance of a smell in general or of the peculiar smell of ammonia in particular, unless someone told him so or he had smelled it for himself. (71)

Bedau (2010) says:

Our inability to have found any plausible micro-level explanation for conscious mental states might reflect just our ignorance, but another possibility is that these phenomena really are strongly emergent. (60, note 3)

And Nagel (1974) says:

[W]e have at present no conception of what an explanation of the physical nature of a mental phenomenon would be. Without consciousness the mind-body problem would be much less interesting. With consciousness it seems hopeless. (436)

Indeed, existing arguments offered as supporting the Strong emergence of consciousness nearly all rely, one way or another, upon the supposition that consciousness, or certain of its characteristic features (or aspects¹), lies beyond the explanatory reach of any lower-level physical goings-on.

As the reader may suspect, the presence of even an insuperable or ‘in-principle’ explanatory gap can’t be the whole supporting story, however, since—to take one previously discussed example—many complex non-linear phenomena are clearly physically acceptable, in spite of having features that are, in some relevant sense of ‘insuperability’ (as, e.g., beyond the access of any empirically possible determination), insuperably unpredictable or otherwise unexplainable in lower-level physical terms. What’s moreover at issue in discussions of consciousness is that the explanatory gaps are taken to be metaphysically significant, in reflecting not just broadly mathematical barriers to explanation, such as non-linearity and associated sensitivity to initial conditions, but rather that certain characteristic features of consciousness—notably, subjective or qualitative aspects of conscious experience—depart so greatly from lower-level physical features that this divergence provides reasonable grounds for thinking that no physicalist account of consciousness of either reductive or non-reductive (i.e., Weakly emergent) varieties could possibly be correct.² In being so motivated, explanatory gap arguments in

¹Talk of ‘characteristic features’ of consciousness or conscious states should be understood as broadly neutral on exactly such features enter into such states (in particular, it isn’t intended to suggest that characteristic features of conscious states are second-order features of features).

²An early expression of this line of thought can be found in Ewing’s (1951) reasons for rejecting the identity theory: “Nineteenth-century materialists were [...] inclined to identify thinking, and mental events generally, with processes in the central nervous system or brain. In order to refute such views I shall suggest your trying an experiment. Heat a piece of iron red-hot, then put your hand on it, and note carefully what you feel. You will have no difficulty in observing that it is quite different from anything which a physiologist could observe, whether he considered your outward behaviour or your brain processes. The throb of pain experienced will not be in the least like [...] anything described in textbooks of physiology as happening in the nervous system or

favour of the Strong emergence of consciousness deserve independent consideration. As I will argue, however, these arguments are ultimately unconvincing, for reasons that have not been previously much explored, notwithstanding the considerable critical attention that has been given to these arguments. More generally, I will argue that while it remains a live empirical possibility that consciousness is Strongly emergent, at present we have no compelling philosophical or empirical motivation for taking this to actually be so.

On the assumption that consciousness isn't (doesn't turn out to be) Strongly emergent, might it be Weakly emergent rather than ontologically reducible? Here I again present an underexplored reason to endorse a positive answer to this question. The argument proceeds via the claim that qualitative conscious states—e.g., states of conscious awareness of colours or pains—are typically determinable rather than (maximally) determinate, in a way that defensibly renders them suitable (again, assuming that they are not Strongly emergent) for being realized in determinable-based fashion, and hence Weakly emergent.

The overall conclusion, as in the previous chapter, is that there are reasons to think that consciousness is at least Weakly emergent from, as opposed to ontologically reducible to, lower-level physical goings-on.

brain. I do not say that it does not happen in the brain, but it is quite distinct from anything that other people could observe if they looked into your brain. [...] We know by experience what feeling pain is like and we know by experience what the physiological reactions to it are, and the two are totally unlike. [...] the difference is as plainly marked and as much an empirical matter as that between a sight and a sound. The physiological and the mental characteristics may conceivably belong to the same substance [...] but at least they are different in qualities, indeed as different in kind as any two sets of qualities." (101–102).

7.1 Is consciousness Strongly emergent?

In this section, I consider two different forms of argument that have been or might be offered in support of consciousness's being Strongly emergent. The qualifier 'might be' reflects that there is some variation in what proponents of these arguments take their conclusions to be. For present purposes, what is important is that the considerations these authors raise can and have been offered in support of taking consciousness to be physically unacceptable, and hence (along with other premises) taking consciousness to be Strongly emergent, in the sense operative in the associated schema for emergence. I'll start by considering Nagel's and Jackson's arguments, which fall broadly under the rubric of 'knowledge arguments'; I'll then turn my attention to Chalmers's 'conceivability' or 'zombie' argument.

7.1.1 The knowledge arguments

The suggestion that even complete knowledge of physical goings-on might not suffice for knowledge of certain aspects of conscious experience can be found throughout history. Echoes of the basic idea can be found in Leibniz's (1714) 'Mill argument':

[W]e must confess that perception, and what depends upon it, is inexplicable in terms of mechanical reasons, that is through shapes, size, and motions. If we imagine a machine whose structure makes it think, sense, and have perceptions, we could conceive it enlarged, keeping the same proportions, so that we could enter into it, as one enters a mill. Assuming that, when inspecting its interior, we will find only parts that push one another, and we will never find anything to explain a perception. (§17; GP: VI, 609/AG: 215)

Leibniz concluded that perceptual consciousness should be understood as located

in a ‘simple substance’ rather than ‘the composite’. As previously, the British emergentist Broad more specifically offered a version of a knowledge argument, involving a mathematical archangel who knows all the mechanistic facts about chemical goings-on, in support of taking qualitative aspects of conscious experience to be Strongly emergent. More recently, both Nagel (1974) and Jackson (1982 and 1986) have offered arguments aiming to show that one could have complete physical knowledge of some entity or subject matter, but nonetheless fail to know certain facts pertaining to conscious states associated with the entity or subject matter; from this they conclude that physicalism, at least as commonly understood, is false. Coupled with the assumption that states of conscious awareness synchronically materially depend on lower-level physical states, such a conclusion would provide positive motivation for consciousness’s being Strongly emergent.

Nagel’s discussion proceeds by attention to the question of what we ought to be able to understand about the conscious experience of creatures relatively different from ourselves.³ He first connects the notion of consciousness to what he calls ‘the subjective character of experience’:

[N]o matter how the form may vary, the fact that an organism has conscious experience at all means, basically, that there is something it is like to be that organism. There may be further implications about the form of the experience; there may even (though I doubt it) be implications about the behavior of the organism. But fundamentally an organism has conscious mental states if and only if there is something that it is like to be that organism. We may call this the subjective character of experience.⁴ (436)

³See Feigl 1959 for a historical antecedent of attention to this issue, involving a Martian in possession of all the physical facts about humans, who would nonetheless “be lacking completely in the sort of imagery and empathy which depends on familiarity (direct acquaintance) with the kinds of qualia to be imaged or empathized” (431).

⁴See also Farrell 1950 and Feigl 1967, 139–140.

Nagel goes on to more specifically consider the case of a bat—a creature plausibly having conscious experience, but of a very different sort than any to which we are privy—and to raise to salience difficulties in our being able to comprehend what it is like to be that sort of creature. As he observes, there is no clear way to extrapolate from our own perceptual experience to that of a creature who navigates the world using echolocation. And nor, crucially, would any amount of knowledge of the physiology of bats, whether pitched at a lower or higher level of physical goings-on, enable us to gain knowledge of this comparatively alien form of conscious experience. Nagel concludes that such knowledge is beyond our ken, and goes on to suggest that the latter gap spells trouble for any view on which consciousness (and its subjective character) is either ontologically reducible to or non-reductively realized in (that is, Weakly emergent from) lower-level physical goings-on:

While an account of the physical basis of mind must explain many things, this appears to be the most difficult. [...] if physicalism is to be defended, the phenomenological features must themselves be given a physical account. But when we examine their subjective character it seems that such a result is impossible. The reason is that every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view. (437)

Nagel sees here a “divergence between [...] two kinds of conception: subjective and objective” (438). An account of consciousness as physically acceptable must include a treatment of the subjective character of consciousness, but on the assumption that physical theory is solely concerned with the objective or third-person point of view, any such treatment will be, he maintains, “impossible”.

Jackson’s (1986) knowledge argument (for an early variation on the theme,

see Jackson 1982) also aims to establish that there is an insuperable, metaphysically significant gap in relevantly comprehensive knowledge of certain physical goings-on and knowledge of what it is like to have certain conscious experiences. Jackson's focus differs from Nagel's in highlighting the gap as attaching to the qualitative aspects of conscious experiences, as per the following thought experiment:

Mary is confined to a black-and-white room, is educated through black-and-white books and through lectures relayed on black-and-white television. In this way she learns everything there is to know about the physical nature of the world. She knows all the physical facts about us and our environment, in a wide sense of 'physical' which includes everything in *completed* physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this, including of course functional roles. If physicalism is true, she knows all there is to know. For to suppose otherwise is to suppose that there is more to know than every physical fact, and that is just what physicalism denies. [...] It seems, however, that Mary does not know all there is to know. For when she is let out of the black-and-white room or given a color television, she will learn what it is like to see something red, say. This is rightly described as *learning*—she will not say "ho, hum." Hence, physicalism is false. This is the knowledge argument against physicalism in one of its manifestations. (291)

As Jackson qualifies, at issue here is not that Mary comes to learn something about her own experiences, but rather that Mary comes to learn something about the experiences of others: "The trouble for physicalism is that, after Mary sees her first ripe tomato, she will realize how impoverished her conception of the mental life of *others* has been *all along*. She will realize that there was, all the time she was carrying out her laborious investigations into the neurophysiologies of others and into the functional roles of their internal states, something about these people

she was quite unaware of" (292).

How might a physicalist best respond to the knowledge arguments? In what follows, I'll focus on Jackson's version; the application to Nagel's argument and other variations on the theme will be clear. I'll moreover focus on the version of Jackson's argument presented in [Nida-Rumelin 2015](#), which insightfully articulates the key premises and conclusions:

P1: Mary has complete physical knowledge about human colour vision before her release.

C1: Therefore, Mary knows all the physical facts about human colour vision before her release. (P1)

P2: There is some (kind of) knowledge concerning facts about human colour vision that Mary does not have before her release.

C2: Therefore, there are some facts about human colour vision that Mary does not know before her release. (P2)

C3: There are non-physical facts about human colour vision. (C1, C2)

The knowledge arguments have generated an enormous amount of literature, and a full treatment of all the variations on, ramifications of, and responses to these arguments is beyond the scope of this chapter (see [Nida-Rumelin 2015](#) for extensive discussion and references). Here I'll present my preferred strategy, which is apparently not much on the books—namely, to deny P1. Along the way I'll positively contrast this approach with certain other more popular strategies.

The ‘Incomplete Physical Knowledge’ strategy

According to P1 of the knowledge argument, Mary has complete physical knowledge about human colour vision before her release. The strategy of denying P1 is not much considered; hence Nida-Rumelin passes it over, saying, “it seems hard to deny that it is in principle possible to have complete physical knowledge about human colour vision (or about an appropriately chosen part thereof). If so, premise P1 should be accepted as an appropriate description of a legitimate thought experiment”. (Note that Nida-Rumelin’s remarks here presuppose not just that it is in-principle possible to have complete physical knowledge about human colour vision, but more strongly that Mary could have such knowledge prior to her release from the black-and-white room.) A good case can be made for denying P1, however, as per what I hereby dub the ‘Incomplete Physical Knowledge’ strategy.

To start, note that nearly all participants to the debate over the import of the knowledge arguments take for granted that physical knowledge does not include knowledge of subjective or qualitative aspects of consciousness. Hence Nagel links physical knowledge to physical theory, and takes the latter to concern only “objective” phenomena, such that it would be “impossible” for physics to accommodate a subjective point of view, and Jackson motivates P1 on grounds that “it is plausible that lectures over black-and-white television might in principle tell Mary everything in the physicalist’s story. You do not need color television to learn physics or functionalist psychology” (295). Opponents of the knowledge argument similarly take for granted that complete physical knowledge fails to be knowledge of any subjective or qualitative aspects of reality there might be. This is true on the ‘Ability hypothesis’, according to which what Mary gains upon leaving her room is a new ability rather than a new piece of knowledge (see, e.g.,

Lewis 1988, Nemirov 2006). It is true on what Nida-Rumelin calls the ‘Complete Physical Knowledge without Knowledge of all the Physical Facts’ strategy, which proceeds by accepting P1 but denying the inference from P1 to C1 (according to which Mary knows all the physical facts about human colour vision before her release), on grounds that since physical knowledge is not of subjective or qualitative facts, and since some physical facts are subjective or qualitative, Mary’s having complete physical knowledge does not entail her knowing all the physical facts (see, e.g., Harman 1990, Flanagan 1992, and Alter 1998). It is true on the ‘Acquaintance hypothesis’, which again grants that Mary has complete physical knowledge, and moreover knows all the physical facts, prior to leaving her room, but maintains that what she gains upon leaving the room is ‘knowledge by acquaintance’, which in not being ‘informational’, doesn’t threaten physicalism. And it is true on the popular ‘two ways’ or what Nida-Rumelin calls the ‘New Knowledge; Old Fact’ approach, according to which, while Mary does gain new knowledge upon leaving her room, this is simply knowledge about a different, qualitatively informed way of thinking about a fact she already knew, such that the supposition in P1 that prior to leaving her room Mary had complete physical knowledge, sufficient unto knowing all the physical facts, is not undermined.

But need the physicalist agree that physical knowledge is ‘objective’ in the sense, in particular, of failing to be of any subjective or qualitative aspects of reality there may be? No, for two reasons.

First, such a view is in tension with physicalism. The ontological physicalist of whatever stripe maintains that the (lower-level; henceforth this qualification will be assumed) physical goings-on provide a basis for all of natural reality. Given that natural reality includes consciousness and its associated subjective and quali-

tative aspects, the physicalist should correspondingly maintain that some physical goings-on—namely, those that are either identical with (on a reductive physicalist view) or which realize (on a non-reductive physicalist view) these aspects of consciousness—involve the instantiation of subjective and/or qualitative features which (again, given the supposed truth of physicalism) are themselves physical (or physically acceptable) features. The questions of whether these features should be seen as had by the highly complex physical microconfigurations (or pluralities) that accompany the instantiation of the features, and of whether, if so, the associated microconfigurations (or pluralities) should be taken to have a first-person perspective, represent choice points: what answers are given to these questions will depend on further details about the form of physicalism at issue, as well as what account is given of the creatures having such conscious states. However the physicalist answers these questions, they are in any case committed to complex lower-level physical goings-on' involving the instantiation of lower-level physical features which, in being either identical to or such as to realize subjective and qualitative aspects of natural reality, are themselves properly deemed subjective and qualitative.

Now, complete physical knowledge is, as a matter of definition, complete knowledge of any and all actual and/or potential physical goings-on, including knowledge of the potential instantiation of any subjective and qualitative physical features which serve, either by way of identity (on a reductive view) or in some other intimate fashion (on a non-reductive view) for any subjective and qualitative aspects of consciousness there might be.⁵ How is one to gain knowledge

⁵Nothing in this claim turns on whether knowledge has to be broadly propositional—i.e., of states of affairs as opposed to properties or other features—or not, since knowledge of (the potential instantiation) of properties can be translated, if so desired, into propositional terms (as, e.g., knowledge that a certain property would be instantiated under such-and-such circumstances).

of subjective and qualitative physical features? The physicalist can and should maintain, plausibly enough and without any clear threat to their position, that such knowledge can only be gained by acquaintance—that is, by being in position to experience these subjective and qualitative features either directly (on a reductive physicalist view) or indirectly (on a non-reductive physicalist view). Consequently, the physicalist can and should deny (in particular) that Mary has complete physical knowledge about human colour vision before her release—that is, they can and should deny P1.

I will shortly offer a diagnosis of why, in spite of its being in clear tension with the truth of physicalism, P1 has been taken for granted even by opponents of the knowledge argument. First, though, it's worth noting that the Incomplete Physical Knowledge strategy improves on certain alternative strategies of response to the knowledge argument, while arguably preserving the insight common to many of these responses—namely, that the fact that certain kinds of experience are prerequisite to grasping certain features of reality is no barrier to the physical acceptability of these features. Consider, for example, Conee's (1994) case for taking the Acquaintance hypothesis to block the physical unacceptability of phenomenal qualities:

The physical facts may include every fact about qualia. Still, the physical story does ‘leave out the qualia’, in the sense that knowledge of the physical facts does not imply knowledge of the qualia. Gaining knowledge of phenomenal qualities, though, is no more than a matter of making their acquaintance by attentive experience. It requires only entering into a new cognitive relation to the qualities, not learning any new information. It gives us no reason to doubt that everything is physical. (148)

As Nida-Rumelin points out, a proponent of the knowledge argument is likely to

respond that if the having of an experience with the qualitative physical property $[Q]$ in question is required for knowledge of it by acquaintance, such acquaintance is moreover “a necessary condition for being able to know (in the relevant sense) that an experience has Q ”. What has not been previously appreciated is that the physicalist can and arguably should simply grant that acquaintance is a necessary condition for knowing certain physical facts—namely, those providing a subjective or qualitative basis for any subjective or qualitative aspects of consciousness there may be. And similarly for the usual responses to the Ability hypothesis, the ‘two-ways’ strategy, and the ‘Complete Physical Knowledge without Knowledge of all the Physical Facts’ strategy. Rather than stand opposed to the intuitively compelling take on Mary’s response to seeing a ripe tomato—‘so *this* is what it is like to see red!’—according to which she does come to know a new fact after leaving her room, the physicalist can rather simply agree that Mary gains new knowledge, and then ‘modus tollens’ the anti-physicalist conclusion as rather showing that prior to leaving her room, Mary didn’t have complete physical knowledge (of human colour vision, in particular), after all.

Second, the physicalist has grounds for resisting P1 as motivated by a mistaken characterization of the physical goings-on—one which is (a) overly representational, (b) overly restricted, and (c) problematically qualitatively etiolated.

To start, recall that the motivation for P1 reflects the supposition that physics does not contain reference to subjective or qualitative phenomena. Hence it was that Nagel claimed that physical ‘theory’ concerns only ‘objective’ goings-on, and that Jackson claimed that qualitative experience is not needed to learn physics.

The first problem with this motivation is that it characterizes the physical basis in overly representational terms. Yes, physicalists look to physics as a guide to

the compositionally basic goings-on, but representation (not to mention theory) is one thing, reality another. Recall the operative conception of the physical goings-on, according to which these are the goings-on that are treated, approximately accurately, by present or (in the limit of inquiry, ideal) physics (see Wilson 2006a). On this and other physics-based conceptions of the physical, physics gives us a handle on the compositionally basic goings-on, but even in the limit of inquiry, there is no supposition that physical theory will be up to the task of offering a complete description of these goings-on.

The second problem with this motivation is that (and even bracketing the previous point) it presupposes an overly restricted characterization of the physical goings-on. The goings-on explicitly referenced in physics provide the starting point, not the end point, of the lower-level physical base. To be sure, the immediate targets of physical theory—e.g., subatomic particles or their wave-functional or field-theoretic correlates—do not individually instantiate or otherwise encode subjective or qualitative aspects of consciousness of the sort we enjoy. But that doesn't show that sufficiently complex physical goings-on do not do so, any more than the fact that the goings-on explicitly referred to in physical theory do not individually instantiate or otherwise encode thermodynamic processes and properties shows that sufficiently complex physical goings-on do not do so.

A third problem might be thought to reflect a somewhat more substantive motivation for P1—namely, that (so the story usually goes) insofar as physics is concerned only with the trajectories and other behaviours of the goings-on that are the immediate target of physical theory, and insofar as such behaviours and associated laws are completely described in structural, non-qualitative terms, no amount of complex or even causal combination could ever eventuate in the instantiation of

physical properties up to the task, so to speak, of either either being or serving as a physical basis for such qualitative aspects of consciousness. Structure, one might say, can only beget more structure.

Now, this is a more substantive motivation for P1, but drawing upon the previous responses, the physicalist can reasonably maintain that it mischaracterizes the physical goings-on, even at low orders of complexity, as completely qualitatively etiolated.

Such an abstract, completely structuralist conception of the compositionally basic entities is not forced by physical theory. On the contrary: it is common for physicists to speak of particles or other lower-level physical phenomena as ‘feeling’ forces, where the reference here might be thought to be both qualitative and sensitive to perspective—the force coming from this direction rather than that, and applied to the object in its location in ways that then enter into the object’s moving in such-and-such a way. This degree of qualitativity, in which a non-complex physical entity is capable of registering a local perspective and associated sensitivity to the environment, in a way that is not entirely qualitatively etiolated, clearly need not be seen as involving any measure of consciousness or mentality, even of a ‘protopsychic’ variety. On the contrary, it reflects a naturalistic point of view according to which physical goings-on are real and substantial, and when they move, they do so not because, e.g., they must play a predetermined part in the harmony of the spheres, but rather for some local salient reason or reasons—i.e., one or more felt forces, or other form of interactive influence. This understanding of non-complex physical goings-on suggests that we should distinguish qualitativity and consciousness: sometimes these go together, as in creatures like us; but qualitativity is a weaker notion, that may be present, e.g., in the non-conscious

response of a particle to the forces acting on it.

Now, from recognition of this amount of qualitativity in the physical goings-on it's still a considerable leap to get to the instantiation of subjective or qualitative aspects of consciousness. But given that some explanatory gaps (e.g., in the case of complex non-linear systems) do not have anti-physicalist import, the burden is on the anti-physicalist to establish that the seeming explanatory gap between physical goings-on and subjective and qualitative aspects of conscious experience must be taken to have such import. The strategy of the knowledge arguments crucially proceeds by way of P1 and the associated characterization of complete physical knowledge, as failing to be knowledge of any subjective or qualitative aspects of consciousness there may be. But the physicalist can and should reject this characterization, both as effectively begging the question against their view (since for the physicalist, complete physical knowledge must include knowledge of those physical features that are identical with or serve as a subjective or qualitative basis for such aspects of consciousness), but also on grounds that there's no independent reason to accept such a characterization.

I conclude that the knowledge arguments do not provide compelling reason to think that consciousness and its associated subjective and qualitative aspects are actually physically unacceptable, nor that any such goings-on are Strongly emergent.

7.1.2 The conceivability argument

I turn now to the conceivability argument advanced and developed by Chalmers (1996, 1999, 2009, and elsewhere), according to which the conceivability of zombies—creatures which are functional and physical duplicates of creatures like

us, but which are lacking in any conscious mentality—is taken, in combination with certain other commitments, to establish the physical unacceptability of consciousness.

Like the knowledge arguments, the conceivability argument relies on the presence of an explanatory gap, as needed to make room for the conceiving in question. As Levine (1983) puts it:

If there is nothing we can determine about C-fiber firing that explains why having one's C-fibers fire has the qualitative character that it does—or, to put it another way, if what it's particularly like to have one's C-fibers fire is not explained, or made intelligible, by understanding the physical or functional properties of C-fiber firings—it immediately becomes imaginable that there be C-fiber firings without the feeling of pain, and *vice versa*. We don't have the corresponding intuition in the case of heat and the motion of molecules—once we get clear about the right way to characterize what we imagine—because whatever there is to explain about heat is explained by its being the motion of molecules. (359)

Chalmers's argument goes beyond the knowledge arguments, however. To start, his argument relies on the positive conceivability of zombies, as opposed to the mere absence of explanation or associated failures of knowledge. That said, an appeal to conceivability alone isn't much of an advance on the previous arguments, since just as physicalists typically do not deny that there are explanatory gaps between consciousness and lower-level physical goings-on, neither need physicalists deny that zombies are conceivable, as long as they are not forced to take such a broadly epistemic fact to have metaphysical import (in particular, of the anti-physicalist variety).

