1 History and Background

1.1 Surveying the Landscape

Certain things commence in the most unexpected ways. According to Hippolytus, a Roman writer of the third century CE, Anaximander (fl. 571 BCE) held that living creatures arose from moisture evaporated by the sun. Given the crude state of science at the time, and that the leading alternative seems to have been that Prometheus and Epimethius created life on Zeus' command, this venturesome hypothesis warranted further consideration. Less worthy of consideration was a rumor, circulating in the seventeenth century, that if one put sweaty undergarments and husks of wheat together in an open-mouthed jar and waited 21 days, the action of the sweat on the husks would generate mice. Like the belief circulating among sailors that maggots were propagated by rotting meat, such speculations were not put to rest until Louis Pasteur's 1859 experiments undercut any basis for belief in spontaneous generation.

Those exploded claims were emergentist theses. Their common basis was that radically different sorts of things "emerged" from an ontologically simpler foundation in ways defying rational expectation. If that was emergentism's finest fruit, the view would be no more than a quaint historical relic. And, as we shall see below, more recently even sober emergentist claims have been derailed by scientific advances. But there remain open questions which seem to show that emergentist options still have a pulse in some quarters. The one I want to explore in this work is the question of conscious properties arising from brute, unthinking matter. Not only do I estimate that this is still an open question; I shall argue that there is no equally plausible non-emergentist alternative. Consciousness isn't the only live emergentist possibility. Normativity and color are two prominent instances

of as-yet-unresolved and potentially unresolvable puzzles. Although each comes up later as a side issue (color in section 1.7, semantic norms in section 7.4), I will say nothing further about those prospects in this work. Rather, I will concentrate exclusively on the topic of conscious properties (more generally, conscious tokenings).

How to fit consciousness into a largely material world is—like "Why is there something rather than nothing?"—a perennial question that is hard to dismiss. Questions of this kind tempt some into supernatural resolutions. But, aside from the fact that a supernatural solution assumes the prior existence of what should be part of the question, they are too facile for thoughtful inquiries. There seems to be no alternative to carefully probing the natural world for an answer.

As a continuous school, emergentism entered Anglophone philosophy in 1843 with the publication of John Stuart Mill's *System of Logic*.[1] (The term 'emergentism' came later.) Its earlier advocates regarded themselves as staunch defenders of naturalism, although nowadays that may seem paradoxical. (On this change in fashion, see footnote 6.) Philosophical emergentists have always been in the minority, and had almost completely disappeared by the 1960s. But there has been a mini-resurgence of the view in some areas, especially those concerning a subclass of natural laws. However, the view I shall be recommending for reconsideration doesn't quite fit the mold of either the earlier or more recent versions. As I have noted, I am confining attention to emergentism in the philosophy of mind, and in particular to aspects of consciousness. Upon becoming aware of unwarranted assumptions that have made emergentism seem mysterious, I hope that it will be clear why this should not be unseemly even to the tough-minded.

I sidestep some basic questions about the phenomenology of consciousness or how we come to be apprised of it. I will rely on its self-evident character, without trying to decide whether awareness of consciousness is attained by introspection, knowledge without observation, natively, by intuition, or by some other method. Indeed, I can provide doubters no firmer demonstration of its reality than to suggest that they pinch themselves.

Salient recent accounts of conscious life that conflict with emergentism are considered in part II, where I discuss reasons for rejecting what I take to be their best versions. Here (and perhaps generally in philosophy), *a priori*

1. My quotations are from the third edition (1851).

demonstrations seem out of reach, but I believe we can achieve what Mill elsewhere (*Utilitarianism*, chapter I) called "considerations . . . capable of determining the intellect."

Before turning to concerns rooted in the philosophy of mind, a few remarks about the general character of emergentism are in order. This will be followed by a brief historical summary of what I am calling *classical emergentism*. The summary does scant justice to emergentism's contributions to metaphysics,[2] but I hope it will acquaint readers with both the similarities and the differences between that tradition and the view elaborated here. To orient the reader, I also describe some of the reasons the movement has fallen into disfavor.

1.2 Emergentism Depicted

Originally emergentism covered a broad array of topics, from chemical compounds to sensible qualities. Although it was never philosophically ascendant, for much of the first half of the twentieth century it was taken seriously even by those who opposed it.

Emergentists and their opponents generally shared a hierarchical picture of explanation, dependence, and the sciences—known also as a *layered conception of reality*. Suppose we proceed from relatively simple constituents and their interactions to their complex constructions. It is customary to think of the more complex as occurring at a higher level than the less complex, the lower infusing the higher with its features. Unfortunately, there may be no way of achieving a coherent unified portrayal of levels for every purpose to which this image has been put. Still, the picture has an intuitive pull, and for our purposes we needn't fuss over details. A generic hierarchy will simplify the exposition without placing any of the positions at a disadvantage. We might start with atoms as basic constituents—"swarms of atoms" for C. Lloyd Morgan (1923, p. 35)—and might then proceed to molecules, thence to chemical compounds, and eventually to macroscopic physical objects. Or we might go from individual molecules, to genes, organelles, cells, tissues, organs, and finally to organisms. Or perhaps, emphasizing a hierarchy of sciences, we might start from particle physics, then, omitting a number

2. For more thorough treatments, see McLaughlin 1992, McLaughlin 1997, and Kim 1992. For Mill's contribution, see also E. Nagel 1961.

of intermediate steps, proceed to chemistry, geology, biology, economics, psychology, and sociology.

From those bare materials, let's devise this rough sketch of emergentism: On some occasions, a novel outcome results from increased complexity in that outcome's base. To explicate the novelty in question, we might say that the lower levels on which emergent phenomena rest contain nothing that could enable one to anticipate the former's initial occurrence in terms of the lower level's types of intrinsic features, its structure, or a comprehensive examination of both.

Despite the fact that the view has been described in terms of what one can anticipate, the emergentism that will occupy us is intended as an *ontological* thesis. As the point is sometimes put, it is "over and above" its base. Various writers, among them Chalmers (2006) and Crane (2001), have distinguished an ontological variety of emergentism from an epistemological variety, regarding the latter as a worthwhile variation. However, a popular objection to classical emergentism has been that its ontological claims are transformed into an epistemic thesis about what one can expect. From that angle, the view is then readily dismissed as a form of temporary scientific ignorance. Whatever the value of an epistemic variety, the interest here is in ontology. The term 'emergentism' is nowadays tossed about regularly in scientific and philosophy-of-science circles. Those uses may bear to some extent on the brands of current interest, but in ways not easily summarized. Some of them, though not all, are epistemic. Moreover, classical emergentists have written in ways suggesting both kinds of view. However, the emergentism targeted here is an ontological thesis, and it is part of my task to distinguish that view from its epistemic kin.

Ordinarily what is taken to emerge is a *property* (or a feature, or a characteristic, or a quality). My treatment is focused largely on *instantiated* properties. Other versions might emphasize events (including static events or states), relational properties, processes, facts, or even individuals. However, most of the subsequent inquiry should be applicable, *mutatis mutandis*, to all, or many, of those aspirants; when the choice between them is not relevant, I often use neutral terms such as 'token' and 'aspect'.

Chemical compounds were once emergentists' favorite illustration. Various other things the view embraced at one time or another include life, organic functioning, lawlike generalizations (including stochastic laws), consciousness, (intentional) mental states, secondary qualities, norms,

purposes, organized collectives such as nations, and other social or conventionally enabled phenomena. Items on this list do not share a common base from which each is supposed to emerge, although given the transitivity of supervenience some may claim that each is ultimately resolved into interactions among fundamental particles. Confining ourselves to immediate explanations, for some theorists life arises from structural properties of proteins, perceptual awareness from 40-hertz semi-oscillatory firings of neurons, colors from light and reflectance potential, norms perhaps from conventions, which themselves arise from tacit agreements (or acquiescences), and nations from the cooperative interactions of inhabitants. Certain emergentists have claimed that at least some of these phenomena are novel structures arising from material that lacked structures prefiguring their appearance.

Of course, not every first appearance of a complex aspect counts as emergent. Details matter. For example, if simple mereological composition will suffice to explain a complex constructed from its parts, the whole is not emergent. As C. D. Broad notes (1925, p. 62), if we have two forces acting on a particle at an angle to each other, we "find by experiment that the actual motion of the body is the vector-sum of the motions it would have had if each had been acting separately." Or, if two chunks of matter could be combined into a single aggregate, their combined rest mass would be the sum of the rest masses of the original chunks. A label that seems to have caught on among classical emergentists for describing the non-emergent is 'resultant'. Soon we will need a more rigorous statement of those differences, but this should be enough background to enable us to follow the movement's early development.

1.3 Classical Emergentism

Emergentism flourished from roughly the middle of the nineteenth century to the middle of the twentieth. Mill broached the topic by introducing a distinction between two modes of causal interaction: the mechanical and the chemical. He gives "the name Composition of Causes to the principle exemplified in all cases in which the joint effect of several causes is identical with the sum of their separate effects" (1851, book III, chapter VI, p. 374). This is intended to cover all mechanical causation, not only when causes augment one another, but also when one cause is partially thwarted by a

second, thereby altering without obliterating the first's impact. Whereas the Composition of Causes is the general rule, there are exceptions; they issue in the products that later came to be known as emergent. According to Mill, "the chemical combination of two substances produces, as is well known, a third substance, with properties different from those of either of the two substances separately or of both of them taken together" (ibid., p. 374). Mill (ibid., pp. 377–380) mentions "heteropathic laws" as covering cases "in which the augmentation of the cause alters the kind of effect." Shortly after Mill, the contrastive term 'homopathic' was coined for mechanical causal laws. We can easily extend the heteropathic/homopathic distinction beyond laws to particular causes. In fact, without further explanation Mill uses the terms 'heteropathic causes' (p. 380) and 'heteropathic effects' (pp. 377, 381).[3] George Lewes later coined the title 'emergent', which eventually replaced Mill's 'heteropathic' as the consensus designation.

Mill's late contemporary, Alexander Bain gestures toward the same phenomenon when he distinguishes force, which is always mechanical, from collocation. He writes that "the mixing of materials, and its union of force, are not the same fact" because "we cannot fully predict the characters of the compound form [its] elements" (1887, p. 270). They "yield a new product, where the combining elements are not recognizable." Bain also acknowledges its broader range: "The analogy of Chemical Combination has been applied to mental and social combinations."

Emergentism was taken up as a more-or-less self-conscious cause early in the twentieth century. Its hallmark thesis was that emergent phenomena resist an expected mechanical explanation in their bases. Samuel Alexander, a leading figure in the movement, declared that such phenomena should be taken with "natural piety," which then became a watchword among those writing on the subject.[4] As Broad remarked with characteristic mordant wit, these results "must simply be swallowed whole with that philosophic jam which Professor Alexander calls 'natural piety'" (1925, p. 55). On the other hand, Charles Lloyd Morgan preferred a similarly reverent exhortation: when encountering emergence, "consider and bow the head" (1923, p. 4).

3. When first treating effects rather than just laws as heteropathic, Mill attaches a footnote reading "Anet ch. vii, §1." But I find no mention of heteropathic *effects* there or any earlier in Mill's *System of Logic*.

4. The phrase seems to have originated in the epigraph to Wordsworth's "Intimations of Immortality": The child is father of the man; / And I could wish my days to be / Bound each to each by natural piety.

Aside from Morgan, we may count Arthur Lovejoy among influential early-twentieth-century emergentists, and several notable authors joined him. Also, a group of present-day chemists, biologists, and philosophers of science declare themselves for emergentism. In some instances, their forms of emergentism are different, both in content and in ontological implications, from the present concern with conscious properties; in yet others, there is overlap. It would be unproductive here to try to cover the many varieties going under the title 'emergentism'. In what follows, I shall pick my way cautiously through that material for points and theses that intersect with our narrower concern, but shall focus only on several closely related views.

1.4 Problems and Refinements

It did not take long for emergentism's critics to find weaknesses. Indeed, some were spotted by its followers.