The primary advance of Chalmers's argument rather lies in his situating the conceivability of zombies in an independently motivated framework—‘epistemic

two-dimensionalism', or E2D, to be explicated in more detail shortly—according to which certain facts about meaning, which are supposed to be *a priori* accessible, can be used to identify certain facts about modality, expressing or encoding what is genuinely metaphysically possible (necessary, contingent, impossible). It is commonly assumed that the mode of *a priori* access to meanings—more precisely, the mode of *a priori* access to ‘intensions’: functions from scenarios, or worlds, to extensions—that enters into the E2D strategy proceeds by way of conceiving. Consequently, commitment to the E2D strategy, and to this strategy’s being implemented by means of a conceiving-based epistemology of meanings/intensions, provides an independent basis for taking the conceivability of zombies to have anti-physicalist metaphysical import, as a case-in-point of a systematic connection between conceivability and possibility.

Reflecting the crucial role that E2D plays in his argument, Chalmers has come to call his argument, more accurately, ‘the two-dimensionalist argument against materialism’. The argument proceeds essentially as follows:

1. It is conceivable that there is a world which is physically exactly like our world, but in which there is no consciousness.
2. If the world described in (1) is conceivable, then it is genuinely possible.
(E2D)
3. If the world described in (1) is genuinely possible, then physicalism is false.
4. Therefore, physicalism is false.
5. In particular, consciousness is physically unacceptable (and moreover might be Strongly emergent).

In aiming to provide independent motivation for taking explanatory gaps to have anti-physicalist metaphysical import, Chalmers's argument is different from, and to my mind improves on, the knowledge arguments. Nonetheless, as I'll presently argue, Chalmers's line of thought is also ultimately unsuccessful, and hence fails to motivate taking consciousness to be physically unacceptable, much less Strongly emergent.

The focus of my critical attention is on the second premise.⁶ I'll first present Chalmers's two-dimensionalist argument in more detail, highlighting the attractiveness of the E2D strategy that lies at the argument's core. Next, drawing on Biggs and Wilson (2017), I'll suggest that there is an alternative, and superior, way in which the E2D strategy might be implemented—namely, by appeal to an abduction-based rather a conceiving-based epistemology of the meanings (intensions) entering into this strategy. I'll then argue that it is far from clear that the genuine possibility of zombies, or the associated Strong emergence of consciousness, is output from E2D, when this framework is implemented using abduction rather than conceiving. One might wonder, as against this line of thought, whether abduction would be suited for purposes of implementing E2D, given that

⁶Rejecting the first premise seems likely to lead directly to stalemate, and rejecting the third premise is ultimately indecisive. In re the latter: one might deny that the genuine possibility of a zombie world suffices to establish the falsity of physicalism, on grounds that the premise presupposes that any relation between conscious states and lower-level physical states compatible with physicalism must be one according to which the latter metaphysically necessitate the former; but this assumption is incorrect. Recall, e.g., that it would suffice for the physical acceptability of conscious states that the latter be functionally realized by physical states, in such a way that the powers of the former are, on any given occasion, a proper subset of those of the latter; but if powers are only contingently associated with features, then the physical states that realize conscious states in worlds with our laws of nature might not do so in worlds with different laws of nature—compatible with, in particular, the existence of zombie worlds. This response to Chalmers's two-dimensional argument is ultimately indecisive, however, since he might maintain that the argument applies even when the holding of the actual physical laws is built into the specification of worlds in which the lower-level physical base goings-on are present.

(as above) the access to the meanings which are in turn supposed to provide a basis for access to modal truths is supposed to proceed in a priori fashion. Here again, I draw on joint work with Biggs (Biggs and Wilson 2016*b*), where we argue that, contra common assumption, abduction is an a priori mode of inference—as a priori as conceiving, in particular. I conclude that, like the knowledge arguments, Chalmers’s two-dimensional argument fails to establish that consciousness is actually physically unacceptable, much less Strongly emergent.

Epistemic two-dimensionalism

What is missing from the previous knowledge arguments is independent good reason to take the explanatory gaps that they highlight to have anti-physicalist metaphysical import. Chalmers’s intended route to such good reason takes as its starting point Kripke’s (1972/1980) undermining of the traditional supposition that modal truths are (always) a priori accessible. Kripke argues, more specifically, that many necessary truths about individuals and natural kinds can be known only a posteriori. Some such truths—e.g., those expressed by ‘Hesperus is Phosphorus’ and ‘water is H₂O’—involve identity, to which reductive physicalists appeal as holding between goings-on at seemingly different levels of the scientific hierarchy; and one might suppose that realization relations of the sort to which non-reductive physicalists appeal might be similarly accessible only a posteriori. As such, Kripke’s results might be thought to undercut any hope that a priori investigations might shed light on whether or not some goings-on are physically acceptable, and more generally to undercut the prospects of our having much, if any, modal knowledge about broadly scientific goings-on.

These consequences would be undesirable, not just for purposes of assessing

the truth of physicalism, but insofar as much theory and practice (philosophical, legal, semantic, scientific, etc.) presupposes that we can know modal truths independently of, or at least prior to the end of, empirical inquiry. We might thus hope that, notwithstanding Kripke's results, the traditional link between necessity and a priority could be restored, to at least some considerable extent.

This is the promise of epistemic two-dimensional semantics (E2D), advocated by Chalmers (1996, 2006*b*, 2009), Chalmers and Jackson (2001), and others.⁷ E2D can be heuristically seen as refining and generalizing Frege's suggestion that there are two kinds of meaning: sense and reference. On the Fregean picture, sense represents an aspect of meaning that is *a priori* accessible to a competent speaker, whereas reference represents an aspect of meaning that may fail to be so accessible; hence, for example, Frege plausibly supposed that a competent speaker could know *a priori* the senses of 'Hesperus' and 'Phosphorus' as 'the first star seen in the evening' and 'the first star seen in the morning', respectively, while not knowing the empirical fact that these expressions have the same referent (namely, the planet Venus).

E2D similarly aims to characterize two distinct aspects of meaning, each represented as intensions: functions from worlds or scenarios to extensions. First are 'primary intensions', corresponding to an aspect of meaning that, like Fregean sense, can be accessed from the armchair, i.e., can be known *a priori*. Second are 'secondary intensions', corresponding to an aspect of meaning that, like Fregean reference, can in at least some cases be accessed only through additional experience, i.e., can be known only *a posteriori*, reflecting the dependence of such intensions (such referential aspects of meaning) on how things are at whatever

⁷See, e.g., Peacocke 1993, Chalmers 1996, Boghossian 1996, and Gertler 2002.

world is in fact actual, and which may enter (as per Kripke's results) into certain a posteriori necessary truths. The promise of E2D ultimately lies in the suggestion that primary and secondary intensions are connected in such a way as to provide a basis for a priori knowledge of a wide range of modal truths, including a conditional such basis for the truths at issue in Kripke's discussion. For example, while it is a posteriori that, e.g., water is necessarily H₂O, our access to appropriate intensions, on the E2D strategy, provides a basis for our knowing a priori that *if* water (or the watery stuff) is actually H₂O, *then* water is necessarily H₂O.⁸ Here and in related cases, the a posteriori contribution to the necessity claim at issue is limited to discharging the antecedent of a conditional that is known a priori—again, compatible with our having a priori access to a great deal of modal knowledge. There is more one might say here about the technical details of implementing E2D, but for present purposes this sketch suffices (see [Biggs and Wilson 2017](#) for more detailed discussion).

Granting that we have independent reason to accept the E2D framework and associated strategy for reforging the link between a priority and modality, the question remains of which epistemology of intensions should be taken to be operative in implementing this strategy.

Now, as prefigured above, it has commonly been taken for granted that the epistemology of intensions—the account of how we can or should go about determining what extensions our expressions have, in a given world or scenario—must be one relying on conceiving. Hence [Gertler \(2006\)](#) says:

Conceivability is the only guide to necessity; our concepts, and the intuitions about possibility that derive from them, provide our only grip

⁸Similarly, we can know a priori that *if* water (or the watery stuff) is actually XYZ, *then* water is necessarily XYZ.

on modal claims. [...] [I]t's worth mentioning that modal intuitions—intuitions about what is possible and impossible, which it is the aim of conceivability arguments to reveal—are as important to arguments for reductionism as they are to anti-reductionist claims. Again, reductionism entails that it's impossible for the reduced property to vary independently of the reducing property. Since a claim of impossibility cannot be established by considering the actual world alone (though of course it can be refuted in this way), the reductionist must consider whether certain non-actual scenarios are possible. And the only way to determine this is to use the method of conceivability. (205)

Given a conceiving-based epistemology of intensions (CEI), we are close to the anti-physicalist conclusion of Chalmers's conceivability argument. E2D implies that if ‘zombie’ has a positive extension at some world, then zombies are possible. CEI implies that ‘zombie’ has a positive extension at some world if one can conceive of a zombie world. So, E2D coupled with CEI implies that, if one can conceive of a zombie world, then zombies are possible. But as per the first premise of Chalmers’s argument (which I am granting) one can conceive of a zombie world. So zombies are possible, which given the third premise of the two-dimensional argument against materialism/physicalism (which again I am granting for the sake of argument) implies that consciousness is not physical, and moreover might be Strongly emergent.

This is an interesting line of thought, but as I'll now argue, we have both negative and positive reasons to reject it.

Conceiving-based vs. abduction-based epistemologies of intensions

Let's start by asking: why think that the only epistemic access to the extensions of our concepts or terms in other scenarios is via conceivability, as per CEI? Why not, in particular, allow that such identification might proceed via inference to the

best explanation, or abduction? Why not allow that various theoretical desiderata or abductive principles—ontological and ideological parsimony, plausibility, explanatory fruitfulness, compatibility with other beliefs, unifying power, ability to resolve certain problematics, and the like—can come into play in determining the extensions of certain concepts or terms in non-actual scenarios? After all, on the face of it such considerations might well be relevant.

Plausibly, abduction has not been taken seriously as a basis for implementing the E2D strategy on grounds that, if the strategy is to do the job of reforging an *a priori* accessible link between meaning and modality, the operative epistemology of intensions must be *a priori* (involve an *a priori* mode of inference), and while conceiving is *a priori*, abduction is not. As Biggs and I argue in our (2016b), however, abduction *is* *a priori*. Considerations of space prevent my rehearsing these arguments here.⁹ By way of heuristic motivation I note three points.

First, a common characterization of what it would be for some claim to be known *a priori* adverts to such knowledge being available ‘from the armchair’; but abductive deliberation can and does proceed from the armchair, via consideration of how best to maximize satisfaction of the usual abductive principles (ontological and ideological parsimony, plausibility, fruitfulness, etc.).

Second, for purposes of implementing E2D what is crucial is that we can have *a priori* access to conditional claims of the form, e.g., ‘if the actual watery stuff is H₂O, then it is necessary that water is H₂O’, and so on. As it happens, abduction is often applied, in scientific contexts, to empirically gained data; but nothing prevents the possibility of abductive deliberation being directed at hypothetical rather than actual goings-on. Hawthorne (2002, 252) makes a similar point, sug-

⁹Note to readers: should I put a summary of this argumentation into an appendix to the chapter?

gesting that abduction can deliver a priori justification for belief in a conditional whose antecedent describes an ‘experiential life history’ and whose consequent is whichever theory best explains some aspect of that life history; see also Cohen 2010, 152–153 and Wedgwood 2013). On the face of it, then, there is no barrier, in particular, to using abduction as a means of gaining a priori knowledge about extensions in hypothetical scenarios or worlds, just as conceiving is supposed to allow one to do.

Third, perhaps the main objection to taking abduction to be an a priori mode of inference proceeds via the supposition that the epistemic value of abductive principles, hence the epistemic value of applications of abduction in line with these principles, contingently depends on whether, e.g., the world in which abduction is being applied (actually or hypothetically) is one where, e.g., ontologically more parsimonious theories are likely to be true. Such a line of thought fails, however, since neglecting that abductive principles are always operative *ceteris paribus*, or other things being equal. Once this qualification is taken into account, scenarios which are sometimes presented as showing that adherence to an abductive principle can, for contingent reasons, lead one epistemically astray (if, e.g., it turns out that there are more kinds in a world than a theory motivated by conformity to the principle of parsimony countenances) are in fact scenarios where the *ceteris paribus* condition was violated, with the upshot being that, perhaps surprisingly, the epistemic value of abductive principles and their application is not subject to empirical disconfirmation. There’s a great deal more to say here, which Biggs and I can and do say elsewhere (besides our 2016b, see also our 2016a, 2017, and in progress). I direct the interested reader to these discussions, and henceforth assume, what I take to have elsewhere established, that abduction is a priori—as a

priori, in particular, as conceiving.

I next want to highlight the primary independent reason for thinking that, given that abduction is a priori, we should look to abduction rather than conceiving as a guide to the intensions at issue in the E2D strategy.¹⁰

This reason reflects what might be called ‘access-based’ objections to E2D, according to which the common presence of indeterminacy or inconsistency in our concepts indicates that we commonly fail to have a priori access to the intensions of natural kind expressions (among others), to such an extent that the prospects of an E2D-led reforging of the link between meaning and modality are fatally undermined. As Biggs and I argue in our (2017), however, these access-based objections are not to E2D *per se*, but rather to E2D when implemented using CEI; for while conceiving alone does not have the resources to resolve indeterminacy or inconsistency in our concepts, abduction does have such resources. More generally, Biggs and I argue that when E2D is combined with AEI rather than CEI, the full range of access-based objections can be satisfactorily answered. In what follows, I discuss as an illustrative case-in-point one salient access-based objection, rehearsing our reasons for thinking that while a conceiving-based epistemology of intensions alone cannot overcome the objection, an abduction-based epistemology of intensions can do so. Afterwards I return to Chalmers’s conceivability argument, and the question of whether an implementation of E2D provides reason to think that zombies are genuinely metaphysically possible.

According to one pressing access-based objection, widespread conceptual indeterminacy—better: underdetermination¹¹—undermines the prospects for E2D. The line of

¹⁰For a detailed and more general comparison between abductive vs. conceiving-based epistemologies of modality, see [Biggs and Wilson in progress](#).

¹¹Though the objection and responses have been pitched as involving conceptual ‘indeterminacy’, ‘underdetermination’ is a better term for the phenomenon at issue, since ‘indeterminacy’

thought here is that implementing the E2D strategy requires that extensions of expressions (e.g., natural kind terms) of the sort about which we hope to regain modal knowledge requires that these expressions have a priori-accessible extensions in every possible scenario; but in cases of conceptual underdetermination the associated expressions lack antecedent extensions in some scenarios. Moreover, the concern runs, such underdetermination is widespread. That there is widespread conceptual underdetermination is suggested by Mark Wilson's (1982 and 2006b) arguments to the effect that applications of natural kind predicates can and frequently do depend on arbitrary factors. In one of Wilson's toy cases, whether members of an isolated tribe take airplanes to be in the extension of their predicate 'bird' (or linguistic correlate) depends on whether the first airplane they encounter is overhead (in which case, airplanes are judged to be birds) or is rather on the ground (in which case, airplanes are judged not to be birds). If such historical accident partly determines whether 'bird' applies to airplanes, then, Wilson reasonably assumes, the full range of the predicate's extension was not antecedently determined. And, Wilson suggests, such underdetermination is a feature of natural kind predicates more generally.

Chalmers grants that there may be underdetermination of the sort at issue in Wilson-style cases. In his (1996), he says "There may of course be borderline cases in which it's indeterminate whether a concept would refer to a certain object if a given world turned out to be actual", and more recently, Chalmers grants that it is plausible that "later extensions [of the kind of expressions at issue in Wilson's cases] depend on idiosyncratic developments, and verdicts about such cases are not determinately prefigured in a user's original use of an expression" (Chalmers

has connotations associated more specifically with vagueness or metaphysical indeterminacy, of the sort discussed in the last chapter.

2012, 231). Chalmers maintains, however, that such underdetermination is “no problem” for E2D, as in the continuation of the passage from his 1996 discussion:

There may of course be borderline cases in which it’s indeterminate whether a concept would refer to a certain object if a given world turned out to be actual. This is no problem: we can allow indeterminacies in a primary intension, as we sometimes allow indeterminacies in reference in our own world (364).

But while Chalmers is right that E2D can tolerate some conceptual underdetermination, it cannot tolerate widespread underdetermination, on pain of undermining (as Biggs and I previously put it) E2D’s *raison d’être* of providing a basis for our a priori knowledge of a wide range of modal truths, including conditional such knowledge of necessary a posteriori truths (such that, e.g., we can know a priori that *if* the actual world is one where water, or the watery stuff, is H₂O, then it is necessary that water is H₂O). As such, widespread conceptual underdetermination would undermine E2D—at least if such underdetermination is insuperable.

Might Chalmers maintain either that conceptual underdetermination is not really widespread, or that in any case such underdetermination is superable using the resources of conceiving? Neither strategy is promising.

First, conceptual underdetermination is widespread, for both natural kind and other predicates (more generally, expressions). Besides the Wilson-style cases and arguments for this conclusion, Biggs and I note two further sources of conceptual underdetermination. The first concerns cases of vague predicates of the sort associated with Sorites sequences (e.g., ‘red’, ‘bald’, ‘rich’, ‘cell’, ‘part’, etc.). As Raffman (1994) compellingly argues, the application of such predicates can be and often is typically determined in a given case by arbitrary contextual and/or psychological factors. For example, the breaking point in where one stops

(or starts) applying the predicate ‘red’ in a given colour-spectrum sequence may non-systematically depend on where along the spectrum one starts (as per the phenomenon of hysteresis) as well as on arbitrary psychological factors (fatigue, boredom, etc.). For Wilson-style reasons, the dependence of extensions on arbitrary historical and other factors supports thinking that the associated expressions do not come with antecedently fixed extensions—i.e., are to some extent conceptually underdetermined. But those discussing the phenomenon of vagueness often register its being likely that *most* expressions in ordinary language (including scientific language) are vague.

A second source of widespread conceptual underdetermination reflects the historical scientific record, which supplies many cases of underdetermination involving natural kind terms. For example, ‘acid’ initially was taken to refer to only oxygenated substances, but was later applied to HCL, for theoretical reasons now largely discarded; dispute remains over whether Newtonian uses of ‘mass’ apply in relativistic contexts; the decision to classify whales as mammals was a controversial affair; there’s the infamous resolution declassifying Pluto as a planet (about which there is ongoing dispute). In these and many related cases, it’s worth noting, the antecedent underdetermination has been resolved (when it has been resolved) not through accident, but rather, through non-demonstrative reasoning of one sort or another—typically, broadly abductive reasoning.

It seems clear, then, that conceptual underdetermination is widespread. Hence such underdetermination will be ‘no problem’ for E2D only if the operative epistemology of meanings/intensions is capable of resolving the underdetermination. Now, Chalmers (1996) does consider whether conceivers might be able to eliminate underdetermination by foreseeing relevant accidents. In Wilson’s toy case,

for example, Chalmers says that one “might try to classify these two different scenarios [airplane first seen in the sky or on the ground, respectively] as different ways for the actual world to turn out, and therefore retain a fixed, detailed primary intension” (364). On this strategy, the fully determinate primary intension of ‘bird’ includes planes in its extension if the tribe members first see a plane overhead, but not if they first see it on the ground. Either way, according to the suggestion, the underdetermination in the tribe’s expression ‘bird’ is resolved. And Chalmers might attempt to implement a similar strategy as a means of overcoming conceptual underdetermination involving vague or scientific predicates or other expressions.

Such a strategy has traction against the underdetermination objection only if conceivers can foresee how intensions are sensitive to accidental or arbitrary factors. But as Biggs and I see it, a deeper lesson of Wilson’s case, as well as of Raffman’s discussion, is that the influence of such factors cannot be foreseen, at least not by conceiving alone. Determinism and such aside, those using conceiving alone might apply ‘bird’ differently even relative to the same ‘accident’. After all, there are any number of respects of dissimilarity between airplanes and birds, even when the former are in flight, and a minor difference in attention to these features might result in a different decision about whether ‘bird’ applies to a flying airplane. And as Raffman points out, the factors entering into the application of a vague predicate are not just arbitrary but are unsystematically so, as is reflected in different applications of vague expressions even against relevantly the same background conditions. We can register, post-hoc, extensions resulting from whatever decision was in fact made, but in cases of arbitrary determination there is no way to antecedently identify these extensions and corresponding intensions

through conceiving alone—there is simply no fixed extension to ‘conceive’, even taking relevant circumstances into account.

Similarly, there is no case to be made that even an idealized conceiver could determine the extension, in every scenario, of natural kind expressions pertaining to ‘mass’, ‘planet’, and so on. Again, as a matter of historical fact the considered extensions of these terms is heavily informed by abductive considerations, taking into account ideological and ontological parsimony, compatibility with existing beliefs or theories, systematicity, unification, fruitfulness, etc. But as Chalmers and others implementing E2D via a conceiving-based epistemology emphasize, conceiving does not involve these sorts of abductive or ampliative resources.¹² Correspondingly, conceiving alone is simply not up to the task of overcoming conceptual underdetermination in these cases.

By way of contrast, an abduction-based epistemology of intensions is up to the task of overcoming conceptual underdetermination, in the cases of scientific expressions and more generally. For, when considering whether to apply an expression in a given scenario, abductors can consider not only historical accident and psychological variability, but also any non-demonstrative rational grounds that might push towards one extension rather than another. Hence abduction, unlike conceiving, can be productive (it is ampliative, after all), allowing those who are identifying intensions to consider how the concept *should* be applied, given the usual abductive principles and associated theoeretical desiderata, even when

¹²For example, Chalmers and Jackson (2001) are explicit (when arguing, in particular, against Block and Stalnakers 1999 claim that the justification for the conditional claims output by E2D might rely on broadly abductive considerations) that in their view abductive considerations do not play any ‘essential justificatory’ role in the cases at hand; see especially pp. 342–350). See Biggs and Wilson 2017, §3.1 for discussion of positive and negative understandings of conceiving as entering into E2D.

the application conditions are antecedently to some extent underdetermined. Since abduction can, in a rational way, go beyond what expressions antecedently encode, AEI has the potential to overcome conceptual underdetermination, extending applications of natural kind and other predicates and expressions to new scenarios or situations, on ultimately rational grounds. For example, an idealized version of a competent user of ‘bird’ who has never before seen an airplane would plausibly be in position to consider and compare theories of the intensions of ‘bird’ by attention to which theory would be, among other desiderata, most ontologically and ideologically parsimonious, unifying (e.g., of experience of flying entities), plausible, explanatorily fruitful, and so on. Similarly, for the cases of ‘mass’, ‘Pluto’, ‘water’, and so on. Here it is worth noting, by the way, that there is no barrier to an idealized abductor engaging in such extensions of the terms/concepts at issue, for this is more or less what language-users do as a matter of course.

Similar results attach to other access-based objections.¹³ Even focusing just on underdetermination, however, the upshot is clear: E2D when implemented using a conceiving-based epistemology of intensions faces difficulties, to which an abduction-based epistemology of intensions, in having ampliative resources of the sort that conceiving fails to have, is in position to respond.