First, scientific progress has placed some of emergentism's former claims *in extremis*,[5] and typically they occurred in cases taken as emblematic of the movement. Chemical combination and organic life may be the first subjects that occur to most when thinking about classical emergentism, and for good reason. Although there were also emergentist treatments of the mental—and in spite of the title of Broad's celebrated opus *The Mind and Its Place in Nature*—in the view's heyday the examples of chemical compounds and living things (including evolutionary novelties) were its most prominent themes. But discoveries in quantum mechanics introduced explanations of chemical bonding via electromagnetism, and the disclosure of the DNA composition of genes has opened the way to explanations of crucial features of life such as self-replication, both of which emergentists had declared resistant to physicalist treatment.[6]

Next, objections have been raised to ways of defining the position. Of course, any attempt at a philosophical definition or analysis may be plagued by allowing in too few of the intended cases or too many of the unintended ones. Efforts to fill these gaps regularly end up in trivialization

5. McLaughlin 1992, p. 54 ff.

6. Taking life to be emergent was originally a *materialist* response to the doctrine of vitalism, which had explained life via immaterial factors such as *élans vital* or ectoplasms.

or vicious circularity, not to mention baroque qualifying clauses that can drain a view of its initial attraction. Emergentism is no less prey to those dangers than are the general run of attempts, with rare exceptions, to provide a challenged concept's illuminating necessary and sufficient conditions. But, beyond that, there have been complaints directed at distinctive aspects of emergentism. Here are two notorious examples.

A popular specification of emergentism latches on to unpredictability upon first appearance. As was noted earlier, some have charged that this turns what was advertised as an ontological discovery into a claim about our current understanding. Our inability to predict an outcome upon first occurrence would be a fragile basis for the thesis, one that scientific progress has regularly overturned. This alone makes it unwise to transform emergentism into an epistemic thesis. However, leading emergentists didn't mean anything that chancy by 'unpredictability'. Predictability for them wasn't clearly distinguished from their notion of deducibility. The latter also bore an interpretation that by current standards is outdated: it included derivations requiring generous substitutions of terms. But it was not broad enough to collapse into an epistemic notion. For example, notice in the earlier quote from Bain, he writes that "we cannot *fully* predict the character of the compound" (emphasis added). Such qualifications were almost always at hand. This is not to deny that our current epistemic position, when carefully reflected upon, can be a useful clue to explanations not being in the offing. But the claim of current interest, and no doubt the one at which classical emergentists were aiming, is ultimately about the nature of external reality, not about current knowledge. Classical emergentists generally have been more guarded in their official pronouncements.

A tempered account of the intended doctrine seems to embody two requirements: (a) a complete, or ideal, characterization of the base from which the aspect emerges, and (b) an inability to deduce (or infer, or predict) an emergent aspect from that ideal base. This version of emergentism has valid historical credentials. The idealized explanations in (a) would disclose, in Broad's terms, what "can be deduced from the most complete knowledge of [the constituent properties] in isolation or in other wholes which are not of [their form]" (1925, p. 61).[7] But this invites the question "What belongs in a total explanation of the base?" Available answers

7. Also see Alexander 1920, volume I, pp. 46–47.

discover new complications. One method might be to build enough into requirement (a) so that its consequences would always be deducible. For example, suppose that a *complete* explanation of hydrogen includes "forming, if suitably combined with oxygen, a compound which is liquid, transparent, etc. Hence, the liquidity, transparence, etc. of water *can* be inferred from certain properties of chemical constituents." (Hempel and Oppenheim 1948, in Hempel 1965, p. 260) (Also see Beckermann 1992.) Similar dispositions or tendencies may be devised for inclusion in any base properties. With regard to pain, suppose its base includes the firing of C-fibers, activity in the somatosensory and prefrontal areas of the brain, etc. of a more complete organism. We might then ascribe to a complete or idealized knowledge of, say, C-fibers the conditional power to produce pain in its subjects when firing in the relevant conditions and background. That would enable us to infer pain from its dependence on nothing more than a combination of its base properties.

Emergentists were not unaware of this objection. Broad explains what he calls a "trans-ordinal law" as follows:

> A and B would be adjacent, and in ascending order if every aggregate of order B is composed of aggregates of order A, and if it has certain properties which no aggregate of order A possesses and which cannot be deduced from the A-properties and the structure of the B-complex *by any law of composition which has manifested itself at lower levels*. (p. 78, emphasis added)

From there it is possible to claim that the exceptions, the emergent aggregates, all use laws which are unmanifested at a lower level. But it is ordinarily much easier to rule out certain clear violators than it is to devise a formula covering all and only acceptable aggregates. Or, given that the Hempel-Oppenheim objection relies on mentioning relational properties, another suggestion might be that at each level an aggregate include only intrinsic properties. But, alas, that move is doomed. It is difficult to take *structure* into account without relational properties, and we do need structure (that is, form) in our base: X inside Y is different from Y inside X. A further suggestion might be to allow certain relational properties as long as they don't relate aggregates at different levels. Aside from the difficulty of delineating a principled distinction between levels, this will prevent us from including any potential for combinations of items. One might settle for ruling out certain cases without requiring a general description of admissible versus inadmissible laws. It is a challenge to show that those

exclusions are not *ad hoc*; another is to show why this is satisfactory in the absence of any principle.

Yet it could be the particular approach rather than emergentism itself that is problematic here. To get a better grasp of the state of the dispute, consider two ways, following Broad (pp. 24–25), of regarding the physical base properties.[8]

As the emergentist will insist, the physical base properties in certain combinations have the potential to issue in its emergent properties. Broad calls such properties when they are unrealized "latent." (This doesn't imply that *all* latent properties would, when realized, issue in emergent ones.) Taking our cue from that, we may regard a description of the (physical) base including both its manifest *and* latent properties as *the latent base*, or, if you prefer, *the latent description of the base*. An inability to show how to avoid taking the latent base for the base *period* fuels the Hempel-Oppenheim objection. However, a feature of this base as described is that no one is in a position to include its various latent properties until they have been combined in what Sydney Shoemaker (2002, p. 54; 2007, p. 76) has called "emergent-engendering ways."