¹³One of these is Melnyk’s (2008) objection that on the sort of internalist account of concept possession that is most promising for implementing the E2D strategy, competent language users can and typically do have at least some incorrect dispositions to apply such expressions, both due to fallibility and to distance from the end of inquiry, in such a way as to lead to inconsistency in applying the application. Again, as Biggs and I argue in our 2017, abduction, but not conceiving, has resources enabling correct from incorrect dispositions to be sorted.

Revisiting the possibility of zombies

This result in hand, I now want to return to the status of Chalmers's two-dimensionalist argument against materialism, and in particular, to the question of whether the independently desirable E2D framework supports taking zombies to be metaphysically possible. Effectively, the question is: are there worlds that are physical and functional duplicates of our own (including, let us grant, where the physical laws are the same as our own) in which the intension of 'zombie' has a non-null extension?

We can start by observing that on an abductive epistemology of intensions, an attempt to answer this question will proceed by attention to any relevant considerations, which may include but are not restricted to any modal intuitions we might have. The usual theoretical desiderata—parsimony, plausibility, fruitfulness, systematicity, and so on—will be relevant, as will be considerations such as the desirability of properly accommodating the causal efficacy of consciousness, and (perhaps relatedly) the desirability of conforming to what our best sciences give us reason to believe. Accordingly, different 'theories' of the intensions at issue will rate better or worse, depending on how well these do at accommodating or satisfying these diverse considerations. A theory on which the extension of 'zombie' has a non-null extension scores positively, in accommodating the seeming conceivability of zombies. On the other hand, insofar as such a non-null extension implies that consciousness is Strongly emergent or otherwise physically unacceptable, the associated theory of the intension of 'zombie' is less ontologically parsimonious than one on which the extension of 'zombie' is null at the relevant worlds. The former theory is also less systematic than the latter, so far as accommodating the possibility that conscious states can have physical effects is concerned, for it is

unclear how non-physical goings-on can interact with physical goings-on (as was originally pressed by Princess Elizabeth of Bohemia against Descartes; see [Kim 2015](#) for recent discussion), and though my own view is that this can be made sense of via the notion of a fundamental mental interaction, in any case such an account requires further ontology and ideology. Alternatively, one might couple a theory on which ‘zombie’ has a non-null extension with the denial of the assumption that conscious states can cause physical effects, as Chalmers sometimes seems to suggest; but since such a denial is highly implausible, given our experience, and moreover unsystematic, given that no other empirical goings-on are epiphenomenal, this theory too suffers by way of comparison with one taking ‘zombie’ to have a null primary intension.

The final abductive comparison of theories of the intension(s) of ‘zombie’ will involve not just these, but many other considerations. Even without entering further into this issue, however, it is clear that an abductive rather than (merely) conceiving-based approach to the question at issue is very far from indicating that ‘zombie’ has a non-null extension—if anything, the all-things-considered weight seems likely to be on the other side. To be sure, a positive empirical result, indicating the presence of a novel fundamental interaction, might trump all these other considerations, but at present we are lacking such evidence. Hence even granting both the independent desirability of the E2D framework and the conceivability of zombies, nothing yet follows about whether zombies are metaphysically possible. Chalmers’s conceivability argument, like the knowledge argument(s), thus fails to establish that consciousness is physically unacceptable, nor that it is, more specifically, Strongly emergent.

7.2 Is consciousness Weakly emergent?

Let us suppose that in the fullness of empirical inquiry, it is established that consciousness is not Strongly emergent: no fundamental novelty is operative at the level of complexity associated with any conscious mental states. In that case, is there presently good reason to think that any conscious states are Weakly emergent from, as opposed to ontologically reducible to, the lower-level physical states upon which they depend?

As prefigured above, my strategy here starts by attention to the fact that qualitative conscious features—e.g., states of conscious awareness of colours or pains—are reasonably taken to be typically determinable rather than (maximally) determinate, and by going on to argue that (again, on the assumption that conscious states are not actually Strongly emergent) such determinable conscious states are plausibly seen as actually realized by determinate lower-level physical states, in accord with a determinable-based account of non-reductive realization—that is, Weak emergence. As we'll see, the latter claim faces certain challenges, which I will aim to address.

7.2.1 Determinable perceptions

Among the most salient conscious mental states are those which we might call ‘qualitative’, in constitutively involving experience of features such as colours, tastes, textures, pains, and the like. There are (at least) two motivations for taking the qualitative aspects of conscious experiences to be determinable rather than maximally determinate, rendering the states themselves appropriately seen as determinable rather than maximally determinate.

The first motivation appeals to Sorites phenomena, as indicating that we fail to perceive fully determinate instances of many properties, including colours, tones, and textures. Recall that Sorites sequences consist in instances of such features which are pairwise indiscriminable. The notorious Sorites paradox or paradoxes consist in observing that, if one starts with an instance that is (judged to be), e.g., clearly red, and given that the next patch is indiscriminable from the first, the next patch must also be (judged to be) red; similarly for the second and third patches (if the second patch is red, then since the third is indiscriminable from the second, the third must also be red); and on and on, in what is sometimes called a ‘forced Sorites march’, until one is obliged, contradictorily, to (judge) red a patch that is clearly not red—e.g., a patch that is clearly orange. As Fales (1990) notes, that the qualitative aspects of perceptual experiences are to some extent determinable rather than maximally determinate “is a conclusion which seems forced upon us by the fact that each member of a series of colors, etc., may be perceptually indistinguishable from its immediate neighbors but easily distinguishable from more distant members of the series” (172).

The second motivation reflects that our perception of macro-objects and their features typically fails to register microdeterminate details. As mentioned this in Ch. 1, when discussing the *prima facie* motivations for there being metaphysical emergence, under the rubric of ‘perceptual unity’:

Though the macro-entities of our acquaintance are, the scientists tell us, materially constituted by massively complex and constantly changing micro-configurations (ultimately involving whatever fundamental physical goings-on there might be), macro-entities do not perceptually appear to us as massively complex, constantly changing, configurations of micro-phenomena. A tree, for example, does not look like a complicated structure of cells or tissues, much less like a buzzing array of sub-atomic particles or other physical fundamenta; rather, a

tree looks like a comparatively unified entity, both at and over time; and the same is true for other familiar macro-entities.

These points hold true if qualitative features of macro-objects (colour, shape, texture, etc.) are at issue, rather than the objects themselves. Again, we do not experience the shapes of macro-objects in fully microscopically determinate detail; rather we experience these shapes as to some extent determinable.

Two further points about these broadly perceptual motivations are worth noting. First, the motivations here do not presuppose or depend on the claim that the objects that are perceived themselves fail to be determinate in the relevant respects. My own view, as per the previous chapter, is that certain (many? most?) features of ordinary objects (including but not limited to their boundaries) are in fact metaphysically indeterminate, which when combined with my view of metaphysical indeterminacy does entail that at least some features of objects that are perceived are determinable (though whether these are the same features as those which are experienced as determinable in perception is a further matter). But even if the objects perceived are themselves completely determinate in all (or the relevant) respects, it would remain that qualitative conscious states involving perceptual experience of such objects would still be less than maximally determinate. As I note in my (2012), in discussing how perception provides motivation for thinking that determinables exist,

Perhaps the (instances of) properties perceived are really maximally determinate, and only perceptual features or modes of presentation are determinable; but features of perceptual experience are also aspects of reality, so the larger point remains. (5)

It seems reasonable to suppose, then, that at least some qualitative conscious states are less than maximally determinate, due to these states' having qualitative

aspects that are less than maximally determinate.

Second, the phenomena at issue suggest that there is no hope here of providing a deflationary—anti-realist or reductionist—treatment of the determinable qualitative conscious states at issue. That we have qualitative conscious states (qualitative experience) isn't, in my view, up for grabs—it's as or more epistemically foundational in our experience as any part of reality; hence any argument for eliminating such states would be less compelling than the sheer fact of our experience.¹⁴ Nor is it plausible to suppose that these determinable states might be given some or other reductive treatment. There are a number of routes towards the rejection of any such reduction (see, e.g., Wilson 2012), but for present purposes it suffices to observe that any reductive treatment of a determinable state, whether this (implausibly) involved an identification of one determinable type with a single determinate type, or (more plausibly) the identification of one determinable type with a disjunction or some other complex of determinate types, would entail that every token of a determinable type was identical with a token of a determinate type—and indeed, a token of a maximally determinate type. But the phenomena associated with qualitative conscious states undercuts the hope of any such reduction, since as per the motivations above, our token qualitative conscious experiences are not maximally determinate. It is thus moreover reasonable to assume that at least some conscious experiences, understood as having constitutive qualitative aspects, are characterized in irreducibly determinable terms.

Now, as previously, one account of realization of the sort plausibly satisfying the conditions in the schema for Weak emergence is a determinable-based account, according to which it suffices for the non-reductive realization of a feature

¹⁴This is a common, and to my mind compelling, reason to reject the sort of eliminativism about mental states endorsed by the Churchlands (in, e.g., Churchland 1986 and Churchland 1981).

(property, state, etc.) that the feature be a determinable of lower-level physical determinates. So, if the qualitative conscious states at issue can be seen as determinables of lower-level physical determinates, we will be in position to conclude that such conscious features are (at least) Weakly emergent.

7.2.2 The objections from mental multiple realizability and mental super=determinates

Seeing qualitative conscious states as Weakly emergent, by lights of a determinable-based account, requires that it make sense to see lower-level physical states as determinates of determinable conscious mental states. [Ehring \(1996\)](#), [Funkhouser \(2006\)](#), and [Walter \(2006\)](#) argue, however, that this does not make sense.¹⁵ The common line of argument in these discussions is along the following lines:

1. The determinable/determinate relation has feature F
2. The relation between qualitative conscious states and lower-level physical states does not have F
3. Therefore, the relation between qualitative conscious states and lower-level physical states is not the determinable-determinate relation.
4. Qualitative conscious states are not realized, in determinable-based fashion, in lower-level physical states.

¹⁵[Worley \(1997\)](#) and [Funkhouser \(2006\)](#) also argue that certain mental states are not appropriately seen as determinables of lower-level physical determinates; but since these discussions focus on beliefs, they are not directly relevant to the present question of whether qualitative conscious states can stand in a determinable/determinate relation to lower-level physical states. See [Wilson 2017](#) for discussion of Worley's and Funkhouser's concerns.

This line of thought, if successful, leaves open that there is some other route to the actual Weak emergence of consciousness, but would undercut what to my mind is the most promising—i.e., determinable-based—route to that conclusion. My strategy here will thus be to defend a determinable-based treatment of qualitative conscious states against Ehring's and Walter's objections; here I draw upon the discussion in my (2009).

There are two instantiations of the above argument schema relevant to the question of whether qualitative conscious states might have lower-level physical determinates.

The first appeals to the feature of the determinable-determinate relation according to which determinates of a determinable differ ‘in respect of’ their determinable. This feature reflects that the distinctive form of specification whereby a determinate is more specific than a given determinable is supposed to contrast with the genus-species relation as well as, more generally, the conjunct-conjunction relation (see Wilson 2017 for historical discussion). The latter specification relations are compatible with the increase in specificity involving the conjunctive addition of some independent property, as in the case of an account of the species *human* as involving the genus *animal* conjoined with the differentia *rational*, or in the relation between a state of affairs of something’s being round and the conjunctively specified state of affairs of that thing’s being red and round. The specification at issue in paradigm cases of the determinable/determinate relation is not properly understood in such conjunctive terms; for example, *scarlet* is not appropriately analyzed as a conjunctive combination of *red* and some other property; rather, determinates are in some sense more specific ‘in respect of’ their determinables. One way in which this more intimate variety of specification gets elucidated is in terms

of the determinable's having certain 'determination dimensions', along which the determinable can be rendered more determinate (as per [Funkhouser 2006](#)). For example, it is common to suppose that the determination dimensions of *colour* are hue, saturation, and brightness, with different determinates of *colour* differing in respect of colour by differing in respect of how they are specified along one or more of these determination dimensions.

Now, according to the argument from mental multiple realizability, that conscious mental states are typically multiply realizable rules out that lower-level physical states differ in respect of any such conscious mental state, hence rules out taking conscious mental states to be determinables of lower-level physical determinates:

[T]he physical realizers of the mental will not differ mentally at all, as they should if they are determinates of the requisite mental states.
([Ehring 1996](#), 474)

Mental properties are said to be multiply realizable precisely because distinct physical realizers can be exactly the same with respect to the mental property they realize, ([Walter 2006](#), 219)

Note that the problem being raised here is not that we can't make sense of determinates of a given conscious state as being exactly the same with respect to the mental property they realize—there's no problem with this (perhaps, e.g., the multiple determinates are exactly similar in sharing the determinable or its powers, or in their essences containing the determinable essence, as Yablo suggests in his [1992](#)). Rather, the concern is that there is no clear way to see qualitative conscious mental states as having determination dimensions that can be further specified by their multiple lower-level physical realizers. Hence, for example, if the determination dimensions of perceived colours are hue, saturation, and brightness, there

is no clear way of understanding how perceived values of these features could be further specified by a lower-level physical realizers: it's not as if the lower-level physical realizers are capable of more precise colour perception! And similarly for other qualitative conscious states, such as pain.

The second instantiation of the above argument schema relies on the feature of the determinable/determinate relation according to which some determinables admit of maximal specification; in such cases, there is a maximal determinate or ‘super-determinate’ of the determinable. According to the argument from mental super-determinates advanced by Ehring (1996), some qualitative conscious states are super-determinate. As Ehring puts it:

[S]uppose that M is a fully determinate type of mental state. For example, make M a precise state of searing pain such that there is no room for further specification of this mental state *qua* pain state. [...] Suppose that M and “being in pain” have a physical [realizer], P . Suppose that P is a determinate of M and “being in pain”. [If M is realized by P in determinable-based fashion] M cannot be a fully determinate pain state. This is so because there are further determinates of that pain state [...] in the form of P . But in fact, M is a fully determinate pain state by hypothesis. Thus the physical [realizers] of mental properties are not determinates of that which they realize since if that were true, M would not be a fully determinate pain property.
(473)

Here the concern is clear—namely, that taking qualitative conscious features to be realized in determinable-based fashion by lower-level physical features would falsely imply that certain qualitative mental super-determinates could be further determined.

7.2.3 Responses to the objections

These concerns are each important, but as I will now argue, they can be addressed, given a proper understanding of the determinable-determinate relation.

A Powers-based account of determination

Determinables (of whatever ontological category) are less specific than their determinates. In [Wilson 1999](#), I argued that this increase in specificity reflects a proper subset relation between the sets of powers of the types of features involved, as follows:

Powers-based determination (first pass): Feature P is a determinate of feature Q iff the set of powers associated with Q is a proper subset of the set associated with P .

Here the idea is that a determinate is more specific than its determinable in being associated with a more specific set of powers. Hence it is, e.g., that in virtue of being scarlet, a patch can do more (e.g., get Alice the picky pigeon to peck at it) than the patch can do simply in virtue of being red.

This first-pass proposal has the virtue of ensuring that there is a contrast between the determinable/determinate relation and the disjunction/disjunct relation (another specification relation with which determination is traditionally taken to contrast), for as previously discussed, disjuncts are associated with more powers than associated disjunctions. As it stands, however, the first-pass proposal does not ensure the contrast between the determinable/determinate relation and the genus/species and conjunct/conjunction relations. Again, it is crucial to the determinable/determinate relation that the determinate cannot be understood or analyzed as a conjunction of the determinable and some other property; in Karen

Bennett's memorable terms, the relation between determinables and determinates is not properly understood along lines of a 'cupcake model', on which determinates are determinables with frosting on top.

We may preserve this contrast on a powers-based approach (improving on the first-pass proposal) by stipulating that the powers in the complement of the sets associated with a determinate and any of its determinates, respectively, are not associated with a set that is associated with any property, as per:

Powers-based determination (second pass): feature P is a determinate of feature Q iff Q is associated with a proper subset of the powers associated with P , and the set of powers had by P but not by Q is not associated with any property.

The second pass account characterizes a specification relation that appropriately contrasts with both the disjunction/disjunct and conjunct/conjunction specification relations. Henceforth, I'll just refer to this as 'Powers-based determination'.

Addressing the objection from multiple mental realizability

Recall that the deeper concern at issue in the objection from multiple mental realizability reflects, first, a conception of the intimate ('in respect of') form of specification associated with the determinable/determinate relation as involving increased specification along one or more determination dimensions of the determinable; and second, the supposition that diverse physical realizers of a qualitative conscious state cannot differ from each other along mental determination dimensions of the conscious states at issue, contra the 'in respect of' feature of the determinable/determinate relation.

My response proceeds by providing empirical reason to think that what determination dimensions are associated with a given determinable conscious state

(perceived colour, pain) is science-relative, in such a way that we can make sense of such states as having purely ‘psychological’ determination dimensions relative to certain sciences (e.g., ‘normal’ colour science or some branch of psychology), and as having physical determination dimensions relative to other sciences (e.g., metameric colour science or pharmacology). Such relativization makes in-principle room for multiple physical realizers of a qualitative conscious state to have explicitly physical as well as psychological determination dimensions. I then use Powers-based determination to fill in how this might be, from a metaphysical point of view.

Take perceived colour, for example. Hue, saturation, and brightness suffice to characterize perceived colours in normal light conditions, to normally sighted creatures more or less like us. Interestingly, however, things that appear to be the same colour under normal light conditions may appear to be different colours under different light conditions. The explanation for this phenomenon, called ‘metamerism’, has to do with broadly physical features of the objects and light at issue.¹⁶ Most notably, what colour we perceive an object to be will be a function of the spectral power distribution (SPD) of the light hitting the retina, specifying the power of the light at each wavelength in the visible spectrum; this SPD is itself is a function of the SPD of the light incident on the object’s surface, and the surface reflectance properties of the object. Different SPDs of light hitting the retina may give rise to the same “tristimulus values” (effectively: hue, saturation, and brightness); hence it is that samples that appear the same in normal light conditions may appear different in other conditions, or that samples that appear different in normal light conditions may appear the same in different light conditions. Metamers are

¹⁶See, e.g., Wandell 1993.

colour appearance properties that are individuated, in part, by the relevant broadly physical features—let's assume these are the retinal SPDs—needed to accommodate the phenomenon of metamerism, such that colour appearance properties not distinguished by hue, saturation, and brightness are distinguished by the relevant broadly physical features.

As I argue in detail in [Wilson 2009](#), metamers are reasonably taken to be colours—just colours seen (no pun intended) through a lens of a finer and partly physical grain. Here I'll just mention two considerations in favour of this claim. First, metamers are part of the broader field of colour science, and they are characterized as colours in that science. In particular, colour science is not concerned only with colours as individuated by hue, saturation, and brightness.¹⁷ On the contrary, considerable colour research is aimed at understanding colours as individuated by broadly physical features such as retinal SPDs, as relevant to digital photography, screen displays, car interiors, etc.¹⁸ Second, the role that retinal SPDs play in this research appears to be compatible with taking colours to be themselves characterized by the relevant broadly physical features. After all, as above, colours are understood as perceptual properties; and retinal SPDs are clearly part of the process of colour perception—in particular, retinal SPDs are input into the colour-sensitive cones, which then output the tristimulus values. Whether or not the input/output function here is causal or rather ‘filter-like’, in any case there seems to be no in-principle barrier to characterizing (specific kinds of) colours in terms of the broader process of visual perception—especially since the broader process and associated features are required to *fully* characterize colour appearances (in particular, metamerism).

¹⁷See, e.g., [Wyszecki and Styles 1982](#).

¹⁸See, e.g., [Judd and Wyszecki 1975](#).

Similar remarks apply to other qualitative conscious states, such as pain. For example, [Funkhouser \(2006\)](#) suggests that states of pain have mental determination dimensions along lines of feel and intensity; but it seems reasonable to suppose that pains that are exactly similar with respect to these psychological determination dimensions might be physically specified, as sciences such as pharmacology take for granted.

I draw two morals from these sorts of case studies. First is that qualitative conscious states, such as states of perceiving colours or experiencing pains, may have physical as well as psychological determination dimensions. Second is that determination dimensions may be science-relative: different sciences may treat the same feature as having different determination dimensions—effectively treating the same feature at different levels of specificity. Relative to normal appearance colour science, *colour* has determination dimensions of hue, saturation and brightness; relative to metameric colour science, *colour* has further determination dimensions. Note that there's nothing mysterious about taking determination dimensions to be science-relative, from a physicalist perspective of the sort we are presently considering. That different sciences may treat the same determinable as having different determination dimensions simply reflects that different sciences and their associated laws may treat certain phenomena at different levels of metaphysical grain.

The observation that contemporary science suggests that conscious mental states may have physical as well as psychological determination dimensions provides a wedge for responding to the objection from mental multiple realizability. Still, the question remains: can we make metaphysical sense of this, compatible with qualitative conscious states being determinables of lower-level physical

determinates?

Here's where Powers-based determination comes in.¹⁹ The basic idea is that the phenomenon of a given determinable's being associated with increasingly fine-grained determinable dimensions, relative to a given science, can be straightforwardly understood in terms of non-conjunctive specification of a determinable's powers. Relative to purely psychological determination dimensions, reflecting sensitivity to the set of powers associated with a multiply realized determinable, determinates of the determinable may be exactly alike. Relative to a finer-grained set of determination dimensions, reflecting sensitivity to powers additionally possessed by the physical realizers of the determinable, determinates of the determinable will not be exactly alike. On this approach, what it is for determinates to be determined 'in respect of' a determinable in the sense relevant to ensuring that the determinable/determinate relation is a distinctive specification relation is cashed in terms of the determinates sharing the powers of the determinable (associated with the psychological determination dimensions) but differing with respect to powers going beyond these (associated with the more fine-grained determination dimensions reflecting physical as well as psychological distinctions). All this is compatible both with the multiple realizers being exactly similar with respect to the determinable's psychological determination dimensions in virtue of sharing the powers had by the determinable—and correspondingly, with qualitative conscious mental states being Weakly emergent, as per a determinable-based account of realization.

¹⁹As I discuss in my (2009), there are likely other approaches to the determinable/determinate relation (as per, e.g., the proposal set out in Funkhouser 2006) that can make sense of science-relativity of determination dimensions.

Accommodating mental super-determination

Recall that the concern at issue in the objection from mental super-determinates is that taking qualitative conscious features to be realized in determinable-based fashion by lower-level physical features is incompatible with the intuitive possibility of there being qualitative mental super-determinates, since implying, falsely, that these could be further determined.

My response again starts with the observation that different sciences may treat a single determinable as having different determination dimensions, such that mental properties may be super-determinate relative to a purely psychological science, while being further determined relative to a lower-level science. What is super-determinate relative to one science may not be super-determinate relative to another.

This much provides a wedge for responding to the objection from mental super-determinates. Still, the question remains: can we make metaphysical sense of this, compatible with qualitative conscious states being determinables of lower-level physical determinates?

Again, Powers-based determination provides a comprehensible metaphysical basis for accommodating the phenomenon at issue. What counts as a super-determinate depends on what determination dimensions are at issue. Relative to one set of determination dimensions, reflecting sensitivity to powers associated with the determinable set (or certain supersets thereof), a given qualitative conscious state might be characterized as a super-determinate of a given qualitative conscious state. Relative to a finer-grained set of determination dimensions (reflecting sensitivity to powers in larger supersets of the determinable set) that same feature might not be appropriately characterized as a super-determinate of that

state.

All this is compatible both with there certain qualitative conscious states being super-determinate with respect to certain psychological determination dimensions, but being further determinable with respect to further, partly physical determination dimensions—and with qualitative conscious mental states being Weakly emergent, as per a determinable-based account of realization.