The other way of regarding the base is to include only those properties that are manifest. In addition to its shape, color, size, weight, texture, and odor, this will include many if not all of the base's conditional powers[9]—what it can cause, again perhaps in combination with other properties, and what can cause it. For example, an object's having a certain bulk indicates that it can crush a fly, even if nothing with those manifest properties ever crushed a fly. Call this *the manifest base*, or *the manifest description of the base*. However, distinguishing a manifest base from a latent base in a principled way is fraught with difficulties.

The problems reviewed thus far have to do with achieving a respectable statement of the position. More particularly, they result from understanding what it is to be deducible from a physical base or from the requirement that we divide the physical base into two sorts and then rely on using only one of them (the manifest base) to bring out emergence's defining features. I introduce an alternative characterization in section 1.6, and elaborate it in chapter 2. Although it demands no such division of descriptions, it does

8. I am generalizing, as Broad intended, from his illustration of silver chloride and its components, Ag + Cl.

9. See Shoemaker 2003.

not escape other criticisms that have been leveled against emergentism. A final set of problems awaits all forms.

A leading problem has been how to find a causal role for *any* independent mental tokens. It is generally agreed that in order to have any effect, including a mental one, a mental property must have an effect in the physical world. (For dissenting views, see Gibbons 2006; Stephan 2002; Craver and Bechtel 2007.) That involves the notorious problem of *downward causation*, the causation of something at the physical level on which the mental depends. Moreover, every physical effect presumably has a physical cause, and it has been argued that the physical cause trumps the supposed mental one. Don't a mental property's causal powers reduce to those of the ultimate bases on which it rests? And if emergent properties have no causal role, what reason have we to believe there are such things? These questions are discussed in section 3.7 and probed further in chapter 4. One might also ask what it means for something constituted by its supervenience (or realization) base not to be identical with it. That question is also addressed in chapter 3. Powerful intuitions driving physicalism expose another challenge to emergentism. Whereas some classical and recent emergentists have declared that their views are consistent with physicalism, forms of present concern sharply distinguish them. Physicalists of this stripe may claim that emergentism makes a mystery of its relation to physical reality. In part II I examine those and other leading naturalist views to see if they contain persuasive grounds for rejecting emergentism. Here I simply note the battery of challenges that my highlighted position faces. Their cumulative effect may help explain why it is so difficult for current versions of emergentism to get admitted to the ranks of serious options.

1.5 Emergence and the Mental

Under *conscious* properties and/or states I include not only phenomenal consciousness, but also what Ned Block (1995, 1996) has termed "access consciousness." Phenomenal consciousness is the undergoing of a conscious state.[10] Leading examples are sensations such as a headache, a tingle in one's leg, an itch, drowsiness, orgasm, an after-image, and perceptual

10. Although I take this to be nothing beyond undergoing a first-order state (e.g., Block 1995; Dretske 1993), these points are not in conflict with the view that consciousness is a second-order monitoring of its first-order episode.

experiences such as of a blue patch or the taste of cinnamon. Some of these states are representational, but, it is contended, there is a "what it is like" (Nagel 1974) that resists consignment to their representational contents. Access-conscious properties are those that one has directly available for use in reasoning and in other (verbal and non-verbal) behavior. Block calls them "inferentially promiscuous." They contrast with phenomena, such as blindsight or a Freudian unconscious, in which a subject's behavior might be directed by intentions and motives to which she has no unmediated access. Although for the bulk of this essay my choice of examples will concentrate on phenomenal properties (and then mostly on pain, because it has been focal in these discussions), the points raised are intended to apply to both forms of consciousness. Access consciousness isn't seriously contemplated until chapter 5.

No special note needs to be taken of the occasional claim that access consciousness already involves phenomenality. The point to be emphasized is that the mental phenomena that concern us cover this spread of cases, its extension. In fact, if access consciousness is just a species of phenomenal consciousness, that should make it easier to draw lessons for the former from the conclusions reached about the latter. Indeed, it may even extend those conclusions to the bulk of mental properties.

1.6 Relevant Emergentist Theses

Finally, here is the version of emergentism that constitutes our central topic. It can be summed up in three theses, to be elaborated as the discussion proceeds.

Imagine that E is a representative sample of the properties of concern. To be emergent, E must meet the following set of conditions:

(1) E is *dependent* on different sorts of a non-emergent base in a way made manifest by E's *supervenience* (or *realization*) on those same properties.

(2) There is *no* further (minimal) *explanation* of why E is supervenient on (or dependent on, or realized by) that non-emergent base, viz., the relationship is brute.

(3) E is a cause (of both mental and physical aspects) in ways in which there is no sufficient cause in context at the levels of E's non-emergent base(s).

I leave open whether the supervenient or realization base is itself a property, a collection of properties, or something else. Our dealings are predominantly with properties, but we need make no commitment on the general issue.

Each of the three conditions cries out for further clarification and defense. Beginning with (1), there are a number of distinguishable forms of supervenience. To what extent does (1) depend on which of them is chosen? Why suppose that any of the forms establishes the dependence of the supervening property? In fact, as we shall see in chapter 3, realization turns out to be more central to emergentism than supervenience.[11] (The realization relation of concern is that in which something is realized *in* something else, as a statue being realized in marble, not that in which something is realized *by* something else, as Smith's profit being realized by last year's investment.) Regarding (2), the claim is that *no* further minimal explanation exists. This is required to distinguish the present view from the epistemological interpretation stating that we do not possess (or will never possess) the desired explanation. That view was set aside earlier. Still it remains that our never being able to achieve an explanation could result from our limitations rather than from the nature of cognitively indifferent reality. Only the latter is of interest for the ontological thesis now being examined. Two additional questions are pertinent: How can anyone in our present circumstances claim with any confidence that *there is no* further explanation? What can be meant by the *sufficiency* of a cause? Questions of this order concerning supervenience and explanation are addressed in chapter 2 to the extent that our limited concerns dictate.

Conditions (1)–(3) isn't the only form in which emergentism appears in the current philosophical literature. Some emergentists reject (1) on the grounds that once an emergent aspect arises it radically transforms the base on which it depends; the base then disappears and is absorbed into its emergent product. Earlier, Alexander (1920, volume II, p. 9) suggested as much: "The neural process which carries thought becomes changed into a different one when it ceases to carry thought." (Recall also Bain's remark that "the combining elements are not recognizable" in the collocation.)

11. Distinctions between supervenience and realization are explored in chapter 2. However, because the extensions of each in cases of present interest overlap almost completely, we are able to overlook their differences for much of this discussion. What matters is a constitutional dependence in both.

Others (e.g., Gillett (2002)) suppose that (3) by itself, or (3) with the addition of (1), suffices for emergentism. Those views are compatible with some robust forms of physicalism. Finally, minimal forms of emergentism, called "weak emergentism" by Bedau (1997), need only (1) and (2). Some of those variations are considered in subsequent chapters.

We may illustrate weak emergentism with a favorite ploy of opponents of physicalism: the possibility of zombies. (See Chalmers 1996.) Zombies are physically, and perhaps behaviorally, indistinguishable from us, but lack phenomenological properties, commonly known as qualia. Their behavior is directed by their physical components, and if it is possible to have beliefs without qualia, zombies may even falsely believe that they have conscious states and properties. If we are not zombies, it must be by virtue of nonphysical features we possess. Under (1) those features do not float free of the physical world, and on (2) their intrinsic nature is not explicated by the features of that world. But nothing about our differences from zombies shows our additional features to have any causal powers that are not contributed by our physical bases. For that we need condition (3).

Weak emergentism has been too timid a view for most emergentists, indeed even for typical non-emergentist commentators. On the weak view, despite our differences from zombies, our distinct mental lives might be epiphenomenal, and this, it has been held, is unacceptable. (See below.) Any form of the view that includes (3) could be known as *strong emergentism*,[12] which I shorten to *emergentism* because the only versions seriously considered in this work ascribe causal relevance to emergent aspects. But, as I noted earlier, any emergentism incorporating (3) encounters serious problems about the causation of the mental. Before tackling that issue in earnest (in chapters 3 and 4), a few initial remarks about causation may be in order.

An additional distinction is relevant to article (1). Supervenience, or realization, is synchronous with what supervenes on it or is realized by it. We are concerned here only with varieties of emergence meeting that requirement. This isn't an arbitrary stipulation; it captures what seems to me to be the main tendency of the current view. Certain classical emergentists may regard the base as preceding its emergent. It hasn't even prevented some current emergentists from making similar claims. For example, in

12. This differs from Chalmers' (2006) taxonomy, in which 'strong' and 'weak' designate, respectively, ontological and epistemic emergentism.

evolutionary versions of the doctrine, the base may be an earlier form of plant or animal giving rise to a novel form. In yet other versions the base is specified as the (efficient) cause of the emergent, and it is generally supposed that a cause precedes its effect. Those variants aren't ruled out, but they aren't part of the view under discussion. Our concern is with a theory in which the emergent is supposed to supervene on or be realized in its base, both synchronic relations. This doesn't discharge all criticism of (1). In the next chapter I will briefly discuss a challenge to (1) that raises an objection to an independent base, but it concerns only the interplay of synchronous factors.

Recall that the discussion largely concerns *properties*. One may wonder how, when the topic turns to causation, properties, rather than, say, events, can be the relata of prime interest. However, this treatment of properties is meant to cover tokens or instances of properties as much as their types. The properties under consideration, unless we are discussing type-type identity theories, are instances, also known by some as 'tropes'. They are as individual as the particulars to which they are ascribed. Our interest is in the red of *this* tomato, not redness in general or the redness of tomatoes as such. When broaching the topic of causation, my remarks should also apply, *mutatis mutandis*, to events and states. Depending on one's further views, the difference may turn out to be merely terminological. On one popular account, events are instantiations of properties. That interpretation creates a convergence between issues about the causal prowess of properties and states. If we are to remain resolute realists about causation, the particular case ought to be of special interest. Our properties and/or states must be causally efficacious (or causally relevant) if our causal generalizations are to have this realist bite.

Although this affords us some latitude in discussing issues interchangeably in terms of properties and events, we needn't suppose that the difference between an event and its property is *never* relevant. But the differences would be more pronounced if, as in many discourses, it had been assumed from the start that the events discussed were all particular occurrences and properties were all universals or types.

Sticking with the theme of causation, consider epiphenomenalism of the mental—the view that it never causes anything, although it is an effect of other causes. That doctrine is not without advocates, but it also has a considerable dialectical burden. One solution to the causal difficulties besetting

the mental is the view that a mental property gains a causal role by virtue of its identity with a physical property. If so, the causal efficacy resides in the property under its physical description, not in its mental aspect—what has been called "type epiphenomenalism." The problem—known sometimes as the qua problem (Maslen et al. 2009)—was noted by Broad (1925, p. 473): "Epiphenomenalism . . . simply says that mental events either (a) do not function at all as cause-factors; or (b) that, if they do, they do so in virtue of their physiological characteristics and not in virtue of mental characteristics."

Why should epiphenomenalism seem so implausible here? On a popular account, I can become *empirically* acquainted with an X only if X figures causally in my experience.[13] Suppose the same is true of our conscious episodes. If they could not be causes of our judgments about them, it would be perfectly mysterious why we ever supposed we had them in their mental semblance (or, for that matter, how we could "suppose" anything at all). Against this, some have claimed that conscious phenomena are different; there is no internal sensory faculty by which to detect them. A few have even claimed them to be *a priori* and thus not regulated by the conditions governing experience of the empirical world. However, these differences fail to shake the conviction that if conscious properties couldn't play any role in a self-awareness of them, it would be hard to see how we should have happened upon the belief that there were such properties. Even if a conscious property's empirical credentials are tainted, it is only a *contingent* truth, say, that my thumb hurts. Thus, it is not something I can excogitate in the manner of a typically necessary *a priori* truth, such as that 2 + 3 = 5. It would be the contingency of the pain, rather than its paradigmatic empirical character, that would demand its causal role here. Even if, as some suggest (e.g., Horgan and Kriegel), awareness is an intrinsic feature of an occurrent phenomenal state, the second-order knowledge of that state would invoke a causal connection to the state.

I have yet to complete this initial summary of emergence. But before attending to that, perhaps a quick review of the variety of positions in the last century or so on the nature of the mental, painted in very broad strokes, will locate emergentism more definitely for the ensuing discussion.