7.3 Concluding remarks

Let's sum up the results of this chapter. Unlike the phenomena considered in the two previous chapters, consciousness or associated mental states has been frequently offered up as a good—indeed, the best—candidate for an actual Strongly emergent phenomenon. But, I've argued, neither the knowledge argument(s) nor the conceivability arguments establish that conscious features are physically unacceptable, and nor do they establish that such features are, more specifically, Strongly emergent. While the Strong emergence of consciousness remains a live empirical possibility, the presently best case for consciousness's being emergent is one according to which, due to the irreducibly determinable features of qualitative conscious states, consciousness is Weakly emergent, by lights of a determinable-based account. As I've argued, such an understanding of qualitative conscious states can be defended against various objections, given a proper appreciation of the science-relativity of determination dimensions and a proper understanding of the determinable/determinate relation, as per *Powers-based determination*. I conclude that consciousness is at least Weakly emergent.

Chapter 8

Free will

Free will (or free agency), if such there be, involves free choosing: the ability to mentally choose an outcome (an intention to φ , or a φ -ing), where the outcome is ‘free’ in being, in some substantive sense, up to the agent of the choice. Free will has frequently been taken to be core to what it is to be a person, of either human or other varieties, in part (though not exclusively) because such agency seems to be a prerequisite for persons’ being autonomous in the way seemingly crucially relevant to achieving certain moral, aesthetic, and other goods. In this last chapter, I consider reasons for thinking that free will, as it actually exists, might be either Weakly or Strongly emergent.

I start (§1) by drawing on Bernstein and Wilson 2016 in order to set up a useful framework for investigating into the status as emergent of free will. Recall, to start, that the schemas for Weak and Strong emergence were initially motivated as associated with two specific responses to the problem of higher-level causation; namely, non-reductive physicalism and Strong emergentism. A common focus of this problem concerns the status of mental features (events, states, properties, and

so on); but in the usual case the mental features at issue are qualitative or intentional mental features for which free choice is supposed not to be at issue. More generally, debates over the status of free will have tended to proceed in relative independence from debates over the status of mental features whose governance by natural law is taken for granted. As Bernstein and I argue, however, the problematics underlying the free will and the mental causation debates are appropriately seen as special cases of a more general problem, concerning whether and how mental features of a given type may be efficacious, *qua* the types of feature they are (qualitative, intentional, freely deliberative), given their apparent causal irrelevance (i.e., failure of distinctive efficacy) for effects of the type in question. (The literature I will be discussing typically focuses more specifically on mental *events* of a given type, so I will present the parallel in these terms.)

That the free will and mental causation debates can be seen as special cases of a more general problem serves to suggest certain parallels between positions in the respective debates, which parallels are useful for purposes of assessing whether free will is either Weakly or Strongly emergent. In particular, as Bernstein and I argue, a representative range of compatibilist (or ‘soft determinist’¹) accounts of free will implement a strategy that is structurally similar to the ‘proper subset’ strategy that, I have argued, is core and crucial to non-reductive physicalist accounts of realization, and more generally, to Weak emergence. I start by presenting and extending this result, arguing that the compatibilist’s proper subset strategy is reasonably taken to indicate that compatibilist free will, were it to ex-

¹In general, compatibilists maintain, as Lewis (1981) puts it, that “soft determinism may be true”. Beyond this commitment, individual compatibilists may aim simply to undermine arguments for incompatibility, or moreover to provide a positive conception of the compatibility at issue (see McKenna and Coates 2008 for discussion). Here by ‘compatibilism’ I mean ‘positive compatibilism’.

ist, would be Weakly emergent. I go on to argue that a representative range of libertarian treatments of free will are appropriately seen as committed to such agency's involving a fundamentally novel power, such that Libertarian free will, were it to exist, would be Strongly emergent. In a final section, I consider the prospects of there being emergent free will, suggesting that free will of the compatibilist, Weakly emergent variety is common, and providing a new argument for there moreover being free will of the Libertarian, Strongly emergent variety.

8.1 The generalized problem of mental quausation

In our (2016), Bernstein and I argue that (certain understandings of) the problems of free will and of mental causation can be seen as special cases of a more general problem, concerning whether and how mental events of a given type may be efficacious, *qua* the types of feature they are—qualitative, intentional, freely deliberative, and so on—given their apparent causal irrelevancy for effects of the type in question.

In making this connection, we generalize what Horgan (1989) calls as “the problem of mental quausation”.² As Horgan presents it, this problem is a refinement of the problem of mental causation (a special case of the problem of higher-level causation presented in Chapter 2). The latter problem is sometimes pitched as the problem of how a (real, distinct, synchronically materially dependent) qualitative or intentional mental event *M* might be efficacious *at all*, given that any effect *E* it might be seen as causing is, by *Physical Causal Closure*, already caused by a physical event *P*. But as Horgan notes, there is a quick route to

²Though the focus here is on mental goings-on, the more general problem of ‘quausation’ here can be seen as applying to any seemingly higher-level goings-on.

gaining M 's efficacy—namely, by identifying M with P , as per the usual reductive physicalist response—that leaves open what is arguably the deeper question underlying the original problem:

Even if individual mental events and states are causally efficacious, are they efficacious *qua* mental? I.e., do the mental types (properties) tokened by mental events and states have the kind of relevance to individual causal transactions which allows these properties to figure in genuine causal explanations? (47)

What is needed to adequately vindicate the efficacy of the mental, Horgan suggests, is that mental events be shown to be distinctively efficacious: efficacious *qua* mental—and more specifically, “*qua F*”, where F is schematic for a specific type of mental event” (50). As such, it is mental ‘quausation’, not mental causation *per se*, that is most deeply challenged by the possible truth of *Physical Causal Closure*. How could a mental event M be efficacious *vis-á-vis* an effect E , in virtue of being qualitative or intentional, given that E was causally determined by physical events and associated laws in ways that seem to preclude M 's being causally relevant in either of these respects?

Though Horgan's discussion targets mental events for which freedom is not at issue, the deeper concern about whether and how mental events can be seen as causally relevant—that is, efficacious in virtue of, or *qua*, the distinctive mental types they are—lies also at the core of what is sometimes called ‘the consequence argument’, ‘the problem of free will and determinism’, or just ‘the problem of free will’. This problem highlights a seeming tension between an intuitive conception of free will as involving the ability to freely choose an outcome φ , and the broadly scientific thesis of *Determinism*, according to which every event is a consequence

of the laws of nature and the state of the world at any time.³ *Determinism* admits of different interpretations;⁴ what is at issue here is a reading involving causal determination of present or future events by prior states (broadly construed), as follows:⁵

Causal Determinism: With the exception of any first events there might be, every event is a causal consequence of the laws of nature and the state of the world at any prior time.

The possible truth of *Causal Determinism* leads to a question: If every event *E* (e.g., an intention to φ , or a φ -ing) purportedly caused by a mental event of free choosing *M* is, by *Causal Determinism*, a causal consequence of the laws of nature and prior states, what causal role is left for *M* to play, *vis-à-vis E*? The answer, according to (a causal reading of) the consequence argument, is that no role is left.

As van Inwagen (1983) puts it:

If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born; and neither is it up to us what the laws

³One might wonder whether *Determinism* can be ruled out of court, on grounds that quantum mechanics, which is likely true, is indeterministic; but this would be premature, both because there are deterministic interpretations of quantum mechanics (see, e.g., Bohm 1952), and because the present incompatibility of quantum mechanics and general relativity suggests that we are not quite yet in position to infer to the likely truth of quantum mechanics. (See O'Connor 2002, 4 for similar observations.) Of course, even if *Determinism* is actually false, one might be interested in understanding what bearing the truth of *Determinism* would have on free will. More importantly, the parallel doesn't crucially depend on the assumption of *Determinism*, since concerns about the causal relevance of free will also arise if the outcomes of seemingly free choices are the product of indeterministic (e.g., quantum) laws.

⁴Steward (2015) distinguishes between interpretations based on entailment between propositions at one time by propositions at another time, and those based on causation between events.

⁵Our focus here thus excludes compatibilists whose rejection of causation or causal production between events—Leibniz, and perhaps also Hume—would entail their rejection of a causal reading of *Determinism*.

of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.⁶ (16)

The question and *prima facie* negative answer constitute what van Inwagen calls “the problem of free will and determinism”, and what Bernstein and I call, for short, ‘the problem of free will’.

In the problem of free will, it is not the mere efficacy of deliberative mental events (as, perhaps, the most proximate causes of intentions or actions in a causally determined chain of events) that is at issue; indeed, that events of choosing are efficacious is typically taken for granted in the free will debates. Rather, what is at issue is whether events of choosing can be efficacious *qua* free—again, in eventuating in outcomes that are, in some substantive sense, up to the agent—under circumstances where *Causal Determinism* is presumed to be true. How could an event of mental choosing M be efficacious *vis-á-vis* an event E , in virtue of being *free*, given that E was causally determined by laws of nature and events antecedent to M in ways that seem to preclude M ’s being causally relevant in this respect?

This parallel suggests that the problem posed by *Physical Causal Closure* for qualitative and intentional mental events, and the problem posed by *Causal Determinism* for events of seemingly free choosing, are each instances of a suitably general problem (or question, if you like) of mental quausation:

How can a mental event M of a given type be efficacious *vis-á-vis* an event E , in virtue of being the type of mental event it is, given that there is reason to think that events of M ’s type are causally irrelevant to the production of events of E ’s type?

⁶Van Inwagen’s remarks are directed at an entailment-based reading of *Determinism*, but they apply, *mutatis mutandis*, to a causal reading (simply insert ‘causal’ before ‘consequences’).

Note that in drawing this parallel, Bernstein and I are not suggesting that there is any interesting parallel between the problem of free will and the *unrefined* problem of mental (more generally, higher-level) causation; nor are we suggesting that the problem of free will is an instance of the problem of mental quausation as applied to qualitative or intentional mental events; after all, there is no clear tension between *Causal Determinism* and mental quausation involving mental causes for which freedom is not at issue. Our point is simply that both the problem of free will and the refined problem of mental causation (again, typically directed at qualitative and intentional mental events for which free will is not at issue) are specific cases of a suitably general problem of mental quausation, whereby the causal relevance (distinctive efficacy) of a mental event of a distinctive type—whether intentional, qualitative, or freely deliberative—is called into question by the holding of certain theses (*Physical Causal Closure*, *Causal Determinism*) which are to some extent live possibilities for being true.⁷

The traditional responses to the problem of free will may be categorized by reference to the following free will conditional:

If all events are subsumed by deterministic natural laws, then free mental quausation—the causation of events (e.g., intentions to φ , φ -ings) by mental choosings *qua* free—does not exist.

Hard determinists hold that both antecedent and consequent are true; soft determinists or ‘positive’ compatibilists (henceforth, just ‘compatibilists’) take the antecedent to be true and the consequent to be false; libertarians hold that both antecedent and consequent are false.

⁷Again, thought it is convenient to present the parallel in terms of the problem for free will posed by the assumption of *Determinism*, concerns have also been raised for the causal relevance of free will on the assumption that free choices are the product of indeterministic laws.

Similarly, the traditional responses to the refined problem of mental causation can be categorized by reference to the following mental causation conditional:

If all physical events are subsumed by physical laws, then qualitative and intentional mental causation—the causation of physical events (e.g., bodily movements) by mental events *qua* qualitative or intentional—does not exist.

Eliminativist physicalists and epiphenomenalists hold that both antecedent and consequent are true; reductive and non-reductive physicalists (Weak emergentists) hold that the antecedent is true and the consequent is false; Strong emergentists and substance dualists hold that both antecedent and consequent are false.

Since the problem of free will and the refined problem of mental causation may each be seen as special cases of a more general problem, we might expect there to be parallels between the primary responses to each problem, corresponding to parallels in the stances taken towards the components of the corresponding conditional. As I will argue in the following two sections, this is indeed the case for compatibilism and libertarianism, respectively.⁸

8.2 Compatibilism and Weak emergence

I start with compatibilism, according to which the antecedent of the free will conditional (*Causal Determinism*) is true (or in any case might be true), but the consequent (denying the existence of free will) is false. Here two positions—reductive and non-reductive physicalism—take the same stance as regards the refined mental causation conditional. As previously discussed, however, reductive versions of

⁸There is also a parallel between hard determinism and eliminativism; see Bernstein and Wilson 2016 for discussion.

physicalism, which identify the mental events at issue with physical events, face immediate difficulties in making sense of how mental events can be efficacious *qua* the qualitative or intentional types they are, with (as per the discussion of deflationary or reductive accounts of higher-level phenomena in Chapter 3) the usual strategies being to offer pragmatic or purely epistemic accounts of the desired “higher-level” efficacy. Given this, and since the deeper concern at issue in the problem of free will is with whether there is a genuinely metaphysical basis for free mental quausation, we can cut to the chase of considering whether there is an interesting parallel between non-reductive physicalism (Weak emergentism), which standardly aims to make metaphysical sense of distinctive mental (higher-level) efficacy, and compatibilism.

8.2.1 Weak emergence and the non-reductive physicalist’s proper subset strategy

At this point, the reader is well familiar with the characteristic strategy of the non-reductive physicalist response to the problem of higher-level quausation, as encoded in the schematic conditions in Weak emergence:

Weak emergence: Token apparently higher-level feature S is weakly metaphysically emergent from token lower-level feature P on a given occasion just in case, on that occasion, (i) S synchronically materially depends on P ; and (ii) S has a non-empty proper subset of the token powers had by P .

As setup for motivating the structural parallel with compatibilism, it is worth recalling how satisfaction of these conditions—most crucially, satisfaction of the Proper Subset condition on the token powers of the higher- and lower-level fea-

tures at issue—operates to secure, not just the reality and efficacy of the higher-level goings-on, but also their causal relevance/distinctive efficacy. Again, the core idea is that there are two ways for a higher-level (here, mental) feature M to be distinctively efficacious as compared to its base feature P . One way is for M to be associated with a *new power*—a power that P doesn't have; this is the form of distinctive efficacy at issue in the Strong emergentist strategy, to which we will return. A second way reflects that M is associated with a *distinctive power profile* consisting of a proper subset of the powers associated with P ; this is the form of distinctive efficacy at issue in the Weak emergentist strategy.

Recall also that there are (at least) two strategies which might be used to motivate taking the having of a distinctive power profile conforming to Weak emergence to provide a basis for S 's being efficacious *qua* the type of higher-level feature it is: first, by observing that such a power profile provides a basis for difference-making or “proportionality” considerations (where the effect would still have occurred and been caused by S even had S been differently realized); second, by observing that such a power profile may be associated with laws (or a system of laws) which are comparatively insensitive to the details that are needed for the lower-level physical laws governing P to operate. Applied to the qualitative or intentional mental events associated with the problem of mental quausation: even if every token power of a given such mental event M , on a given occasion, is identical with a token power of its physical realizer P on that occasion, M can be distinctively efficacious—that is, causally efficacious *qua* qualitative or intentional, in virtue of M 's distinctive power profile tracking the comparatively abstract causal level of grain associated with (non-agential) psychology.

8.2.2 The compatibilist's proper subset strategy

Interestingly, as Bernstein and I argue, seemingly diverse compatibilist accounts implement a structurally similar ‘proper subset’ strategy for responding to the problem of free will. To prefigure: in the case of compatibilist free will, the operative strategy involves characterizing events of seemingly free choosing as associated with only a proper subset of the causal determinants of the outcome (effect) at issue, in such a way as to provide a principled basis for the claim that the choosing is efficacious *qua* free.

Hawthorne and Pettit’s (1996) taxonomy of compatibilist strategies serves as a useful basis of operations. They start by noting:

All compatibilists agree that every choice has antecedents and [...] that this fact puts freedom of choice in doubt. How can a choice be made freely if it is the product of independent antecedents? The response they make is that some possible antecedents are better than others from the point of view of free choice and that a choice is free to the extent that its antecedents, or at least its relevant antecedents, satisfy the inherently vague condition of leaving it up to the agent. (191)

In schematic form:

X chooses freely to φ if and only if the relevant antecedents of the choice leave the φ -ing up to X . (191)

As above, for present purposes a choice to φ (the outcome of an event of choosing) may be either an intention to φ , or a φ -ing.

Hawthorne and Pettit identify three main compatibilist accounts of what it is for a choice to be “up to an agent”, associated with the notions of freedom as underdetermination, ownership, and responsibility, respectively. The accounts

vary to some extent as regards which antecedents are supposed to be relevant to establishing whether the choice was up to the agent, in the intended sense. As Bernstein and I argue, however, each of these accounts plausibly impose satisfaction of a certain ‘proper subset’ condition, as key to their strategy of response to the problem of free will.

I start by motivating the condition. First, reflecting endorsement of (the live possibility of) *Causal Determinism*, the compatibilist accepts the following condition on the causal antecedents of any outcome of a free mental deliberation:

Causal Antecedents Condition: The total causal antecedents of an event of free choosing M completely determine the outcome of M (e.g., a choice to φ).

As a first pass, the compatibilist strategy requires that a free mental choosing M satisfy the following proper subset condition:

Subset of Causal Antecedents Condition (first pass): The relevant causal antecedents $\{C\}$ of a free mental choosing M are (i) a non-empty proper subset of the total causal antecedents of M , which (ii) satisfy the condition of leaving the outcome of M up to the agent.

Moreover, as will shortly become clear, if the compatibilist’s strategy of identifying the relevant causal antecedents of M is to make sense of the idea that these antecedents leave the choice up to the agent, then the relevant antecedents must more specifically satisfy the following (final pass) proper subset condition:

Proper Subset of Causal Antecedents Condition: A free mental choosing M resulting in a choice to φ satisfies the following: (i) M has relevant causal antecedents $\{C\}$ which are a non-empty proper subset of the total causal antecedents of M , and (ii) it is possible that a choosing M' of the same type as M occur, having relevant antecedents $\{C'\}$ of the same type as $\{C\}$, but where the *total* antecedents of M' are such

as to completely determine the outcome of M^9 as either a choice not to φ or as the absence of a choice to φ .

After arguing that the following three compatibilist accounts aim to satisfy this condition, I will go beyond Bernstein's and my discussion, to make the case that the compatibilist's *Proper Subset of Causal Antecedents Condition* is not just structurally similar to, but can moreover be plausibly understood as entailing, the *Proper Subset of Powers Condition* operative in the schema for Weak emergence. This will set up for the later discussion of whether compatibilist free will (hence persons characterized as having such agency) is at least Weakly emergent.

Freedom as underdetermination

On underdetermination accounts, a choice to φ is the result of a free choosing M iff M could have resulted in a choice not to φ . How could this be, given that the choice to φ was determined, as per the *Causal Antecedents Condition* (encoding *Determinism*)? The underdetermination approach proceeds by identifying a subset {C} of the causal antecedents of the choice to φ relative to which it was left open whether or not M would result in a choice to φ . As [Hawthorne and Pettit \(1996\)](#) put it:

Taken as a whole, the antecedents of any choice will necessitate that choice under a deterministic picture and compatibilists of this stripe must take the relevant antecedents to be a subset of the totality. But which subset? (193)

The relevant subset of antecedent events will include the choosing event, along with events tracking the standing beliefs and desires of the agent at the time of

⁹As per the *Causal Antecedents Condition*; the possibility of indeterministic scenarios is put aside here.

choosing, and events tracking whether the choosing took place under conditions of physical restraint, threat, etc. (c.f. Ayer 1954). From broader perspectives, the relevant antecedents might also include events tracking cultural influences, past trauma, or other psychological, social, psychiatric, neurological, etc., conditions holding of the agent. In general, Hawthorne and Pettit note,

The line will be that an agent is free to the extent that the antecedents that can or have to be countenanced in that perspective leave the choice underdetermined. [...] To be free, if you like, is to be free relative to that stance. (193–4)

Here the relevant antecedent events must be a *proper* subset of the causal antecedents that, as per the *Causal Antecedents Condition*, completely determine the choice, since only if the subset is proper is there any hope that the subset of antecedents will leave that choice undetermined. Moreover, the assumption that the subset of antecedents leaves the choice undetermined plausibly entails (indeed, has as its content) that it is possible that a choosing M' of the same type as M occur, having relevant antecedents C' of the same type as M 's relevant antecedents C , but where the *total* antecedents of M' are such as to completely determine the outcome of M' as either a choice not to φ or the absence of a choice to φ .

This last entailment just is the *Subset of Causal Antecedents Condition*. As such, a freedom as underdetermination form of compatibilism explicitly implements a proper subset strategy, characterizing a mental choosing M as associated with a proper subset of its causal antecedents, and using this association to accommodate M 's being distinctively efficacious, *qua* free, *vis-à-vis* the ensuing choice.

Freedom as ownership

A second compatibilist approach takes freedom to be a matter of ownership:

The ownership line takes a choice to be up to an agent to the extent that it is not due to anyone or anything other than the agent themselves; it is a choice that the agent owns, a choice with which the agent identifies, and not something forced upon them. Suppose that the relevant antecedents in the adjudication of free will are taken to be [...] beliefs and desires. [Then] an agent φ s freely just in case their beliefs and desires combine to lead—or at least lead in ‘the right way’ (see [Davidson 1963](#)) to their φ -ing. ([Hawthorne and Pettit 1996](#), 194)

Underdetermination by the relevant antecedents is not explicitly required here, since an agent could own or identify with completely determined intentions. But, Hawthorne and Pettit argue, if the ‘ownership’ line is to be viable, it will have to ensure underdetermination by these antecedents, and so satisfaction of the *Subset of Causal Antecedents Condition*.

Why so? To start, note that if, for example, I am brainwashed with beliefs and desires leading to my choice to φ , this intention cannot be seen as the effect of free mental choosing. A well-known response (see [Frankfurt 1971](#)) requires that choices result from desires that the agent X desires, at the second order, to have and be moved by. The brainwashing problem will re-arise, however, unless “the action issues from desires that the agent has some measure of second-order control over” ([Hawthorne and Pettit 1996](#), 195). As [O’Connor \(2002\)](#) puts it:

We can [...] imagine external manipulation consistent with Frankfurt’s account of freedom but inconsistent with freedom itself. [...] one might discreetly induce a second-order desire in me to be moved by a first-order desire—a higher-order desire with which I am satisfied—and then let me deliberate as normal. Clearly, this desire should be deemed “external” to me, and the action that flows from it unfree.

These considerations indicate that one needs to ensure, somehow, that the formation of second-order desires is up to the agent; and the natural Compatibilist approach will be to restrict attention to a proper subset of the antecedents determining the desire—e.g., those relevant to whether the agent’s choosing was constrained by other persons, or by other psychological, social, psychiatric, neurological, etc., conditions. In other words, to accommodate the needed control of second-order desires on a “freedom as ownership” picture, a proper subset of the antecedents of the choice to φ must be specified, relative to which it was underdetermined that the agent had the second-order desires they had; hence underdetermined that the agent would identify with the first-order beliefs and desires leading to the agent’s choice to φ ; hence underdetermined that the agent’s choosing would result in a choice to φ . Such underdetermination in turn entails (indeed, has as its content) satisfaction of the *Subset of Causal Antecedents Condition*.

More generally, here again a proper subset strategy is implemented, whereby a mental choosing M is associated with a proper subset of its causal antecedents, in service of making room, as per the *Subset of Causal Antecedents Condition*, for M to be causally relevant *vis-á-vis* the ensuing choice.