13. A thesis defended in Vision 1996.

1.7 Theories of the Mental I: Eliminativism

Strictly speaking, eliminativism declares that nothing in the empirical world correlates closely enough to our mentalist vocabulary to warrant taking the latter as more than a useful fiction. Occasionally the view is explicit, often only implied. But in each embodiment it constitutes an *irrealism* about mental, including conscious, properties. Here I use the term 'eliminativism' broadly to cover an aggregate of irrealist options that may go under titles such as instrumentalism, error theory, and even some forms of functionalism. With the exception of a certain relevant form of functionalism, these options also tend to be physicalist in spirit; but unlike the physicalisms sketched below in sections 1.9 and 1.10, they reject outright or diminish beyond recognition the reality of conscious aspects. In a more comprehensive review one would be obliged to examine this collection in greater detail, but nothing that encyclopedic is undertaken here. One reason is that it would distract us from the main target of the exposition, the standing of emergentism. Another reason is that the zeitgeist seems to indicate that the realist alternatives are the leading naturalist views when issues relevant to emergentism are aired. This section contains some further remarks on the irrealist alternatives, but afterwards eliminativism largely drops out of my deliberations (save for a brief reappearance in the epilogue). However, readers may justly wonder what entitles me to be so cavalier in dismissing this view. So as an apologia I set forth one reason for not pursuing it further: namely, once phenomenal experience enters the inquiry, it is a mistake to offer an account of it, including a debunking one, that omits the 'what it is like' of phenomenal states.

First, a qualification. There is a vast literature on consciousness, and a substantial portion of it contains a much more extensive and thorough treatment than I can offer here.[14] In fact, it would be a mistake to take what I say in this section for anything as grand as an account of phenomenal experience. But I can briefly explain why the 'what it is like' of conscious experience is indispensable to any competitive theory in which implications are drawn regarding it. This still leaves in the field a number of competing views, including dualism and other physicalist theories described below. My reasons rule out only what I have labeled as irrealist views about the qualia of conscious sensation.

14. For a list of the ways in which 'conscious' has been employed, see Lycan 1996.

Irrealism diagnoses sensations such as pain very poorly. To illustrate, compare the options for one's theories of pain qualities with those for heat and color.

It is a fair guess that humans first became acquainted with heat as a sensation, ranging in intensity from comfortable warmth to extreme hurtfulness. Even if this isn't an accurate history of the race, it certainly seems to be the way young children come to understand heat. If adults warn "Hot!" as they point to items such as a stove, a pavement, or boiling water, how might that register for young children before they have had an unpleasant sensation? Despite this manner of becoming acquainted with heat, the common heat found in our workaday environment has been discovered to be molecular motion. Although we are introduced to heat via a range of feelings, it is in fact a wholly non-subjective feature. We have, as John Searle put it (1992), carved off the surface features of heat to uncover our definition.

An extension of this reasoning to the sensation of pain may appear to establish irrealism about pain's inherent phenomenology. Just as scientific progress led to marginalizing the feeling of heat, it should do the same for pain. But now consider color. While color is also consciously experienced, the question of whether it is a mind-independent feature of the world remains *sub judice*,[15] both in philosophy and in the relevant sciences. It is not that we lack sufficient information about the physical basis of color experience. Our knowledge here is not inferior in kind to that which we have for heat. If it were merely a question of discovering color's objective correlates—say, surface reflectances for non-luminous objects—everyone should agree, as they do for heat, that this is what color *is*, nothing more. But the debate still rages about whether these mind-independent features are the colors themselves or, as in the case of secondary-quality and error theories, merely experiential contents triggered by an object's non-chromatic features. This is not the place to try to resolve that dispute. But if heat exemplifies the standard by which to decide these cases, the philosophical issue should dissipate once there is general agreement on the scientific facts. Nevertheless, the dispute over color continues even in precincts in

15. X is mind-dependent $=_{df.}$ X is not (/no longer) cognized → X is not (/no longer). (Variations in accounts of mind-dependence will depend on the modal strength of the implication, on how dispositions to cognize are viewed, and on one's views about that definition's tensed forms.)

which there is broad agreement on the science. For whatever reason, some consciously felt properties already have a built-in slot for the non-subjective qualities with which they are to be identified even before the science arrives, whereas others do not.

Why does the objective or mind-independent definition of heat make sense? A possible explanation is that heat does many things other than cause sensations. It fries eggs, expands metals, starts fires, dries up puddles, boils water, melts ice, blisters fingers, both nurtures and kills plant life, and so on. This enables us to intelligibly imagine a world in which an isolated race of intelligent beings have in their environment the same variations in kinetic molecular motion, but do not feel it, never have felt it, and have no inkling that there might be creatures who do feel it (save by conducting Nagelian-like thought experiments about, say, possible bat-like creatures that might have a sense to detect it). Of course, this 'thing' still burns and causes injuries, nourishes these beings, and in sum has all the non-sentient effects on their bodies that heat has on ours, which undoubtedly causes them to take note of it just as we take note of the effects of vitamins. Moreover, we can also imagine that they have the word 'heat' in their vocabulary, signifying the same phenomenon that our similar-sounding word signifies. They simply do not detect its presence or absence in their sentient lives.

I find no reason to suppose that this scenario is incoherent, either intrinsically or in conjunction with current thermal physics. Chalmers (1996, p. 45) claims that in leaving out the sensation of heat "part of the phenomenon is left unexplained." But how does this leave the account any more incomplete than if we had left out instead its ability to blister skin? Was there no heat before there were sentient beings? Or was the mere disposition to cause a sensation if there were sentient beings sufficient for the concept? Indeed, doesn't our current understanding of heat already leave out what may be many equally standard features of it, such as its capacity for interacting with the potentially vast number of elements in the universe with which we are likely to be unfamiliar? My imagined scenario seems to demonstrate that heat sensations are historically and epistemically important for heat's discovery, but not any more significant than its other features for an account of what heat *is*. (Again, it is not obvious that a comparable imaginary setting replacing heat with color would have a similar result, at least according to secondary-quality theorists. Could they

allow that there were colors in any but the most attenuated dispositional sense in a world with no sentient creatures?)