Freedom as responsibility

On the “freedom as responsibility” approach, a choice to φ is the result of a free choosing M iff the agent of the choosing could be held responsible for the outcome of M . The criteria for an agent’s being responsible might advert to prevailing systems of law and morality, or (as per Strawson 1962) to the participant or reactive attitudes characteristic of human interactions. As Hawthorne and Pettit point out, this approach too requires that the choice at issue be underdetermined:

To hold an agent responsible in certain choices is to think that it is not inevitable either that they get things right or that they get them wrong—either that they do well or that they do ill—and so it is to believe that there is a sense in which they could have done otherwise [...] . (197)

The relevant antecedents in this case would then include those relevant to determining whether the agent was deliberating under conditions where they would, by the lights of the prevailing system of law, morality or interaction, be held responsible for the outcomes of their choosings. Again, these might cite events tracking whether various physical, psychological, neurophysiological, etc., conditions or constraints were in place antecedent to or concurrent with the choosing.

Here again, the account (i) identifies a relevant proper subset of M 's causal antecedents, and (ii) requires that, relative to these antecedents, the outcome of M could have been different, as a way of making sense of M 's being efficacious, *qua* free. In other words, a freedom as responsibility account implements a proper subset strategy, encoded in satisfaction of the *Subset of Causal Antecedents Condition*.

8.2.3 Deepening the parallel: a powers-based interpretation of the compatibilist's proper subset condition

Non-reductive physicalist and compatibilist positions thus each respond to their respective problems by characterizing the mental events at issue as associated with a proper subset of the “causal determinants” of their associated effects. As Bernstein and I present the structural similarity, the determinants are not the same: in the one case, these are powers; in the other, these are causal antecedents. This much establishes a structural similarity in strategies: in each case, associating

the mental event M with (only) the relevant proper subset of causal determinants provides a basis for showing that M is causally relevant to the production of the effect E in question, *qua* the type of mental event M is.

As I will now argue, however, the parallel is even deeper: the compatibilist strategy can be understood as entailing the holding of a proper subset relation between token powers associated with two complex, broadly synchronic events, corresponding to, first, the mental choosing M in combination with the relevant antecedents of M (call this complex event C'), and second, the mental choosing M in combination with the total antecedents of M (call this complex event C).

To start, note that there is no in-principle problem with associating a set of token powers with either C' or C . After all, the events, entities, processes, or other goings-on at issue in debates over the status of higher-level entities are often complex and typically temporally extended; and there is no in-principle problem with assigning powers to such goings-on (reflecting, e.g., the operative laws of nature). As such, we are in position to see that the compatibilist's proper subset strategy entails that C' has fewer token powers than C . For this strategy plausibly entails that, on the one hand, every token power of C' is identical to a token power of C , but that C' has fewer powers than C ; and this is because, while C has the power to result in a choice to φ , C' does not have this power—for C' could occur in circumstances in which M (or an event of M 's type) would have resulted either in a choice not to φ or in the absence of a choice to φ .

On this reconstruction of the compatibilist strategy, it would not be the seemingly free choosing itself that would be appropriately deemed Weakly emergent, but rather the seemingly free choosing in combination with the relevant proper subset of antecedents. Such a result makes sense, on a compatibilist picture. Un-

like the case of a Weakly emergent qualitative feature, for example, the status of the event of choosing as genuinely free requires that the choosing be associated with causal antecedents that in the relevant sense leave the outcome up to the agent; hence on a compatibilist view, the freedom of a given act of choosing can be seen as constituted by the occurrence of a complex temporally and causally extended event, consisting in the choosing in combination with the relevant causal antecedents.

The previous considerations indicate that free will on a compatibilist account can be seen as satisfying each of the conditions in the schema for Weak emergence—that is, as Weakly emergent.

This result provides a new basis for addressing a frequently voiced concern about the compatibilist's strategy—namely, that identification of a given subset of causal antecedents won't make sense of how the choosing could be free, since the mere presence of a subset of antecedents doesn't establish that M 'selects' or 'determines' the outcomes of the choosing.

To start, the compatibilist, like the non-reductive physicalist, will grant that M 's distinctive efficacy *vis-á-vis* the effect at issue doesn't proceed by way of M 's having a distinctive power: just as *Physical Causal Closure* blocks taking a qualitative or intentional mental event M to have a new power (that would be Strong emergence, not physicalism), so too does *Causal Compatibilism* block taking a mental event of choosing M to have a new power (that would be libertarianism, not compatibilism). Even so, just as the non-reductive physicalist has alternative ways of motivating the distinctive efficacy of qualitative and intentional mental events—either as tracking difference-making considerations (if the physical realizer had been slightly different, I would still have been thirsty), or as tracking

a distinctive comparatively abstract psychological level of causal grain—so too may the compatibilist maintain that even in the absence of new powers to ‘select’ or ‘determine’ outcomes, M may be distinctively efficacious *vis-à-vis* those outcomes, either in tracking difference-making considerations (if the causal antecedents of my choice had been slightly different, I would still have chosen as I did) or as tracking a distinctive broadly psychological level of causal grain. The distinctive form of causal relevance identified by non-reductive physicalists—that is, that encoded in the schema for Weak emergence—appears to be, *mutatis mutandis*, just what the compatibilist needs. Of course, compatibilism faces other challenges (a point to which I will later return), but in any case it is worth noting that the parallel to non-reductive physicalism is useful in clarifying just what the compatibilist strategy for achieving the autonomy of free will is supposed to be.

8.3 Libertarianism and Strong emergence

I next turn to libertarianism (a.k.a. ‘incompatibilism’),¹⁰ according to which both the antecedent (*Causal Determinism*) and the consequent (denying the existence of free will) of the free will conditional are false. More specifically, I will argue that there is a parallel between Libertarianism, on the one hand, and Strong emergentism as standardly directed at qualitative and non-intentional mental states, on the other.

An initial point of similarity between libertarianism and Strong emergentism is that each view rejects the broadly empirical thesis causing trouble for the sup-

¹⁰Here again we can distinguish incompatibilist views which simply deny the compatibility of free will and Determinism and those which moreover aim to give some positive account of the nature of incompatibilist free will; at issue in this discussion are what we might call ‘positive incompatibilist’ views.

position that the higher-level goings-on are causally relevant in the intended sense.

As [Clark and Capes \(2017\)](#) put it:

To have free will is to have what it takes to act freely. When an agent acts freely—when she exercises free will—it is up to her whether she does one thing or another on that occasion. A plurality of alternatives is open to her, and she determines which she pursues. When she does, she is an ultimate source or origin of her action. So runs a familiar conception of free will.

Incompatibilists hold that we act freely in this sense only if determinism is false.

Similarly, Strong emergentists maintain that an appropriate understanding of the efficacy of certain goings-on requires the falsity of *Physical Causal Closure*.

The deeper point of similarity, however, lies in the positive accounts given of the existence and causal relevance of the higher-level goings-on at issue. Recall that the conditions in the schema for Strong emergence, which takes its original inspiration from the Strong emergentist strategy for responding to the problem of higher-level causation, require that a Strongly emergent feature have a fundamentally novel power not had (or only indirectly had, in virtue of being a necessary precondition of the Strongly emergent feature) by the lower-level physical feature upon which it minimally nomologically depends, which novel power in turn provides a principled metaphysical basis for both the ontological and causal autonomy of the Strongly emergent feature. As I will shortly argue, a representative range of libertarian accounts are reasonably seen as committed to free will's satisfying these conditions, and in particular to taking free will to be associated with a novel fundamental power—namely, the power to freely choose to φ , where as previously, a choice to φ (an outcome of an event of choosing) may be either an intention to φ , or a φ -ing.

In making this case, I'll help myself to a commonly acknowledged tripartite taxonomy of Libertarian accounts, as falling under noncausal, event causal, and agent causal varieties. As [Clark and Capes \(2017\)](#) note:

The incompatibilist theories that have been offered fall into three main groups, depending on which type of indeterminism (uncaused events, nondeterministically caused events, agent- [or substance-] caused events) they require. Further variations among accounts concern where in the processes leading to decisions or other actions they require indeterminism and what other conditions besides indeterminism they require.

(See also [O'Connor 2002](#).) To fix ideas, I will primarily (though not exclusively) focus on representative instances of the accounts, as proposed by Ginet, Kane, and O'Connor, respectively. Each of these accounts can be seen as offering a different positive account of what makes a given act of choosing free in the relevant (strong, incompatibilist) sense, against the common backdrop assumption of the rejection of *Determinism*. The case for Libertarian free will's being understood as Strongly emergent is most straightforward for the two causalist accounts, so I'll start with those; I'll then make a case that even so-called 'noncausalist' accounts are plausibly committed to free will's involving a (fundamentally) novel power, as Strong emergence requires.

Before getting started, a couple of observations that will sometimes enter into what follows.

First, though it's more or less common ground that any variety of free will worth the name (whether compatibilist or incompatibilist) has to make sense of a choice to φ being in some sense 'up to the agent', the issue of what we might call 'agential control' is often interwoven with what sort of free will might make room for moral responsibility, with concomitant detailed discussion of the role

of character, reasons, values, self-identification, conscious and unconscious intentions, and so on (call these the ‘moral virtues’, for short). This is even more the case in presentations or defenses of libertarian accounts, perhaps reflecting a purported primary concern with such accounts that locating free will in some sort of indeterminacy would undermine agential control and hence the associated basis for moral responsibility. My own view is that this mixing of the metaphysics of free will with the question of how such agency bears on the moral virtues is ill-advised. A more systematic approach, it seems to me, is to start by getting clear about what options the libertarian has for accommodating simple cases of seemingly free choosing, such as a case in which one considers whether to throw a piece of chalk in the air and then seemingly freely determines to do so (or not), leaving for later treatment the question of how the available options comport with the deeply complex further issues of self-definition, values, moral responsibility, and so on.

Second, and by way of partial explanation, perhaps, of the too-quick and overly detailed introduction of attention to moral virtues in these discussions, an oft-stated concern with libertarian accounts is that, insofar as they reject *Causal Determinism*, they must be committed to thinking that the outcomes of free choices are somehow indeterministically caused, via quantum or other ‘chance-y’ processes, in a way that would undermine agential control and hence the associated basis for moral responsibility. To be sure, some libertarians (e.g., Kane; see below) embrace a kind of analogy between the indeterminacy of free choice and that of quantum indeterminacy. It is important to realize, however, that the libertarian is under no obligation to suppose that the only account alternative to one on which choices are nomologically deterministic is one according to which they are nomo-

logically indeterministic. On the contrary: to my mind, the natural opposition to the purported incompatibility of free will and *Determinism* is one according to which agents are *transcendentally* free—that is, free, in at least some of their choices, from *either* deterministic or indeterministic laws.

Third, some discussions of libertarianism presuppose, to my metaphysician's mind, an odd assumption about causation—namely, that 'event' causation and 'agent' causation are distinct varieties of causation; relatedly, discussions of agent causation frequently import the assumption that such causation would be causation by a 'substance', which in turn is supposed to be problematic, or in any case unusual. My own view is that there is no need to introduce a new variety of causation, much less a new substance, in order to accommodate the sort of transcendental freedom that agents possess (or should be thought to possess), on a libertarian view.

Insofar as I believe that certain methodological presuppositions of the accounts to follow are problematic, my discussion will at times depart from the letter of certain libertarian views, and advance theses that to my mind do better at metaphysically accommodating their spirit.

8.3.1 Event-causal accounts

Clark and Capes (2017) describe event-causal libertarian accounts as follows:

Compatibilist accounts of free will are typically event-causal views, invoking event-causal accounts of action. The simplest event-causal incompatibilist theory takes the requirements of a good compatibilist account and adds that certain agent-involving events that cause the action must nondeterministically cause it. When these conditions are satisfied, it is held, the agent exercises in performing her action a certain variety of active control (which is said to consist in the action's

being caused, in an appropriate way, by those agent-involving events), the action is performed for a reason, and there remains, until she acts, a chance of the agent's not performing that action.

Standard varieties of such accounts (a.k.a. “centered accounts”) locate the indeterministic causation at issue in the immediate causal antecedents of the choice (as opposed to some prior indeterministic process leading to certain beliefs becoming salient or certain preferences being formed, which beliefs or preferences enter into the process of deliberation).

Does the indeterministic causation at issue involve a fundamentally novel power? Arguably, event-causal libertarians should think so, if they aim to decisively answer the so-called “objection from luck”, to which such accounts are “widely thought to be vulnerable”. As [Clark and Capes \(2017\)](#) put the concern:

If a decision is nondeterministically caused, and if there remains until it occurs a chance that the agent will instead (at that moment) make a different decision, then there is a possible world that is exactly the same as the actual world up until the time of the decision, but in which the agent makes the alternative decision then. There is, then, nothing about the agent prior to the decision—indeed, there is nothing about the world prior to that time—that accounts for the difference between her making one decision and her making the other. This difference, then, is just a matter of luck. And if the difference between the agent's making one decision and her instead making another is just a matter of luck, she cannot be responsible for the decision that she makes.

Consider, for example, Kane's centered causal account.¹¹ Bracketing certain nuances, Kane maintains that a choice to φ is one that involves a “self-forming willing”—an indeterministically caused choice or other action for which the agent is “ultimately responsible” (1996b: 35).¹² Of course, not all free choices need be

¹¹See also [Nozick 1995](#), [Ekstrom 2001](#), and [Franklin 2018](#).

¹²The nuance consists in allowing that a choice might be free even if causally determined, so long as it at least partly resulted from a self-forming willing.

“self-forming”—here we see the unhelpful mixture of the more basic metaphysical question of in what libertarian free will consists with the much more complex question of how our free choices enter into self-constitution and moral responsibility. In any case, as [Clark and Capes \(2017\)](#) point out, it remains unclear how appeal to indeterministically efficacious “willings” is supposed to answer the objection from luck—at least if the indeterminacy of will or effort is, as Kane unwisely suggests, analogous to the sort of indeterministic processes associated with quantum phenomena. Such an appeal not only raises the specter of the objection from luck, but also problematically suggests that free will is, like quantum goings-on, still caught in the net of nomological causation. That’s not what a so-called ‘libertarian’ should say, or so it seems to me.

At any rate, there doesn’t seem to be any reason why an event-causal libertarian such as Kane couldn’t rather maintain that the indeterminacy of agent-involving events of free choosing reflects not an analogy with (much less a basis in) indeterministic quantum goings-on, but rather that agential control involves a fundamentally novel power to choose in a way that is not a matter of either deterministic or indeterministic nomological processes. As the simple cases suggest, the novel power here is not, at least in the first instance, essentially tied to morally loaded notions involving character, morality, or the like. Rather, the power is simply the power to choose, to make a choice to intend or to act, in a way that transcends any nomological goings-on, and for whatever reason the agent happens to find compelling in the moment,¹³ or indeed, for no reason at all (as in the case of the throwing of the chalk). To be sure, as embodied, persons are subject to various physical and psychological limitations: to be capable of choosing to

¹³Here an underappreciated resource might advert to the phenomenon of attention.

intend or to act, in a way that is neither deterministically nor indeterministically nomologically determined, does not mean that all bets, or all laws of nature, are off. Still, as persons we are capable of choosing to φ in a way that is plausibly fundamentally novel, insofar as any non-agents with which we are familiar do not have such a power to freely choose. Or so an event-causal libertarian can and should say.

I interpret Merricks as endorsing such a view. Merricks argues, as regards conscious persons, that “we should say that some of what those objects cause, in virtue of having those properties, lack microphysical causes” (110). Merricks’s claim here is compatible with an event-causal account, and he is clear that he sees free will as involving a power that lower-level physical goings-on do not have—that is, as involving a fundamentally novel power, of the sort that satisfaction of the schema for Strong emergence requires:

Sometimes my *deciding* to do such and such is what causes the atoms of my arm to move as they do. Presumably my so deciding won’t ever be the *only* cause of their moving. There will also be a cause in terms of microphysics or microbiology, in terms of nerve impulses and the like. But at some point in tracing back the causal origin of my arm’s moving (if it is intended), we will reach a cause that is *not* microphysical, that just is the agent’s *deciding* to do something. (110)

Merricks distinguishes his view from that of the British Emergentists, on grounds that they “seem to explain *being emergent* in *epistemic* terms” (111, note 13). However, as previously discussed, that the British emergentists sometimes characterized emergence in epistemic terms reflected their (incorrect) assumption that certain epistemic failures would track fundamental novelty; properly understood, Merricks and the British emergentists are on the same page. More generally, in Merricks’s view, we have a form of event-causal libertarianism that satisfies the

conditions of the schema for Strong emergence.

8.3.2 Agent-causal accounts

Clark and Capes (2017) describe agent-causal Libertarian accounts as follows:

On what are called agent-causal views, causation by an agent is held not to consist in causation by events (such as the agent's recognizing certain reasons). An agent, it is said, is a persisting substance; causation by an agent is causation by such a substance. Since a substance is not the kind of thing that can itself be an effect (though various events involving it can be), on these accounts an agent is in a strict and literal sense an originator of her free decisions, an uncaused cause of them. This combination of indeterminism and origination is thought to capture best the idea that, when we act freely, a plurality of alternatives is open to us and we determine, ourselves, which of these we pursue, and to secure the kind of freedom needed for moral responsibility.

While in some cases talk of an object or other particular entity's causing an effect can be seen as shorthand for talk of the object's having some feature that produced the effect (as when, e.g., 'the rock broke the window' is more specifically understood as 'the rock's having momentum M at the point of contact with the window caused the window's subsequent shattering'), the agent-causal libertarian denies that, for the outcomes of free choices, such a reduction (e.g., to the agent's recognizing certain reasons) is unavailable.

Concerns about agent causation seem to stem from concern about causation by substances rather than events. One might reasonably deny, however, that making sense of agent-causation requires either thinking of agents as 'substances', as opposed to a new kind of object, no more problematic, in principle, than other sorts of objects (molecules, cells, plants, animals) posited by the various special sciences, and which are also supposed to be capable of causing various effects.

No doubt such objects, like persons, cause these effects in virtue of certain of their properties, as per usual. But if the properties are along lines of ‘having transcendently free will’, then there is no threat to the supposition that agents are capable of causing the outcomes of their acts of free choosing.

In any case, in agent-causal accounts there is a clear sense in which free will involves the exercise of a fundamentally novel power. O’Connor (2009a) is explicit on this score, saying that

[A]n adequate account of freedom requires, in my judgment, a notion of a distinctive variety of causal power, one which tradition dubs “agent-causal power”. [...]

The familiar considerations [against locating free will in indeterminacy] lead certain philosophers to conclude that the kind of control necessary for freedom of action involves an ontologically primitive capacity of the agent directly to determine which of several alternative courses of action is realized.

While the tidiness of substance dualism has its appeal, it is in fact optional for the metaphysician who believes that human beings have ontologically fundamental powers (whether of freedom or consciousness or intentionality). For we may suppose that such powers are [...] ontologically emergent powers, ones that are at once causally dependent on microphysically-based structural states and yet ontologically primitive, and so apt to confer ontologically primitive causal power. (191)

What is the power at issue? As desired, the novel fundamental power is one which allows one to choose to act intend to act) in one way rather than another:

One important feature of agent-causal power is that it is not directed to any particular effects. Instead, it confers upon an agent a power to cause a certain type of event within the agent: the coming to be of a state of intention to carry out some act, thereby resolving a state of uncertainty about which action to undertake.

Supposing there is a power of agent causation has the virtue that it seems to avoid this ‘problem of luck’ facing other indeterministic accounts. Agent causation is precisely the power to directly determine which of several causal possibilities is realized on a given occasion.
 [...]

[T]he view posits a fundamental, irreducible power of agents to form intentions.

As above, the posit of such a novel fundamental power is supposed to provide a basis for responding to the luck objection, in virtue of positing “a kind of single-case form of control by means of which the agent can determine what happens in each case”. For present purposes, what is crucial is that O’Connor’s version of an agent-causal form of libertarianism clearly satisfies the conditions in the schema for Strong emergence.

8.3.3 Noncausal accounts

Clark and Capes (2017) describe noncausal libertarian accounts as follows:

Some incompatibilist accounts require neither that a free action be caused by anything nor that it have any internal causal structure. Some views of this type require that a free action be uncaused; others allow that it may be caused as long as it is not deterministically caused. Since any such account imposes no positive causal requirement on free action, we may call views of this type “noncausal”.

As O’Connor (2002) puts it, on a noncausal account, free will is taken to be “entirely noncausal in character and is instead a consequence of intrinsic, noncausal features of the choice itself”. There are several different conceptions of the intrinsic feature at issue. On Ginet’s version (see, e.g., Ginet 1990 and (2002)), this feature is an “actish phenomenal quality,” which he describes (1990, 13) as

its seeming to the agent as if she is directly producing, making happen, or determining the event that has this quality. On McCann's (1998) version, the intrinsic feature at issue is 'intrinsic intentionality'. As Clarke (2003) describes McCann's view:

[I]n making a decision, McCann maintains (1998: 163), one intends to decide—indeed, one intends to decide exactly as one does (e.g., when one decides to A, one intends to decide to A). One's so intending, though intrinsic to the decision, is not a matter of the content of the intention that is formed in deciding; nor is it a matter of one's having any further intention in addition to that formed in making the decision. Rather, McCann holds, it is a matter of a decision's being, by its very nature, an act that the agent means to be performing. (18)

Finally, Stump (1999) endorses a noncausal account along lines of a Thomistic account, which she describes as follows:

What is essential to moral responsibility on Aquinas's view is that a person be the ultimate source of what she does, that her intellect and will be the ultimate causes of her acts. By 'ultimate cause' here, I mean that there is nothing which is prior to that person's acts of intellect and will and which causally determines her intellect and will to be in the states in which they are. If we can trace a causal chain of any sort backward from an agent's act, then the causal chain must originate only in acts of her will and intellect. That is, for any act which the agent does, if there is any causal chain at all of which the act is the effect, then the causal chain must have a first or ultimate cause, and that ultimate cause cannot be anything other than an act of the agent's own will or intellect. (414)

I will now argue that, on plausible construals of each of these variations on the noncausal theme, the sense in which free will is 'noncausal' pertains only to the purported failure of free choices to themselves be effects of prior (deterministic or indeterministic) causes. This much is compatible, however, with free choosings on such an account being 'causal' in the sense of themselves having effects,

and associated powers. To prefigure: this result, coupled with the supposition that the powers themselves are fundamentally novel, as compared to powers of dependence base goings-on, will support taking free will on a noncausal account to conform to the conditions in the schema for Strong emergence.

Consider, to start, Ginet's account, according to which free will involves "an actish phenomenal quality". As Clarke (2003) observes, one might complain that such a conception is compatible with an agent's not really having any "active control":

Whatever the correct characterization of this phenomenal quality, the mere feel of a mental event—the way it seems to the individual undergoing it—although it may be a (more or less reliable) sign of active control, cannot itself constitute the agent's exercise of such control (cf. O'Connor 2000, 25–26). To hold that it does is to render the exercise of active control wholly subjective (nothing more than the way things seem), and this is to greatly diminish the significance of active control. (20)

Even if all this is right, however, and even if the outcomes of acts of choosing (and associated actish phenomenal qualities) are noncausal, in being uncaused, it would remain that such choosings (or associated actish qualities) might be causal in have powers to produce, or to contribute to producing, certain effects. Indeed, one power of an act of choosing might be a power to cause the very phenomenal actish quality itself. Such an effect would conform to the usual supposition of a noncausal account, according to which a choosing itself is not an effect of previous causal factors.

Next, consider McCann's view, according to which noncausal free will involves intrinsic intentionality, and more specifically, where an act of libertarian choosing is one which is characterized as such by the presence of an intention to

decide, as “a matter of a decision’s being, by its very nature, an act that the agent means to be performing”. Here again, and granting that both the act of choosing and the outcome of the choice are noncausal in not being appropriately seen as effects of prior causes, it remains that an act of choosing is causal precisely in that it involves the exercise of a power to choose, or decide.

Finally, consider Stump’s Thomistic view, according to which libertarian free will is noncausal in the sense that, when a person chooses, “there is nothing which is prior to that person’s acts of intellect and will and which causally determines her intellect and will to be in the states in which they are”. Here a given act of choosing may, after all, be caused—so long as those causes are other acts of intellect or will. Most importantly for present purposes, however, this Thomistic view is compatible with taking such acts of intellect or will to be causal in the sense of themselves having powers to cause or contribute to causing certain effects—most saliently, in a case of free choosing, the outcome of the act of choosing.

On a representative range of noncausal accounts of Libertarian free will, then, the characterization of acts of choosing as noncausal reflects that such acts are not the effects of causes prior to the acts. It remains that such acts are causal at least in having powers to cause, at a minimum, an outcome of the choosing, and moreover to cause (if Ginet is correct) a “phenomenal actish quality”, and moreover to contribute to causing any effects associated with the outcome of the choice (e.g., the reaching for a glass of water). This distinctive causal asymmetry—free choices are uncaused, but capable of causing—is reasonably interpreted as suggesting that on noncausal libertarian accounts, acts of free choosing involve new powers, as Strong emergence requires.

8.4 Is free will either Weakly or Strongly emergent?

I turn now to arguing considering whether there is actually any free will of either Weak or Strong emergent varieties.

8.4.1 Is there compatibilist (Weakly emergent) free will?

As above, the compatibilist is plausibly seen as implementing a proper subset strategy, on which the freedom of an act of choosing is ultimately a matter of the act's being associated with a proper subset of the complete and actual causal antecedents of the choosing; and as I argued above, this strategy can in turn be plausibly interpreted as involving certain complex events' having powers satisfying the conditions in the schema for Weak emergence. Given all this, what are the prospects for there actually being free will of the Weakly emergent variety?

The prospects are good. Though free choices are not taken to be part of a higher-level system of laws on either compatibilist or libertarian accounts, a compatibilist account is one manifesting the usual Weak emergentist characterization of special science goings-on as comparatively insensitive to lower-level physical details, in the sense that an agent's reasons for action in a given case float free of many such details (and in particular, are sensitive only to facts about 'relevant' causal antecedents). Since our deliberations and associated acts of choice clearly are insensitive to many microphysical details, then given that free will is understood along compatibilist (Weak emergentist) lines, there is good reason to think that such free will actually exists, and indeed is abundant. Indeed, even libertarians grant that there is actually free will of a compatibilist variety; they just think that's not the only, or most important, kind of free will there actually is. Hence

Clarke (2003) says:

We make decisions and act even if determinism is true; we are thus unlike puppets. And unlike agents who are not persons, we can still act on the basis of our appreciation of practical reasons, including moral reasons. We are also unlike prisoners, in that we can generally go where we want to go. Further, most of us most of the time act quite free from coercive threats and compulsive desires. We are never subject to the direct manipulation of our brains by malevolent neuroscientists. All of this is good, and these goods do not require indeterminism. There is, then, a valuable variety of active control that we can have and exercise even if determinism is true. However, this compatibilist variety of active control falls short of free will. (8)

8.4.2 Is there libertarian (Strongly emergent) free will?

Notwithstanding that there is presumably plenty of what the compatibilist counts as free will, I am inclined to agree with those, like Clark, who think that compatibilist accounts are ultimately unsuccessful in accommodating the core phenomenal and intentional aspects of free will, according to which a freely choosing agent feels, from the inside, to be causally determinative of the outcome in a way transcending any nomologically deterministic or indeterministic goings-on. Even those who aim to reject the appearances as genuine admit as much. Hence Caruso (2013), a determinist, says:

A major part of the folk psychology of free will is the belief that *our conscious intentions cause action*. As Patrick Haggard and Benjamin Libet write, “Most of us navigate through our daily lives with the belief that we have conscious free will: that is, we have conscious intentions to perform specific acts, and those intentions can drive our bodily actions, thus producing a desired change in the external world” (2001, 47). This commonsense intuition plays a major role in our sense of free will and is essential to the *up-to-me-ness* that we associate with free will. It is also well supported by phenomenology. In

normal cases of voluntary behavior, we experience a conscious intention before the onset of action and naturally take the former to be the cause of the latter. When I switch on my reading lamp, for example, I feel as though it is *I*, my conscious self, that controls the movements of my arms and hands through the conscious formation of goals, intentions, and decisions. (189)

Two points about this sort of pretheoretical ‘take’ on our capacity for free will. First, note that the phenomenal and intentional motivations for thinking that our choices are at least sometimes transcendentally free don’t turn on the choice at issue being directed at anything morally or otherwise substantive.¹⁴ I can experience myself as seemingly transcendentally free even if all that is at issue is whether or not to throw a piece of chalk into the air.¹⁵ As such, and even granting that investigating into exactly how free will (assuming it exists) intersects with notions such as reasons, character, and moral responsibility, for present purposes we can restrict our focus to what seems to me to be the prior question of whether we are in position to act in ways that transcend the nomological net. Second, note that this much is already enough to block the too-quick strategy of appealing to indeterministic quantum or other laws as a basis for such freedom. This strategy is often rejected on grounds that an appeal to indeterministic laws would render the outcomes of choosings subject to random processes rather than reasons; but for present purposes it is enough to reject this strategy to note that indeterministic laws are still laws, solidly within the net of nomological goings-on, and hence appeal to such laws cannot provide a realistic metaphysical basis for the seeming

¹⁴This observation is registered in views on which freedom is not itself sufficient for moral responsibility; see Clarke 1992.

¹⁵As such, I would disagree with Clarke’s (2003)’s claim that “For an agent to act with free will, she must be able to regard some considerations as reasons for action” (16), at least if the ‘reason’ has to contain more content than the choice in itself, as in ‘I hereby choose to throw the chalk in the air’.

experience of nomologically transcendent free will.

Now, if the appearance of transcendent free will is to be taken as genuine, free will so understood is not properly accommodated by taking the usual Weak emergentist approach to higher-level special scientific goings-on, for even if it correct (as the compatibilist maintains) that our acts of seemingly free choosing are insensitive to certain (antecedent) micro-level details, it is not this insensitivity that constitutes seemingly free choice, or the experience of such. Indeed, once it is appreciated that compatibilist accounts of free will are implementing a variation on the usual Weak emergentist theme, such that any token power associated with (a complex event having as a part) an act of choosing ends up being identical to a token power associated with a lower-level physical event—to be sure, a highly complex, temporally extended physical event, but no matter—the usual concerns with a compatibilist approach are thrown into high relief. It may be that agents on a compatibilist approach are distinctively efficacious, in ways that complex systems, ordinary objects, and other special science goings-on are distinctively efficacious, in having distinctive power profiles that track difference-making considerations (if I had been differently molecularly configured, I would still have chosen to go to Pasternak for brunch). But this form of distinctive efficacy is not of the transcendent variety that is core and crucial to our phenomenal and intentional experience of agency. Hence I am inclined to agree with Clarke's claim, above, that "[the] compatibilist variety of active control falls short of free will" (2003, 8).

Realistic accommodation of these core phenomenal and intentional aspects and the associated (nomologically) transcendent conception of free will thus requires that such free will involve a fundamentally novel power which, as Liber-

tarians standardly maintain, transcends any nomological net. Either “*a sui generis* agent-causal power (a primitive capacity of a person to form an executive intention to act in a certain way”, as discussed by O’Connor 2009b, or a *sui generis* noncausal variety of active power will do for purposes of such accommodation.

Is there any libertarian (Strongly emergent) free will, so transcendently understood? To start, as with other purported Strongly emergent phenomena, there is a case to be made that such phenomena admit of empirical confirmation (or disconfirmation), at least in principle, by attention to whether the phenomena involves any apparent violations of conservation laws. After all, even if libertarian free will involves a fundamentally novel power, and even if such a novel power transcends not just the nomological net of lower-level physical goings-on, but indeed any nomological net, so long as such a power is at least partly determinative of a given effect, it will presumably involve some transfer of energy or other conserved quantity; and if such a transfer involves some fundamentally novel interaction, then this would provide an in-principle means of empirically verifying the hypothesis that free will is Strongly emergent.

Independent of such a direct route to empirical verification, however, the fact that we have direct introspective access to the phenomenon of seemingly transcendent free will provides the basis for a new argument for there actually being Libertarian—Strongly emergent—free will:

1. We have experience ourselves as seeming to freely choose, in ways transcending any (deterministic or indeterministic) goings-on.
2. In the absence of good reasons to think that our experience of transcendent free will cannot be taken at face value, some form of libertarian Strong emergence about free will must be correct.

3. There are no good reasons to think that our experience of transcendent free will cannot be taken at face value.
- ∴ Some form of libertarian Strong emergence about free will must be correct.

The argument is valid, and premise 1 is clearly true: even hard determinists and compatibilists will agree that we *seem* to freely choose, in ways transcending any nomological net, on at least some occasions. Premise (2) also seems reasonable: if we have clear experience of some seeming phenomenon, we need good reason not to take that experience at face value. In what follows, then, I'll focus on defending premise (3) against certain salient empirical results purporting to show that the phenomenon of free will is in some sense an illusion.

The empirical case against libertarianism

Recent results in empirical neuroscience have frequently been seen as showing that seeming experiences of free will as transcendent cannot be taken at face value.

Hence [Caruso \(2013\)](#) continues:

Although phenomenology supports this commonsense belief [in transcendently free will], empirical evidence in neuroscience now seriously questions it. In fact, a growing number of theorists now conclude that *conscious will*—in the sense of consciously initiated action—is incompatible with the evidence of neuroscience [...] Much of the contemporary case for this conclusion is derived from the experimental work of Benjamin Libet and his colleagues. [...] My thesis will be that the empirical results from neuroscience do in fact reveal that conscious will is an illusion—at least in the cases we can currently study empirically. (189)

In what follows I focus on the ‘Libet cases’ which pose the most serious chal-

lenge to taking our seeming experience of transcendental free will at face value.¹⁶ These studies aim to compare the self-reported time of occurrence of a certain conscious choice to produce a certain physical behaviour, with the time of occurrence of certain unconscious brain states associated with the production of the behaviour. And the concern for transcendent free will is that the evidence has been interpreted as suggesting that the unconscious initiation of the physical behaviour occurs *prior* to the time of the supposed choosing, such that the supposition that our free choices are determinative of our physical actions is illusory. O'Connor (2009b) describes the original setup and interpretation of results as described in Libet 1999:

Libet devised a study in which people are asked to wiggle their finger within a short interval of time (thirty seconds or so). The experimenter instructs them to do so whenever they wish--though spontaneously, not by deciding the moment in advance. Throughout, they are to watch a special clock with a very fast-moving dial (a beam of light) and note its location at the precise moment at which they felt the "urge" or "wish" to move the finger. During the experiment, a device measures electrical activity on the agents scalp. Libet discovered that a steady increase in this activity (dubbed the "readiness potential," or RP) consistently preceded the time the agents cited as when they experienced the will to move. By averaging results over hundreds of experiments, Libet determined that the RP preceded the "experience of will" by an average of some 400 milliseconds, a significant interval in the context of neural activity. Libet and others concluded from

¹⁶See O'Connor (2009b) for discussion of a number of other empirical results that some have taken to problematize taking the seeming transcendence of free choice as genuine; these include, e.g., cases where subjects of induced behaviours confabulate instances of their agency in a post-hoc way, cases where the outcomes of supposedly free choices (about which index finger to move) are influenced by external stimulation of the subject's brain, cases of subjects who report a feeling of agency concerning distal outcomes subsequent to being instructed to have certain negative thoughts (what we might call 'voodoo' cases), and cases where subjects of seemingly voluntary action unaccompanied by any feeling of agency (as in 'alien hand syndrome'). O'Connor compellingly argues that none of these results pose a serious challenge to the libertarian supposition; I direct the interested reader to his discussion.

this result that “conscious will” is not the initiator of voluntary acting but instead a consequence of an unconscious physical process that also (and according to some hypotheses, independently) triggers the action. (176)

Do experimental studies of this sort really establish that transcendentally free will is an illusion? This question continues to be hotly debated, but so far as I can tell, the answer is clearly ‘no’. To start, as O’Connor notes, the setup is one where the agent has already decided to perform a specific action, and while the focus of the experiment is supposedly on the agent’s choice of *when* to perform that action occurs, both the antecedent decision and the restriction on timing render the setup sufficiently non-standard that one might reasonably deny taking the results to suggest anything general about the status of transcendental agency. Relatedly, O’Connor observes that the instructions to the subject to introspect and wait for an unplanned “urge” to occur would likely encourage a passive posture in which “having decided that one will move, one looks for the *urge* to do so in order to act upon it”, which in turn motivates an alternative interpretation of the results according to which “such a preformed intention to act upon the right internal “cue” initiates an unconscious process that promotes the occurrence (or perhaps *evolution*) of a conscious state of desire or intention that is not actively formed” (182).

Indeed, another study by Libet confirms that in such an experimental context, the subject nonetheless has ‘veto’ power over the antecedent ‘urge’ in question. As [Libet \(2002\)](#) himself observes:

[T]he conscious function still had enough time to affect the outcome of the process; that is, it could allow the volitional initiative to go to completion, it could provide a necessary trigger for the completion, or it could block or veto the process and prevent the acts appearance.

There is no doubt that a veto function can occur. The argument has been made that the conscious veto process would itself require preceding developmental processes, just like a conscious sensory awareness. But Libet (1999) argued that the conscious veto in a control function, different from awareness *per se*, need not be a direct product of the preceding processes, as is the case for simple awareness. (292)

As such, one might interpret the supposed antecedent brain activity in the original setup as simply setting in train a deliberative process having as its ‘default’ object the performance of a certain physical action, where the act of free choosing consists *either* in a decision to allow the process to continue to completion *or* in a decision to block the process. All of this is compatible with transcendent free will.¹⁷

Finally, yet another interpretive option remains on the table—namely, one according to which the intention to choose and the brain activity in fact are synchronically initiated, but where it takes just a bit of time for this fact to consciously register as a complete thought in the agent’s mind. Thinking takes time—more time, perhaps, than a choice. On this interpretation, a very small lag—less than half a second—would be a natural concomitant of our mental decision-making processes, again compatible with fully transcendent free will. More generally, it’s clear that Libet’s assumption that “In the traditional view of conscious will and free will, one would expect conscious will to appear before, or at the onset, of the RP [Readiness Potential], and thus command the brain to perform the intended act” (1999, 49) reflects an overly simplistic account of how transcendent free will would actually work.

¹⁷ Mele (2009) also argues that empirical considerations of the sort raised by Libet, Wegner, and others do not support any revisionary conclusions about free will.

8.5 Concluding remarks

I have argued here, first, that there is an important and theoretically powerful connection between debates over the metaphysics of higher-level phenomena and debates over the status of free will (i.e., acts of seemingly free choosing). Following Bernstein and Wilson 2016, these debates can each be seen as specific cases of a general ‘problem of mental quausation’—the problem of how a mental event (or any special science going-on, for that matter) can be efficacious *qua* the type of event it is, given the live possibility of certain theses which threaten to undermine, one way or another, the supposed causal relevance of the event. Drawing on this result, I have here argued that the non-reductive physicalist and compatibilist approaches are not just structurally similar, in each implementing ‘proper subset’ strategies for responding to their respective problematics, but also that compatibilists can plausibly be taken to be implementing the very same proper subset strategy encoded in the schema for Weak emergence, and hence as maintaining that free will is emergent in just this sense. I have similarly here argued that libertarian accounts of free will clearly suppose that free choosings are associated with fundamentally novel powers to bring about the outcomes in question, as per the schema for Strong emergence, such that libertarians are plausibly seen as maintaining that free will is emergent in just this sense.

Finally, I have considered whether free will is actually either Weakly or Strongly emergent. Free will on a compatibilist view, I have argued, is easy to come by, as another case-in-point of the usual Weak emergentist understanding of higher-level features as comparatively abstract or insensitive to lower-level physical details. This result is something of a double-edged sword, however, for unlike many other higher-level phenomena which seem amenable to Weak emergentist treatment,

including complex systems and ordinary objects of the sort appropriately treated by classical mechanics, such insensitivity to lower-level detail seems besides the point of generally accommodating free will, which in at least some manifestations appears to have more to do with an agent's ability to transcend any lower-level physical goings-on than to merely abstract from them. As I have argued, however, our experience of seemingly transcendent free will itself provides good reason to think that we have free will of a libertarian, Strongly emergent variety; and as I have argued, the neuroscientific reasons for rejecting this experience as illusory do not withstand scrutiny. Contrary to common assumption, it is libertarian free will, not subjective or qualitative experience, that provides the best case of a Strongly emergent phenomenon.

I conclude that there is actually free will of both Weak and Strong varieties—a nice result, given the importance of free will as a basis for personal and moral autonomy, and one which, to my mind, provides a fitting closing indication of the importance of metaphysical emergence to our understanding not just of the world, but of ourselves.

Bibliography

- Aizawa, Kenneth and Carl Gillett, 2009. “The (Multiple) Realization of Psychological and Other Properties in the Sciences”. *Mind and Language*, 24:181–208.
- Akiba, Ken, 2004. “Vagueness in the World”. *Noûs*, 38:407–429.
- Alexander, Samuel, 1920. *Space, Time, and Deity*. London: Macmillan.
- Alter, Torin, 1998. “A Limited Defense of the Knowledge Argument”. *Philosophical Studies*, 90:35–56.
- Anderson, P. W., 1972. “More is Different”. *Science*, 177:393–396.
- Antony, Louise M., 2003. “Who’s Afraid of Disjunctive Properties?” *Philosophical Issues*, 13:1–21.
- Antony, Louise M. and Joseph M. Levine, 1997. “Reduction with Autonomy”. *Philosophical Perspectives*, 11:83–105.
- Armstrong, David M., 1978. *Universals and Scientific Realism, Vol II: A Theory of Universals*. Cambridge: Cambridge University Press.
- Audi, Paul, 2012. “Grounding: Toward a Theory of the In-Virtue-of Relation”. *Journal of Philosophy*, 109:685–711.
- Auyang, Sunny, 1999. *How is Quantum Field Theory Possible?* Oxford: Oxford University Press.
- Ayer, Alfred J., 1954. “Freedom and Necessity”. In *Philosophical Essays*, 271–284. London: Macmillan, London and Basingstoke. Reprinted in [Watson 1982](#), p. 15–23.

- Bain, Alexander, 1870. *Logic, Book II & III*. London: Longman's, Green, Reader, and Dyer.
- Baker, Lynne Rudder, 1993. "Metaphysics and Mental Causation". In John Heil and Alfred Mele, editors, *Mental Causation*, 75–96. Oxford: Clarendon Press.
- Baltimore, Joseph A., 2013. "Careful, Physicalists: Mind–Body Supervenience Can Be Too Superduper". *Theoria*, 79:8–21.
- Barnes, Elizabeth, 2006. *Conceptual Room for Ontic Vagueness*. Ph.D. thesis, University of St. Andrews.
- Barnes, Elizabeth, 2010. "Ontic Vagueness: A Guide for the Perplexed". *Noûs*, 44:601–627.
- Barnes, Elizabeth, 2012. "Emergence and Fundamentality". *Mind*, 121:873–901.
- Barnes, Elizabeth and J. R. G. Williams, 2011. "A Theory of Metaphysical Indeterminacy". In Karen Bennett and Dean W. Zimmerman, editors, *Oxford Studies in Metaphysics volume 6*, 103–148. Oxford University Press.
- Batterman, Robert, 1998. "Why Equilibrium Statistical Mechanics Works: Universality and the Renormalization Group". *Philosophy of Science*, 65:183–208.
- Batterman, Robert, 2005. "Critical Phenomena and Breaking Drops: Infinite Idealizations in Physics". *Studies in History and Philosophy of Science Part B*, 36:225–244.
- Batterman, Robert W., 2000. "Multiple Realizability and Universality". *British Journal for the Philosophy of Science*, 51:115–145.
- Batterman, Robert W., 2002. *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford: Oxford University Press.
- Baysan, Umut, 2014. *Realization and Causal Powers*. Ph.D. thesis, University of Glasgow.
- Baysan, Umut, 2016. "An Argument for Power Inheritance". *Philosophical Quarterly*, 383–390.
- Baysan, Umut and Jessica Wilson, 2017. "Must Strong Emergence Collapse?" *Philosophica*, 91:49–104.

- Bedau, Mark, 1997. “Weak Emergence”. *Philosophical Perspectives* 11: *Mind, Causation and World*, 11:375–399.
- Bedau, Mark A., 2002. “Downward causation and the autonomy of weak emergence”. *Principia*, 6:5–50.
- Bedau, Mark A., 2008. “Is weak emergence just in the mind?” *Minds and Machines*, 18:443–459.
- Bedau, Mark A., 2010. “Weak Emergence and Context-Sensitive Reduction”. In Antonella Corradini and Timothy O’Connor, editors, *Emergence in Science and Philosophy*, 6–46. Routledge.
- Bennett, Karen, 2015. ““Perfectly Understood, Unproblematic, and Certain”: Lewis on Mereology”.
- Bennett, Karen, 2017. *Making Things Up*. Oxford: Oxford University Press.
- Berkeley, George, 1710. “A Treatise Concerning the Principles of Human Knowledge”. In A. A. Luce and T. E. Jessop, editors, *The Works of George Berkeley*, volume 2. London: Thomas Nelson and Sons.
- Bernstein, Sara and Jessica M. Wilson, 2016. “Free Will and Mental Quausation”. *Journal of the American Philosophical Association*, 2:310–331.
- Biggs, Stephen and Jessica M. Wilson, 2016a. “Carnap, the Necessary A Posteriori, and Metaphysical Anti-realism”. In Stephen Blatti and Sandra LaPointe, editors, *Ontology After Carnap*, 81–104.
- Biggs, Stephen and Jessica M. Wilson, 2016b. “A Priority of Abduction”.
- Biggs, Stephen and Jessica M. Wilson, 2017. “Abductive Two-Dimensionalism: A New Route to the A Priori Identification of Necessary Truths”. *Synthese*.
- Biggs, Stephen and Jessica M. Wilson, in progress. “Abduction vs. Conceiving in the Epistemology of Modality”.
- Bird, Alexander, 2001. “Necessarily, Salt Dissolves in Water”. *Analysis*, 61:267–274.
- Bird, Alexander, 2002. “On Whether Some Laws are Necessary”. *Analysis*, 62:257–270.