Now let us turn to pain. I can't imagine a similar race of beings who *had* pains, but were constitutionally unable to feel them. Unfelt pain, as many have held, makes no sense. This is not a question of whether peripheral cases in our world might accommodate a sincere report of not feeling pain to be outweighed by neural evidence to the contrary (see, e.g., Lamme), a view on which we needn't take a stand, but whether we could imagine a possible world in which *no* creatures *ever* felt pain, knew of no other creatures who did feel it, but in which nevertheless there were pains only because there were the neural states that currently subserve our pain. If, as I believe, this is not imaginable, we cannot skim off the subjective features of pain and still retain what we thought of as our concept of pain. Some have held that pains have a content to the effect, roughly, that the body has been damaged. This isn't ruled out by the requirement that pain be felt. But bodily damage or its danger cannot be the moral of the tale. We are familiar with very many instances of bodily damage that aren't painful (e.g., various tumors or certain of their stages, the debilitating effects of brain deterioration in senility, leprosy, clogged arteries, or poison ivy damage which causes only itching), so we cannot retain our concept just by hiving off the subjective element and defining it via its tendency to be induced by bodily damage or distress. Indeed, this illustrates an important difference between heat and pain. Heat, as we have seen, comports with its physical manifestations; but the concept of pain persists despite a considerable disconnection with instances of bodily damage. It seems perfectly clear that any concept of pain that indicated *only* bodily damage without sensation would be an altogether different concept from the one we have now. An isolated race of creatures unable to feel pain wouldn't be in pain in any previously recognizable sense.

This is not to say that we are prohibited from redefining pain. We can redefine any concept. But if we do so for pain, it is evident that we have crossed a line that was not crossed for heat. It is not necessary to latch on to any one competing account to explain why this is so. This difference in our treatments of heat and pain may have something to do with the fact that heat manifests itself in so many other salient ways, whereas pain does not. However, the difference is more robust than these remarks; *qua* phenomenon it outlasts all failures in our feeble efforts to explain it. There

is certainly a strong inclination to hold that pain must be felt in all but perhaps a few exceptional cases, whereas sensations of heat seem to us, and should have seemed to us even before the advent of modern science, only contingently related to heat. There may be some who are ready to reject the intuition about pain, courtesy of overarching metaphysical or semantic commitments. But, as Wittgenstein admonished (1953, I, §66), "don't think, but look!" The rest is speculation.

Nothing in this rules out physicalist reduction. It remains possible to discover that pain is identical with physical, functional, or representational X. What is precluded is only that, perhaps per impossibility, Xs without consciousness of pain would no more be pains than Mark Twain without Samuel Clemens would be Mark Twain.

Searle agrees; however, he also states that the difference is trivial, boiling down to the pragmatic fact that we are more interested in the subjective character of pain: "where the phenomena that interest us most are the subjective experiences themselves, there is no way to carve anything off" (1992, p. 121). Following Searle, I believe it is true both that (a) pain's subjective character is its inescapable feature and, apparently unlike heat, (b) the subjective character is what interests us most. But it does not follow that (a) *because* (b). For all Searle says in defense of this pragmatic solution, it falls into the category of an all-too-frequent philosophical defense that we may label "just can't think of anything better." Wherever the explanation has nothing going for it other than the fact that it is the least bad thing we can think up, as philosophers I believe we are well-advised to remain agnostic. Indeed, we could just as easily have said that the reason why the subjective interests us is precisely that it is a necessary truth, and thus impervious to further explanation, that pains are felt. Although this explanation is avoided because it is an unfashionable appeal, at least it has the support of an impressive intuition. Still, even if it is mistaken, the evidence entitles us only to the combination of (a) and (b), not to the second accounting for the first.

Thus I bypass any further examination of irrealisms about consciousness. Let me turn instead to the most popular realist alternatives to emergentism. Of course, substance dualism is the realist view *par excellence*. I shall say something about it, but shall pass on quickly to what I take to be the main realist competitors to emergentism: various forms of physicalism. They are

only briefly sketched here to give us a background. After laying out the case for emergentism in chapters 2–5. They are examined more fully in part II.

1.8 Theories of the Mental II: Dualism

When one speaks of dualism in connection with the philosophy of mind, the original Cartesian variety first comes to mind. It is what I shall mean by plain 'dualism' unless specified otherwise. Historically it spurred the concern with mind that evolved into our current problematic. Dualists state that each person consists of a space-occupying material body plus an immaterial substance, the latter being a non-spatial "receptacle" containing the whole of one's mental life, though dualists customarily allow that *impure* mental states (e.g., answering thoughtfully, driving carefully) also involve bodily motions. On this account there are two major components of the mental realm: a substance and the transient aspects proper to it. Phenomenal and access-conscious states are not themselves immaterial substances, but items that, according to dualists, are capable of taking place within or with the cooperation of their immaterial substances. Pure mental states are not minds, but occur "in" minds. Various other names for immaterial substance have graced the literature, among them 'soul', 'psyche', 'spirit', and '*res cogitans*'.

For a dualist, 'mind' denotes something quite distinct from the brain or any other material entity. In spite of the metaphysical distinction between substances, Descartes believed that two-way traffic between them is rife. But problems of causation, cited earlier in connection with emergentism, recur. In addition to those, many have been baffled by the notion of something in a wholly immaterial realm affecting a state of affairs in the material world (e.g., raising one's arm), or even by brute matter influencing mental life (e.g., perceptual experience).

However, substance dualism plays a minor role in what follows. This is not primarily because of widely circulated misgivings about interaction between antipodal material and immaterial substances, but because the notion of such an immaterial substance, if it is not idle, strongly suggests that one's continued identity depends on the identity of one's soul; and this is extremely tenuous, as witnessed by insuperable difficulties over such a substance's identification, re-identification, and principles of operation.