- Bird, Alexander, 2007. *Nature's Metaphysics: Laws and Properties*. Oxford: Oxford University Press.
- Bliss, Ricki and Kelly Trogdon, 2014. "Metaphysical Grounding". *Stanford Encyclopedia of Philosophy (Winter 2014 edition)*.
- Block, Ned and Robert Stalnaker, 1999. "Conceptual analysis, dualism, and the explanatory gap". *Philosophical Review*, 108:1–46.
- Boghossian, Paul Artin, 1996. "Analyticity Reconsidered". *Noûs*, 30:360–391.
- Bohm, David, 1952. "A Suggested Interpretation of Quantum Theory in Terms of "Hidden Variables", Parts I and II". *The Physical Review*, 85:166–179, 180–193.
- Bokulich, Alisa, 2014. "Metaphysical Indeterminacy, Properties, and Quantum Theory". *Res Philosophica*, 91:449–475.
- Boyd, Richard, 1980. "Materialism without Reduction: What Physicalism does Not Entail". In Ned Block, editor, *Readings in the Philosophy of Psychology*, volume 1, 67–106. Cambridge: Harvard University Press.
- Broad, C. D., 1925. *Mind and Its Place in Nature*. Cambridge: Kegan Paul. From the 1923 Tanner Lectures at Cambridge.
- Byrne, Alex, 1994. *The Emergent Mind*. Ph.D. thesis, Princeton University, Princeton, NJ.
- Byrne, Alex, 2001. "Intentionalism Defended". *Philosophical Review*, 110:199–240.
- Byrne, Alex, 2016. "Inverted Qualia". *The Stanford Encyclopedia of Philosophy (Summer 2016 Edition)*. URL = <http://plato.stanford.edu/archives/sum2016/entries/qualia-inverted/>.
- Calosi, Claudio and Jessica M. Wilson, forthcoming. "Quantum Metaphysical Indeterminacy".
- Camazine, Scott, Jean-Louis Deneubourg, Nigel R. Franks, James Sneyd, Guy Theraulaz, and Eric Bonabeau, 2001. *Self-Organization in Biological Systems*. Princeton, NJ: Princeton University Press.

- Campbell, Keith, 1990. *Abstract Particulars*. Oxford: Blackwell.
- Campbell, Richard and Mark H. Bickhard, 2011. “Physicalism, Emergence and Downward Causation”. *Axiomathes*, 21:33–56.
- Carere, C., S. Montanino, F. Moreschini, F. Zoratto, F. Chiarotti, and D. Santucci, 2009. “Aerial Flocking Patterns of Wintering Starlings, *Sturnus vulgaris*, under different predation risk”. *Animal Behavior*, 77:101–107.
- Caruso, Gregg, 2013. *Free Will and Consciousness: A Determinist Account of the Illusion of Free Will*. Lexington Books.
- Chalmers, David, 1996. *The Conscious Mind*. Oxford: Oxford University Press.
- Chalmers, David, 1999. “Materialism and the Metaphysics of Modality”. *Philosophy and Phenomenological Research*, LIX:473–496.
- Chalmers, David, 2012. *Constructing the World*. Oxford University Press.
- Chalmers, David and Frank Jackson, 2001. “Conceptual Analysis and Reductive Explanation”. *The Philosophical Review*, 110:315–60.
- Chalmers, David J., 2003. “Consciousness and its Place in Nature”. In Stephen P. Stich and Ted A. Warfield, editors, *Blackwell Guide to the Philosophy of Mind*, 102–142. Blackwell.
- Chalmers, David J., 2004. “The Representational Character of Experience”. In Brian Leiter, editor, *The Future for Philosophy*, 153–181. Oxford University Press.
- Chalmers, David J., 2006a. “Strong and weak emergence”. In *The Re-Emergence of Emergence*. Oxford University Press.
- Chalmers, David J., 2006b. “The Foundations of Two-Dimensional Semantics”. In Manuel Garcia-Carpintero and Josep Macia, editors, *Two-Dimensional Semantics: Foundations and Applications*. Oxford University Press.
- Chalmers, David J., 2009. “The Two-Dimensional Argument Against Materialism”. In Brian P. McLaughlin and Sven Walter, editors, *Oxford Handbook to the Philosophy of Mind*. Oxford University Press.

- Chomsky, Noam, 1968. *Language and Mind*. New York: Harcourt Brace and World.
- Churchland, Patricia S., 1986. *Neurophilosophy: Toward A Unified Science of the Mind-Brain*. MIT Press.
- Churchland, Paul M., 1981. “Eliminative materialism and the propositional attitudes”. *Journal of Philosophy*, 78:67–90.
- Churchland, Paul M., 1984. *Matter and Consciousness*. MIT Press.
- Clapp, Lenny, 2001. “Disjunctive Properties: Multiple Realizations”. *Journal of Philosophy*, 98:111–136.
- Clark, Randolph and Justin Capes, 2017. “Incompatibilist (Nondeterministic) Theories of Free Will”. *The Stanford Encyclopedia of Philosophy (Spring 2017 Edition)*. [<https://plato.stanford.edu/archives/spr2017/entries/incompatibilism-theories/>].
- Clarke, Randolph, 1992. “Free Will and the Conditions of Moral Responsibility”. *Philosophical Studies*, 66:53–72.
- Clarke, Randolph, 1999. “Nonreductive Physicalism and the Causal Powers of the Mental”. *Synthese*, 51:295–322.
- Clarke, Randolph, 2003. *Libertarian Accounts of Free Will*. Oxford University Press USA.
- Clayton, Philip, 2006. “Conceptual foundations of emergence theory”. In *The Re-Emergence of Emergence*, 1–31. Oxford University Press.
- Cohen, Stewart, 2010. “Bootstrapping, Defeasible Reasoning, and a Priori Justification”. *Philosophical Perspectives*, 24:141–159.
- Cohen-Tannoudji, Claude, Bernard Diu, and Franck Laloë, 1977. *Quantum Mechanics*, volume 1. France: John Wiley & Sons, 2nd edition.
- Conee, Earl, 1994. “Phenomenal Knowledge”. *Australasian Journal of Philosophy*, 72:136–150.
- Cotnoir, Aaron J., 2013. “Composition as General Identity”. *Oxford Studies in Metaphysics*, 8:294.

- Couzin, I. D., 2007. “Collective Minds”. *Nature*, 4.
- Crane, Tim, 2001. “The Significance of Emergence”. In Carl Gillett and Barry Loewer, editors, *Physicalism and Its Discontents*, 207–224. Cambridge: Cambridge University Press.
- Crane, Tim and D. H. Mellor, 1990. “There is No Question of Physicalism”. *Mind*, 99:185–206.
- Craver, Carl F., 2001. “Role Functions, Mechanisms, and Hierarchy”. *Philosophy of Science*, 68:53–74.
- Crook, Seth and Carl Gillett, 2001. “Why Physics Alone Cannot Define the ‘Physical’: Materialism, Metaphysics, and the Formulation of Physicalism”. *Canadian Journal of Philosophy*, 31:333–360.
- Cunningham, Bryon, 2001. “The Reemergence of Emergence”. *Philosophy of Science*, 68:S62–S75. Supplement: Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association.
- Dasgupta, Shamik, 2014. “The Possibility of Physicalism”. *Journal of Philosophy*, 111:557–592.
- Davidson, Donald, 1963. “Actions, Reasons, and Causes”. *Journal of Philosophy*, 60:685–700.
- Davidson, Donald, 1970. “Mental Events”. In L. Foster and J. Swanson, editors, *Experience and Theory*. Amherst: Massachusetts University Press. Reprinted in Davidson 1980/2001.
- Davidson, Donald, 1980/2001. *Essays on Actions and Events*. Oxford: Oxford University Press.
- Descartes, René, 1641–7/1984. “Meditations”. In Robert Stoothoff John Cottingham and Dugald Murdoch, editors, *The Philosophical Writings of Descartes*, volume II. Cambridge: Cambridge University Press.
- Dosanjh, Ranpal, 2014. *A Defense of Reductive Physicalism*. Ph.D. thesis, University of Toronto, Toronto, Canada.
- Dowell, Janice, 2006. “Formulating the Thesis of Physicalism: An Introduction”. *Philosophical Studies*, 131:1–23.

- Dretske, Fred I., 1995. *Naturalizing the Mind*. Cambridge, MA: The MIT Press.
- Dummett, Michael, 1975. “Wang’s Paradox”. *Synthese*, 30:201–32.
- Ehring, Douglas, 1996. “Mental Causation, Determinables, and Property Instances”. *Nous*, 30:461–480.
- Ehring, Douglas E., 2003. “Part-whole Physicalism and Mental Causation”. *Synthese*, 136:359–388.
- Ekstrom, Laura W., 2001. *Agency and Responsibility: Essays on the Metaphysics of Freedom*. Westview.
- Ellis, Brian, 2001. *Scientific Essentialism*. Cambridge: Cambridge University Press.
- Evans, Gareth, 1983. “Can There be Vague Objects?” *Analysis*, 38:208.
- Ewing, Alfred C., 1951. *The Fundamental Questions of Philosophy*. Routledge.
- Fales, Evan, 1990. *Causation and Universals*. Routledge.
- Farrell, B. A., 1950. “Experience”. *Mind*, 59:170–98.
- Feigl, Herbert, 1959. “The ‘Mental’ and the ‘Physical’”. In Grover Maxwell Herbert Feigl and Michael Scriven, editors, *Minnesota Studies in the Philosophy of Science*, volume 2. Minneapolis: University of Minnesota Press.
- Feigl, Herbert, 1967. *The "Mental" and the "Physical" the Essay and a Postscript*. University of Minnesota Press.
- Feyerabend, Paul K., 1963. “Mental Events and the Brain”. *Journal of Philosophy*, 40:295–6.
- Feynman, Richard, 1963. *The Feynman Lectures on Physics*, volume 1. Boston: Addison Wesley.
- Feynman, Richard, 1965. *The Character of Physical Law*. Cambridge, MA: MIT Press.
- Field, Hartry, 2003. “Causation in a Physical World”. In Michael Loux and Dean Zimmerman, editors, *The Oxford Handbook of Metaphysics*. Oxford: Oxford.

- Fine, Kit, 2001. "The Question of Realism". *Philosophers' Imprint*, 1:1–30.
- Fine, Kit, 2003. "The Non-Identity of a Material Thing and its Matter". *Mind*, 112:195–234.
- Flanagan, Owen, 1992. *Consciousness Reconsidered*. Cambridge, MA: The MIT Press.
- Fodor, Jerry, 1974. "Special Sciences (Or, The Disunity of Science as a Working Hypothesis)". *Synthese*, 28:77–115.
- Fodor, Jerry, 1987. *Psychosemantics*. Cambridge: MIT Press.
- Forster, Malcolm and Alexey Kryukov, 2003. "The Emergence of the Macroworld: A Study of Intertheory Relations in Classical and Quantum Mechanics". *Philosophy of Science*, 70:1039–1051.
- Francescotti, Robert, 2007. "Emergence". *Erkenntnis*, 67:47–63.
- Frankfurt, Harry G., 1971. "Freedom of the Will and the Concept of a Person". *The Journal of Philosophy*, 68:5–20.
- Franklin, Christopher Evan, 2018. *A Minimal Libertarianism: Free Will and the Promise of Reduction*. New York, USA: Oxford University Press.
- Frigg, Roman, 2009. "GRW Theory: Ghirardi, Rimini, Weber Model of Quantum Mechanics". In Klaus Hentschel Daniel Greenberger, Brigitte Falkenburg and Friedel Weinert, editors, *Compendium of Quantum Physics: Concepts, Experiments, History and Philosophy*. Berlin: Springer.
- Funkhouser, Eric, 2006. "The Determinable-Determinate Relation". *Nous*, 40:548–569.
- Garcia, Robert K., 2014. "Closing in on Causal Closure". *Journal of Consciousness Studies*, 21:96–109.
- Gertler, Brie, 2002. "Explanatory Reduction, Conceptual Analysis, and Conceivability Arguments About the Mind". *Noûs*, 36:22–49.
- Gertler, Brie, 2006. "Consciousness and Qualia Cannot Be Reduced". In Robert J. Stainton, editor, *Contemporary Debates in Cognitive Science*, 202–216. Blackwell.

- Ghirardi, G., A. Rimini, and T. Weber, 1986. “Unified Dynamics for Microscopic and Macroscopic Systems”. *The Physical Review*, D 34:470–479.
- Gibb, Sophie C., 2013. “The Entailment Problem and the Subset Account of Property Realization”. *Australasian Journal of Philosophy*, 92:551–566.
- Gillett, Carl, 2002a. “The Dimensions of Realization: A Critique of the Standard View”. *Analysis*, 62:316–323.
- Gillett, Carl, 2002b. “The Varieties of Emergence: Their Purposes, Obligations and Importance”. *Grazer Philosophische Studien*, 65:95–121.
- Gillett, Carl, 2010. “Moving Beyond the Subset Model of Realization: The Problem of Qualitative Distinctness in the Metaphysics of Science”. *Synthese*, 177:165–192.
- Gillett, Carl and Barry Loewer, editors, 2001. *Physicalism and Its Discontents*. Cambridge: Cambridge University Press.
- Ginet, Carl, 1990. *On Action*. Cambridge: Cambridge University Press.
- Ginet, Carl, 2002. “Reasons Explanations of Action: Causal Versus Noncausal Accounts”. In Robert H. Kane, editor, *The Oxford Handbook on Free Will*, 386–405. Oxford University Press.
- Goldwater, Jonah P. B., 2015. “No Composition, No Problem: Ordinary Objects as Arrangements”. *Philosophia*, 43:367–379.
- Grandy, Richard, 2016. “Sortals”. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Winter 2016 edition.
- Grandy, Richard E., 2007. “Artifacts: Parts and Principles”. In Eric Margolis and Stephen Laurence, editors, *Creations of the Mind: Theories of Artifacts and Their Representation*, 18–32. Oxford University Press.
- Haggard, Patrick and Benjamin W. Libet, 2001. “Conscious Intention and Brain Activity”. *Journal of Consciousness Studies*, 8:47–63.
- Hall, Ned, 2004. “Two Concepts of Causation”. In John Collins, Ned Hall, and Laurie Paul, editors, *Causation and Counterfactuals*, 225–276. The Mit Press.

- Harman, Gilbert, 1990. “The Intrinsic Quality of Experience”. In James Tomberlin, editor, *Action Theory and the Philosophy of Mind*, volume 4 of *Philosophical Perspectives*, 31–52. Atascadero: Ridgeview. Reprinted in Harman 1999.
- Harman, Gilbert, 1999. *Reasoning, Meaning, and Mind*. Oxford: Oxford University Press.
- Haug, Matthew C., 2010. “Realization, Determination, and Mechanisms”. *Philosophical Studies*, 150:313–330.
- Hawthorne, John, 2001. “Causal Structuralism”. *Nous*, 35:361–378.
- Hawthorne, John, 2002. “Deeply Contingent a Priori Knowledge”. *Philosophy and Phenomenological Research*, 65:247–269.
- Hawthorne, John and Philip Pettit, 1996. “Strategies for Free Will Compatibilists”. *Analysis*, 56.4:191–201. Hawthorne originally published this article under the name ‘John O’Leary-Hawthorne’.
- Heil, John, 1992. *The Nature of True Minds*. Cambridge: Cambridge University Press.
- Heil, John, 2003a. *From an Ontological Point of View*. Oxford: Clarendon Press.
- Heil, John, 2003b. “Levels of reality”. *Ratio*, 16:205–221.
- Hellie, Benj, 2014. “Love in the Time of Cholera”. In Berit Brogaard, editor, *Does Perception Have Content?*, 241–261. Oxford UP.
- Hellman, Geoffrey, 1985. “Determination and Logical Truth”. *The Journal of Philosophy*, 82:607–616.
- Hellman, Geoffrey and Frank Thompson, 1975. “Physicalism: Ontology, Determination, and Reduction”. *The Journal of Philosophy*, 72:551–564.
- Hempel, Carl, 1979. “Comment at a Symposium on Nelson Goodman’s *Ways of Worldmaking*”. In *Presented at the 76th Annual Meeting of the American Philosophical Association*.
- Hempel, Carl G. and Paul Oppenheim, 1948. “Studies in the Logic of Explanation”. *Philosophy of Science*, 15:135–175.

- Hodgeson, Shadworth, 1962. *The Theory of Practice: An Ethical Enquiry*. London: Longmans, Green, Reader, and Dyer.
- Hooker, C. A., 2004. “Asymptotics, Reduction and Emergence”. *British Journal for the Philosophy of Science*, 55:435–479.
- Horgan, Terence, 1982. “Supervenience and Microphysics”. *Pacific Philosophical Quarterly*, 63:29–43.
- Horgan, Terence, 1989. “Mental Quausation”. *Philosophical Perspectives* 3: *Philosophy of Mind and Action Theory*, 47–76.
- Horgan, Terry, 1993. “From Supervenience to Superdupervenience: Meeting the Demands of a Material World”. *Mind*, 102:555–586.
- Howell, Robert J., 2009. “Emergentism and Supervenience Physicalism”. *Australasian Journal of Philosophy*, 87:83–98.
- Humphreys, Paul, 1996. “Aspects of Emergence”. *Philosophical Topics*, 24:53–70.
- Humphreys, Paul, 1997. “How Properties Emerge”. *Philosophy of Science*, 64:1–17.
- Huxley, Thomas, 1874. “On the Hypothesis That Animals Are Automata, and its History”. *Fortnightly Review*, 95:555–80.
- Jackson, Frank, 1982. “Epiphenomenal Qualia”. *Philosophical Quarterly*, 32:127–136.
- Jackson, Frank, 1986. “What Mary Didn’t Know”. *The Journal of Philosophy*, 83:291–295.
- Jackson, Frank, 1994. “Armchair Metaphysics”. In John O’Leary-Hawthorne and Michaelis Michael, editors, *Philosophy in Mind*, 23–42. Kluwer.
- James, William, 1950/1890. *The Principles of Psychology*. New York, NY: Dover Publications.
- Jaworski, William, 2002. “Multiple-Realizability, Explanation and the Disjunctive Move”. *Philosophical Studies*, 108:289–308.

- Judd, D. and G. Wyszecki, 1975. *Color in Business, Science, and Industry (3rd. edition)*. New York: Wiley.
- Kane, Gordon, 1993. *Modern Elementary Particle Physics: the Fundamental Particles and Forces*. Boulder: Westview Press.
- Kauffman, Stuart A., 1993. *The Origins of Order Self-Organization and Selection in Evolution*. Oxford University Press.
- Kauffman, Stuart A., 1995. *At Home in the Universe the Search for Laws of Self-Organization and Complexity*.
- Kim, Jaegwon, 1984. “Concepts of Supervenience”. *Philosophy and Phenomenological Research*, 45:153–76.
- Kim, Jaegwon, 1989. “The Myth of Nonreductive Materialism”. *Proceedings and Addresses of the American Philosophical Association*, 63:31–47. Reprinted in Kim 1993b.
- Kim, Jaegwon, 1990. “Supervenience as a Philosophical Concept”. *Metaphilosophy*, 21:1–27. Reprinted in Kim 1993b.
- Kim, Jaegwon, 1992a. “”Downward Causation” in Emergentism and Nonreductive Physicalism”. In Ansgar Beckermann, Hans Flohr, and Jaegwon Kim, editors, *Emergence or Reduction?: Prospects for Nonreductive Physicalism*, 119–138. De Gruyter.
- Kim, Jaegwon, 1992b. “Multiple Realization and the Metaphysics of Reduction”. *Philosophy and Phenomenological Research*, 52:1–26. Reprinted in Kim 1993b.
- Kim, Jaegwon, 1993a. “The Non-Reductivist’s Troubles with Mental Causation”. In John Heil and Alfred Mele, editors, *Mental Causation*, 189–210. Oxford: Oxford University Press. Reprinted in Kim 1993b.
- Kim, Jaegwon, 1993b. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.
- Kim, Jaegwon, 1998. *Mind in a Physical World*. Cambridge: MIT Press.
- Kim, Jaegwon, 1999. “Making Sense of Emergence”. *Philosophical Studies*, 95:3–36.

- Kim, Jaegwon, 2006. “Emergence: Core ideas and issues”. *Synthese*, 151:547–559.
- Kim, Jaegwon, 2010. “Thoughts on Sydney Shoemaker’s Physical Realization”. *Philosophical Studies*, 148:101–112.
- Kim, Jaegwon, 2015. “What Could Pair a Nonphysical Soul to a Physical Body?” In Keith Augustine and Michael Martin, editors, *The Myth of an Afterlife: The Case against Life After Death*, 335–347. Rowman & Littlefield.
- Kirk, Robert, 2015. “Zombies”. *Stanford Encyclopedia of Philosophy* (Summer 2015 Edition). URL = <http://plato.stanford.edu/archives/sum2015/entries/zombies/>.
- Kitcher, P. S., 1984. “In Defense of Intentional Psychology”. *Journal of Philosophy*, 81:89–106.
- Klee, Robert, 1984. “Micro-Determinism and Concepts of Emergence”. *Philosophy of Science*, 51:44–63.
- Korman, Daniel Z., 2011. “Ordinary Objects”. *Stanford Encyclopedia of Philosophy*.
- Koslicki, Kathrin, 2008. *The Structure of Objects*. Oxford University Press.
- Koslicki, Kathrin, 2012. “Varieties of Ontological Dependence”. In Fabrice Correia and Benjamin Schnieder, editors, *Metaphysical Grounding: Understanding the Structure of Reality*, 186. Cambridge University Press.
- Koslicki, Kathrin, 2016. “Where Grounding and Causation Part Ways: Comments on Jonathan Schaffer”. *Philosophical Studies*, 101–112.
- Kripke, Saul, 1972/80. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Ladyman, James and Don Ross, 2007. *Every Thing Must Go: Metaphysics Naturalized*. Oxford University Press.
- Lamb, Maurice, 2015. *Characteristics of Non-reductive Explanations in Complex Dynamical Systems Research*. Ph.D. thesis, University of Cincinnati.
- Leibniz, Gottfried Wilhelm, 1714. *The Monadology*.

- LePore, Ernest and Barry Loewer, 1987. "Mind Matters". *The Journal of Philosophy*, 84:630–642.
- LePore, Ernest and Barry M. Loewer, 1989. "More on Making Mind Matter". *Philosophical Topics*, 17:175–91.
- Leuenberger, Stephan, in progress. "The Possibility of Emergence".
- Levine, Joseph, 1983. "Materialism and Qualia: The Explanatory Gap". *Pacific Philosophical Quarterly*, 64:354–361.
- Levine, Joseph, 2001. *Purple Haze: The Puzzle of Consciousness*. New York: Oxford University Press.
- Lewes, G. H., 1875. *Problems of Life and Mind*. London: Kegan Paul, Trench, Turbner & Co.
- Lewis, David, 1966. "An Argument for the Identity Theory". *The Journal of Philosophy*, 63:17–25. Reprinted in [Lewis 1983b](#).
- Lewis, David, 1981. "Are We Free to Break the Laws?" *Theoria*, 47:113–21.
- Lewis, David, 1983a. "New Work for a Theory of Universals". *Australasian journal of Philosophy*, 61:343–377.
- Lewis, David, 1983b. *Philosophical Papers*, volume i. Oxford: Oxford University Press.
- Lewis, David, 1986. *On the Plurality of Worlds*. London: Blackwell.
- Lewis, David, 1988. "What Experience Teaches". *Proceedings of the Russellian Society*, 29–57. Reprinted in [Lewis 1999](#).
- Lewis, David, 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
- Libet, Benjamin, 2002. "The Timing of Mental Events: Libet's Experimental Findings and Their Implications". *Consciousness and Cognition* 1, 11:291–299.
- Libet, Benjamin W., 1999. "Do We Have Free Will?" *Journal of Consciousness Studies*, 6:47–57.

- Locke, John, 1690. *An Essay Concerning Human Understanding*.
- Loewer, Barry, 2001. “From Physics to Physicalism”. In Carl Gillett and Barry Loewer, editors, *Physicalism and Its Discontents*, 37–56. Cambridge: Cambridge University Press.
- Lowe, E. J., 1989. “What is a Criterion of Identity?” *Philosophical Quarterly*, 39:1–21.
- Lowe, E. J., 1998. *The Possibility of Metaphysics: Substance, Identity, and Time*. Oxford University Press.
- Lowe, E. J., 2007. “Sortals and the Individuation of Objects”. *Mind and Language*, 22:514–533.
- MacBride, Fraser, 1999. “Could Armstrong have been a Universal?” *Mind*, 108:471–501.
- Macdonald, C. and Graham F. Macdonald, 1995. “How to be Psychologically Relevant”. In *Philosophy of Psychology: Debates on Psychological Explanation*, update. Oxford University Press.
- MacDonald, Cynthia and Graham MacDonald, 1986. “Mental Causes and Explanation of Action”. In L. Stevenson, R. Squires, and J. Haldane, editors, *Mind, Causation, and Action*, update. Oxford: Basil Blackwell.
- Margolis, Eric and Stephen Laurence, editors, 2007. *Creations of the Mind: Theories of Artifacts and Their Representation*. Oxford University Press.
- Martin, C. B., 1996. “Properties and Dispositions”. In Tim Crane, editor, *Dispositions: A Debate*. London: Routledge.
- McCann, Hugh J., 1998. *The Works of Agency: On Human Action, Will, and Freedom*. Cornell University Press.
- McDaniel, Kris, 2001. “Tropes and Ordinary Physical Objects”. *Philosophical Studies*, 104:269–290.
- McKenna, Michael and Justin Coates, 2008. “Compatibilism”. In Edward Zalta, editor, *Stanford Encyclopedia of Philosophy*. URL = <http://plato.stanford.edu/archives/sum2015/entries/compatibilism/>.

- McLaughlin, Brian, 1992. “The Rise and Fall of British Emergentism”. In Ans-gar Beckerman, Hans Flohr, and Jaegwon Kim, editors, *Emergence or Reduction? Essays on the Prospects of Non-reductive Physicalism*, 49–93. Berlin: De Gruyter.
- McLaughlin, Brian and Karen Bennett, 2018. “Supervenience”. In Edward N. Zalta, editor, *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
- McLaughlin, Brian P., 2007. “Mental Causation and Shoemaker-Realization”. *Erkenntnis*, 67:149–172.
- Megill, Jason, 2013. “A Defense of Emergence”. *Axiomathes*, 23:597–615.
- Mele, Alfred, 2009. *Effective Intentions: The Power of Conscious Will*. Oxford University Press.
- Melnyk, Andrew, 1997. “How to Keep the ‘Physical’ in Physicalism”. *The Journal of Philosophy*, 94:622–637.
- Melnyk, Andrew, 1999. “Supercalifragilisticexpialidocious”. *NOÛS*, 33:144–54.
- Melnyk, Andrew, 2003. *A Physicalist Manifesto: Thoroughly Modern Materialism*. New York: Cambridge University Press.
- Melnyk, Andrew, 2006. “Realization-based Formulations of Physicalism”. *Philosophical Studies*, 131:127–155.
- Melnyk, Andrew, 2008. “Conceptual and linguistic analysis: A two-step program”. *Nous*, 42:267–291.
- Menon, Tarun and Craig Callender, 2013. “Ch-Ch-Changes Philosophical Questions Raised by Phase Transitions”. In Robert Batterman, editor, *The Oxford Handbook of Philosophy of Physics*, 189–212. OUP USA.
- Merricks, Trenton, 2003. *Objects and Persons*. Clarendon Press.
- Messiah, Albert, 1970. *Quantum Mechanics*. Amsterdam: North-Holland.
- Mill, John S., 1843/1973. *A System of Logic*. Toronto: University of Toronto Press. Vols II and III of *The Collected Works of John Stuart Mill*; also available at <http://www.gutenberg.org/files/27942/27942-pdf.pdf>.

- Mitchell, Sandra D., 2012. “Emergence: Logical, Functional and Dynamical”. *Synthese*, 185:171–186.
- Montero, Barbara, 2003. “Varieties of Causal Closure”. In Sven Walter and Heinz-Dieter Heckmann, editors, *Physicalism and Mental Causation*, 173–187. Imprint Academic.
- Montero, Barbara, 2006. “Physicalism in an Infinitely Decomposable World”. *Erkenntnis*, 64:177–191.
- Monton, Bradley, 2004. “The Problem of Ontology on Spontaneous Collapse Theories”. *Studies in History and Philosophy of Modern Physics*, 35:407–421.
- Moore, G. E., 1903. *Principia Ethica*. Cambridge: Cambridge University Press.
- Morgan, C. Lloyd, 1923. *Emergent Evolution*. London: Williams & Norgate.
- Morris, Kevin, 2010. “Guidelines for Theorizing About Realization”. *Southern Journal of Philosophy*, 48:393–416.
- Morris, Kevin, 2011a. “Subset Realization and Physical Identification”. *Canadian Journal of Philosophy*, 41:317–335.
- Morris, Kevin, 2011b. “Subset Realization, Parthood, and Causal Overdetermination”. *Pacific Philosophical Quarterly*, 92:363–379.
- Morris, Kevin, 2013. “On Two Arguments for Subset Inheritance”. *Philosophical Studies*, 163:197–211.
- Morris, Kevin, 2014. “Supervenience Physicalism, Emergentism, and the Polluted Supervenience Base”. *Erkenntnis*, 79:351–365.
- Morrison, Margaret, 2012. “Emergent Physics and Micro-Ontology”. *Philosophy of Science*, 79:141–166.
- Nagel, Ernest, 1961. *The Structure of Science*. London: Routledge & Kegan Paul.
- Nagel, Thomas, 1974. “What is it Like to be a Bat?” *The Philosophical Review*, 83:435–50.
- Nagel, Thomas, 1979. “Panpsychism”. In Thomas Nagel, editor, *Mortal Questions*. Cambridge University Press.

- Nemirow, Laurence, 2006. “So This is What It’s Like: A Defense of the Ability Hypothesis”. In Torin Alter and Sven Walter, editors, *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.
- Newman, David, 1996. “Emergence and Strange Attractors”. *Philosophy of Science*, 63:245–261.
- Ney, Alyssa, 2008. “Defining Physicalism”. *Philosophy Compass*, 3:1033–1048.
- Ney, Alyssa, 2010. “Convergence on the Problem of Mental Causation: Shoemaker’s Strategy for (Nonreductive?) Physicalists”. *Philosophical Issues*, 20:438–445.
- Nida-Rumelin, Martine, 2015. “Qualia: The Knowledge Argument”. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2015 edition.
- Noordhof, Paul, 2010. “Emergent Causation and Property Causation”. In Cynthia Macdonald and Graham Macdonald, editors, *Emergence in Mind*. Oup Oxford.
- Nozick, Robert, 1995. “Choice and Indeterminism”. In Timothy O’Connor, editor, *Agents, Causes, and Events: Essays on Indeterminism and Free Will*. Oxford University Press.
- O’Connor, Timothy, 1994. “Emergent properties”. *American Philosophical Quarterly*, 31:91–104.
- O’Connor, Timothy, 2000. *Persons and Causes: The Metaphysics of Free Will*. Oxford University Press USA.
- O’Connor, Timothy, 2002. “Free Will”. *The Stanford Encyclopedia of Philosophy (Summer 2016 Edition)*.
- O’Connor, Timothy, 2009a. “Agent-Causal Power”. In Toby Handfield, editor, *Dispositions and Causes*. Oxford University Press, Clarendon Press ;
- O’Connor, Timothy, 2009b. “Conscious Willing and the Emerging Sciences of Brain and Behavior”. In Nancey Murphy, George Ellis, and Timothy O’Connor, editors, *Downward Causation and the Neurobiology of Free Will*, 173–186. Springer Verlag.

- O'Connor, Timothy and Hong Yu Wong, 2005. "The Metaphysics of Emergence". *Nous*, 39:658–678.
- O'Connor, Timothy and Hong Yu Wong, 2015. "Emergent properties". *The Stanford Encyclopedia of Philosophy (Summer 2015 Edition)*.
- Owens, David, 1989. "Levels of Explanation". *Mind*, 98:59–79.
- Page, Robert E. and Sandra D. Mitchell, 1990a. "Self Organization and Adaptation in Insect Societies". *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1990:289–298.
- Page, Robert E. and Sandra D. Mitchell, 1990b. "Self organization and the evolution of division of labor". *Apidologie*, 29:101–120.
- Paoletti, Michele Paolini, 2017. *The Quest for Emergence*. Munich: Philosophia.
- Papineau, David, 1993. *Philosophical Naturalism*. Oxford: Basil Blackwell.
- Papineau, David, 2001. "The Rise of Physicalism". In Carl Gillett and Barry Loewer, editors, *Physicalism and Its Discontents*, 3–36. Cambridge: Cambridge University Press.
- Paul, L. A., 2002. "Logical Parts". *Nous*, 36:578–96.
- Paull, C. P. and T. R. Sider, 1992. "In Defense of Global Supervenience". *Philosophy and Phenomenological Research*, 32:830–845.
- Peacocke, Christopher, 1993. "How Are A Priori Truths Possible?" *European Journal of Philosophy*, 1:175–199.
- Pereboom, Derk, 2002. "Robust Non-reductive Materialism". *Journal of Philosophy*, 99:499–531.
- Pereboom, Derk, 2011. *Consciousness and the Prospects of Physicalism*. New York: Oxford University Press.
- Pereboom, Derk and Hilary Kornblith, 1991. "The Metaphysics of Irreducibility". *Philosophical Studies*, 63:125–45.
- Perry, John, 2001. *Knowledge, Possibility, and Consciousness*. Cambridge, MA: The MIT Press, second edition.

- Pettit, Philip, 1995. “Microphysicalism, Dottism, and Reduction”. *Analysis*, 55:141–146.
- Poland, Jeffrey, 1994. *Physicalism; the Philosophical Foundations*. Oxford: Clarendon Press.
- Polger, Thomas W., 2007. “Realization and the Metaphysics of Mind”. *Australasian Journal of Philosophy*, 85:233259.
- Popper, Karl R. and John C. Eccles, 1977. *The Self and Its Brain: An Argument for Interactionism*. Springer.
- Putnam, Hilary, 1967. “Psychological Predicates”. In *Art, Mind, and Religion*, 37–48. Pittsburgh: University of Pittsburgh Press.
- Raffman, Diana, 1994. “Vagueness Without Paradox”. *Philosophical Review*, 103:41–74.
- Raven, Michael, 2015. “Ground”. *Philosophy Compass*, 10:322–333.
- Ravenscroft, Ian, 1997. “Physical Properties”. *Southern Journal of Philosophy*, 35:419–431.
- Reutlinger, Alexander, 2017. “Are Causal Facts Really Explanatorily Emergent? Ladyman and Ross on Higher-Level Causal Facts and Renormalization Group Explanation”. *Synthese*, 2291–2305.
- Robb, David, 1997. “The Properties of Mental Causation”. *Philosophical Quarterly*, 47:178–94.
- Robinson, William, 2012. “Epiphenomenalism”. *The Stanford Encyclopedia of Philosophy*. URL = <http://plato.stanford.edu/archives/sum2012/entries/epiphenomenalism/>.
- Rosen, Gideon, 2010. “Metaphysical Dependence: Grounding and Reduction”. In B. Hale and A. Hoffmann, editors, *Modality: Metaphysics, Logic, and Epistemology*, 109–36. OUP.
- Rosen, J., 2007. “Flight patterns”. *New York Times*. April 22 edition.
- Rosenberg, Alexander, 1994. *Instrumental Biology, or, the Disunity of Science*. University of Chicago Press.

- Rueger, Alexander, 2001. “Physical emergence, diachronic and synchronic”. *Synthese*, 124:297–322.
- Russell, Bertrand, 1912. “On the Notion of Cause”. *Proceedings of the Aristotelian Society*, 13:1–26.
- Salmon, Nathan, 1989. “The logic of what might have been”. *Philosophical Review*, 98:3–34.
- Schaffer, Jonathan, 2003. “Is There a Fundamental Level?” *Noûs*, 37:498–517.
- Schaffer, Jonathan, 2004. “Quiddistic Knowledge”. In Frank Jackson and Graham Priest, editors, *Lewisian Themes*, 210–230. Oxford: Oxford University Press.
- Schaffer, Jonathan, 2009. “On What Grounds What”. In D. Manley, D. Chalmers, and R. Wasserman, editors, *Metametaphysics: New Essays on the Foundations of Ontology*, 347–383. OUP.
- Schaffer, Jonathan, 2010. “Monism: The Priority of the Whole”. *Philosophical Review*, 119:31–76.
- Schroder, Jurgen, 1998. “Emergence: Non-Deducibility or Downward Causation?” *The Philosophical Quarterly*, 48:433–452.
- Searle, John R., 1992. *The Rediscovery of the Mind*. MIT Press.
- Searle, John R., 2001. “Evolution and Progress in Democracies: Towards New Foundations of a Knowledge Society”. volume 31, 75–86. Springer Netherlands. Reprinted in Margolis and Laurence 2007.
- Shoemaker, Sydney, 1980. “Causality and Properties”. In Peter van Inwagen, editor, *Time and Cause*, 109–35. Dordrecht: D. Reidel.
- Shoemaker, Sydney, 1998. “Causal and Metaphysical Necessity”. *Pacific Philosophical Quarterly*, 79:59–77.
- Shoemaker, Sydney, 2000/2001. “Realization and Mental Causation”. In *Proceedings of the 20th World Congress in Philosophy*, 23–33. Cambridge: Philosophy Documentation Center. Published in revised form in Gillett and Loewer 2001, 74–98.

- Shoemaker, Sydney, 2003. “Realization, Micro-Realization, and Coincidence”. *Philosophy and Phenomenological Research*, 67:1–23.
- Shoemaker, Sydney, 2007. *Physical Realization*. Oxford University Press.
- Sider, Theodore, 2011. *Writing the Book of the World*. Oxford University Press.
- Silberstein, Michael, 2009. “Emergence”. In Tim Bayne, Axel Cleeremans, and Patrick Wilken, editors, *The Oxford Companion to Consciousness*, volume 50, 254–257.
- Silberstein, Michael and J. McGeever, 1999. “The search for ontological emergence”. *Philosophical Quarterly*, 50:182–200.
- Simons, Peter, 1994. “Particulars in Particular Clothing: Three Trope Theories of Substance”. *Philosophy and Phenomenological Research*, 54:553–575.
- Simons, Peter M., 1987. *Parts: A Study in Ontology*, volume 100. Oxford University Press.
- Smart, J. J. C., 1958. “Sensations and Brain Processes”. *The Philosophical Review*, 68:141–156.
- Smart, J. J. C., 1981. “Physicalism and emergence”. *Neuroscience*, 6:109–13.
- Sperry, Roger, 1986. “Discussion: Macro- Versus Micro-Determination”. *Philosophy of Science*, 265–270.
- Stalnaker, Robert, 1996. “Varieties of Supervenience”. *Philosophical Perspectives*, 10:221–42.
- Stephan, Achim, 2002. “Emergentism, irreducibility, and downward causation”. *Grazer Philosophische Studien*, 65:77–93.
- Steward, Helen, 2015. “What is Determinism”. *Flickers of Freedom blog*.
- Stoljar, Daniel, 2001. “Physicalism”. *Stanford On-line Encyclopedia of Philosophy*.
- Stoljar, Daniel, 2007. “Distinctions in Distinction”. In Jesper Kallestrup and Jakob Hohwy, editors, *Being Reduced: New Essays on Causation and Explanation in the Special Sciences*, update. Oxford: Oxford University Press.

- Stoljar, Daniel, 2010. *Physicalism*. Routledge.
- Strawson, Peter F., 1962. “Freedom and Resentment”. *Proceedings of the British Academy*, 48:1–25.
- Stump, Eleonore, 1999. “Dust, Determinism, and Frankfurt”. *Faith and Philosophy*, 16:413–422.
- Sturgeon, Scott, 1998. “Physicalism and Overdetermination”. *Mind*, 107:411–432.
- Swoyer, Chris, 1982. “The Nature of Natural Laws”. *Australasian Journal of Philosophy*, 60:203–223.
- Taylor, Elanor, 2015. “Collapsing Emergence”. *Philosophical Quarterly*, 65:732–753.
- Thomasson, Amie L., 2007. *Ordinary Objects*. Oxford University Press.
- Thomasson, Amie L., 2010. “The Controversy Over the Existence of Ordinary Objects”. *Philosophy Compass*, 5:591–601.
- Thompson, Evan, 2007. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press.
- Thompson, Evan and Francisco J. Varela, 2001. “Radical Embodiment: Neural Dynamics and Consciousness”. *Trends in Cognitive Sciences*, 5:418–425.
- Tro, Nivaldo J., Travis D. Fridgen, and Lawton E. Shaw, 2017. *Chemistry: A Molecular Approach*. Toronto: Pearson.
- Tye, Michael, 1990. “Vague Objects”. *Mind*, 99:535–557.
- Tye, Michael, 1995. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.
- Unger, Peter, 1980. “The Problem of the Many”. *Midwest Studies in Philosophy*, 5:411–468.
- van Cleve, James, 1990. “Mind-dust or Magic? Panpsychism versus Emergence”. *Philosophical Perspectives*, 4:215–226.

- Van Gulick, Robert, 2001. “Reduction, Emergence and Other Recent Options on the Mind/Body Problem: A Philosophic Overview”. *Synthese*, 8:1–34.
- van Inwagen, Peter, 1983. *An Essay on Free Will*. Oxford: Clarendon Press.
- Walter, Sven, 2006. “Determinates, Determinables, and Causal Relevance”. *The Canadian Journal of Philosophy*, 37:217–243.
- Walter, Sven, 2010. “Taking Realization Seriously: No Cure for Epiphobia”. *Philosophical Studies*, 151:207–226.
- Wandell, Brian, 1993. “Color Appearance: The Effects of Illumination and Spatial Pattern”. *Proceedings of the National Academy of Sciences*, 90:9778–9784.
- Wasserman, Ryan, 2017. “Material Constitution”. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2017 edition.
- Watson, Gary, editor, 1982. *Free Will*. Oxford: Oxford University Press.
- Wedgwood, Ralph, 2013. “A Priori Bootstrapping”. In Albert Casullo and Joshua Thurow, editors, *The A Priori In Philosophy*, 226–246. Oxford University Press.
- Wiggins, David, 2001. *Sameness and Substance Renewed*. Cambridge University Press.
- Wilson, Jessica M., 1999. “How Superduper Does a Physicalist Supervenience need to Be?” *The Philosophical Quarterly*, 49:33–52.
- Wilson, Jessica M., 2001. *Physicalism, Emergentism, and Fundamental Forces*. Ph.D. thesis, Cornell University, Ithaca, NY.
- Wilson, Jessica M., 2002a. “Causal Powers, Forces, and Superdupervenience”. *Grazer Philosophische-Studien*, 63:53–78.
- Wilson, Jessica M., 2002b. “Review of John Perry, *Knowledge, Possibility, and Consciousness*”. *Philosophical Review*, 111:598–601.
- Wilson, Jessica M., 2005. “Supervenience-based Formulations of Physicalism”. *Nous*, 39:426–59.

- Wilson, Jessica M., 2006a. “On Characterizing the Physical”. *Philosophical Studies*, 131:61–99.
- Wilson, Jessica M., 2007. “Newtonian Forces”. *British Journal for the Philosophy of Science*, 58:173–205.
- Wilson, Jessica M., 2009. “Determination, Realization, and Mental Causation”. *Philosophical Studies*, 145:149–169.
- Wilson, Jessica M., 2010a. “From Constitutional Necessities to Causal Necessities”. In *Classifying Nature: The Semantics and Metaphysics of Natural Kinds*, 192–211. New York: Routledge.
- Wilson, Jessica M., 2010b. “Non-reductive Physicalism and Degrees of Freedom”. *British Journal for the Philosophy of Science*, 61:279–311.
- Wilson, Jessica M., 2010c. “What is Hume’s Dictum, and Why Believe It?” *Philosophy and Phenomenological Research*, 80:595–637.
- Wilson, Jessica M., 2011. “Non-reductive Realization and the Powers-based Subset Strategy”. *British Journal for the Philosophy of Science*, 94:121–154.
- Wilson, Jessica M., 2012. “Fundamental Determinables”. *Philosophers’ Imprint*, 1–17.
- Wilson, Jessica M., 2013a. “A Determinable-based Account of Metaphysical Indeterminacy”. *Inquiry*, 56:359–385.
- Wilson, Jessica M., 2013b. “Nonlinearity and Metaphysical Emergence”. In Stephen Mumford and Matthew Tugby, editors, *Metaphysics and Science*.
- Wilson, Jessica M., 2014. “No Work for a Theory of Grounding”. *Inquiry*, 57:1–45.
- Wilson, Jessica M., 2015a. “Hume’s Dictum and Metaphysical Modality: Lewis’s Combinatorialism”. In Barry Loewer and Jonathan Schaffer, editors, *The Blackwell Companion to David Lewis*, 138–158. Blackwell.
- Wilson, Jessica M., 2015b. “Hume’s Dictum and Natural Modality:”. In Barry Loewer and Jonathan Schaffer, editors, *The Blackwell Companion to David Lewis*, 138–158. Blackwell.

- Wilson, Jessica M., 2015c. “Metaphysical Emergence: Weak and Strong”. In Tomasz Bigaj and Christian Wüthrich, editors, *Metaphysical Emergence in Contemporary Physics; Poznan Studies in the Philosophy of the Sciences and the Humanities*, 251–306.
- Wilson, Jessica M., 2016a. “Are There Indeterminate States of Affairs? Yes”. In Elizabeth Barnes, editor, *Current Controversies in Metaphysics*. Routledge.
- Wilson, Jessica M., 2016b. “Grounding-based Formulations of Physicalism”. *Topoi*.
- Wilson, Jessica M., 2016c. “The Unity and Priority Arguments for Grounding”. In Ken Aizawa and Carl Gillett, editors, *Scientific Composition and Metaphysical Ground*. London: Palgrave-Macmillan.
- Wilson, Jessica M., 2017. “Determinables and Determinates”. *Stanford Encyclopedia of Philosophy*.
- Wilson, Jessica M., in progressa. “Causal Composition”.
- Wilson, Jessica M., in progressb. “The Metaphysics of Fundamental Interactions”.
- Wilson, Mark, 1982. “Predicate Meets Property”. *The Philosophical Review*, 91:549–589.
- Wilson, Mark, 2006b. *Wandering Significance: An Essay on Conceptual Behavior*. Oxford: Clarendon Press.
- Wimsatt, William, 1994. “The Ontology of Complex Systems: Levels of Organization, Perspectives, and Causal Thickets”. *Canadian Journal of Philosophy*, 24:207–274.
- Wimsatt, William, 1996. “Aggregativity: Reductive Heuristics for Finding Emergence”. *Philosophy of Science*, 64:372–384.
- Wimsatt, William C., 2007. “On Building Reliable Pictures with Unreliable Data: An Evolutionary and Developmental Coda for the New Systems Biology”. In Fred C. Boogerd, Frank J. Bruggeman, Jan-Hendrik S. Hofmeyr, and Hans V. Westerhoff, editors, *Systems Biology: Philosophical Foundations*, 103–20. Elsevier.
- Wolff, Johanna, 2015. “Spin as a Determinable”. *Topoi*, 34:379–386.

- Worley, Sara, 1997. "Determination and Mental Causation". *Erkenntnis*, 46:281–304.
- Wyszecki, G. and W. S. Styles, 1982. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. New York: John Wiley and Sons, Inc., second edition.
- Yablo, Stephen, 1992. "Mental Causation". *The Philosophical Review*, 101:245–280.
- Yates, David, 2016. "Demystifying Emergence". *Ergo: An Open Access Journal of Philosophy*, 3.
- Zimmerman, Dean W., 1995. "Theories of Masses and Problems of Constitution". *Philosophical Review*, 104:53–110.