The use of the word 'mind' in the sequel doesn't denote an immaterial substance, but refers to mental life in general, however analyzed.[16]

Another set of positions in the literature has been called "dualism," sometimes as a disparagement by physicalist opponents. Those views might also be labeled "dual aspect theory" and "property dualism." Proponents acknowledge that there is at most one substance, but maintain that in addition to physical properties, that substance also has irreducible mental properties. While this does not resolve the initial enigma about how the mental and the physical can causally interact, the difficulty is now subtler. The mental properties in question have a foot in the physical world; they may be properties of, say, the brain or its activity.[17] The contrast isn't as stark as it would be on Cartesian dualism. But neither is it an issue we can avoid because emergentism is a variety of dual aspect theory, as are certain forms of non-reductive materialism.

Emergentism may have been originally put forward as a more scientifically attuned alternative to dualism, but its scientific standing has since evolved; emergentism's present task is to find a slot alongside the physicalist alternatives that now dominate cognitive studies. Let's complete this rough taxonomy by looking at various physicalisms.

1.9 Variations on Physicalist Themes

A clear majority of cognitive scientists and mainstream philosophers of mind reject substance dualism. Most seem to converge on a loose and tolerant materialism. Beyond that, paths diverge. The basis for calling this consensus "materialist" is that these philosophers agree that our reality bottoms out in material world.[18] Perhaps not everything is explicable; but to

16. This summary doesn't cast a net wide enough to catch every fish. For example, Nida-Rümelin (2007), among others, combines substance dualism with emergentism. Those accounts offer bottom-up theories of experiential subjects as distinct from the subjects' bodies or brains, the latter two being incapable of having conscious properties. (See chapter 2 below.)

17. On certain assumptions about substance, perhaps we should opt for neutral monism. (See Schneider, forthcoming.) That will not require modifying our supervenience thesis.

18. It is hard to know how to classify neo-panpsychists, who speculate that we may find the primordial elements of consciousness at the same fundamental level as the particles of physics. Chalmers (1996, 2002) entertains that hypothesis.

the extent that we have well-grounded explanations, they will contain at least traces of their physical origins. Some hold that the direction of science inexorably points toward the bases of everything else being discoverable in interactions between the most fundamental particles, a thesis sometimes dubbed The Standard View. Stephen Weinberg summed it up in 1974 as follows: "at the present moment the closest we can come to a unified view of nature is a description in terms of elementary particles and their mutual interactions" (p. 50). And nothing in the past 35+ years has undermined a reasonable expectation that things will progress in that direction. Indeed, Sydney Shoemaker considers it an integral part of physicalism. He assumes "a physicalist view according to which all of the facts about the world are constitutively determined by . . . facts about the properties of basic physical entities and how they are distributed in the world" (2007, p. 33). But one need not go to those lengths to reject dualism or emergentism. Various forms of physicalism demand only that the mental be ultimately reducible to, identical with, or explicable in terms of something or other physical. However, for the views of current interest, materialist philosophers have left room for the singular qualitative, first-person features distinctive of our mental lives—that is, qualia. This form of physicalism embraces phenomenal realism.

With only rare exceptions, phenomenal realists who consider themselves physicalists share the view that each mental token is either identical with or fully explained by a physical token. Type physicalists maintain that these identities or explanations can be reductive; token physicalists hold that reductive accounts are not in the offing, although identities or explanations between instances of conscious properties and their bases are realizable. Others may take the supervenience of the mental—emergentism's thesis (1)—to be *sufficient* for physicalism. A physicalism making no additional claims remains compatible with what I am calling *emergentism*. If the supervenience or realization relation is brute (that is, admits of no further account beyond the fact of supervenience or realization), this unadorned explanation of why the mental is really physical is in most estimates disappointing; it falls short of delivering the unified ontology that has been a credo of the dominant strain of materialism. Thus, I concentrate for the most part on brands of physicalism that assert the identity of the mental and the physical, either reductive identity (as with the general run of type physicalisms) or non-reductive identities (as in token or non-reductive physicalism).

Henceforth I'll refer to the view that proposes an identity between mental and physical properties or states at the level of types as *plain*, or *type*, or *old-school* physicalism. As earlier, identity is a stronger requirement than is needed for physicalism, although an identity thesis is its commonest form. Strictly speaking, a physicalist need hold only that physical aspects are able to *explain*, *replace*, or on some versions *necessitate* mental ones. And mere explanation or necessitation has sometimes been regarded as adequate for reduction, a view explored more fully in chapter 8. But it is a delicate balance, perhaps too delicate, to find an intermediate sort of explanation that is strong enough for type physicalism's reductionist ambitions but not sufficient for identity. In fact, those problems mirror the ones that impair the role of the *bridge principles* central to earlier conceptions of theory reduction.[19] It has been observed that, if those principles fall short of stating identities, they are too weak to serve their purpose, and, indeed, that they presuppose an independent existence for the reduced theory and psychophysical laws relating it to the reducing theory.

However, even for physicalism with type identities, further distinctions are in order. One species, commonly known as analytic behaviorism, encounters what are widely regarded as insuperable difficulties; none of its varieties is generally taken to be a live option nowadays (as always in philosophy, ignoring the rare exception). However, a brand of old-school physicalism is still very much alive. It is a descendant of the earlier central-state materialism of Herbert Feigl and J. J. C. Smart, who claimed that we could achieve an identity between mental states or processes and those in the brain or other intrinsic bodily states. That view is the centerpiece of chapter 6. But there are also other forms of physicalist-leaning theories that deserve mention. I begin with a brief glimpse at *representationalism*, or, as it is known in at least one instance, *representationism*.

Representationalists hold, first, that all mental aspects, including phenomenally conscious ones, have intentional or representational contents, and, second, that conscious aspects are features of their contents, sometimes accompanied by limited additional factors. For example, the pain felt upon stubbing one's toe may have an intentional content representing in a particular way damage to or distress in one's toe. Another pain may represent, say, damage or distress to one's shoulder. On a teleological version of the

19. See, e.g., Kim 2000a.

thesis, pain may be designed—say, by evolution—to represent that a certain part of one's body requires attention. Many representationalists, though not all, are semantic externalists (or anti-individualists), holding that environmental, sociolinguistic, or evolutionary forces are central constitutive factors in those intentional contents. Moreover, not all representationalists are, strictly speaking, physicalists. Some settle for a more relaxed brand of naturalism. I list them here, first, because certain of their number consider representationalism as a doorway to physicalism, and, second, because even those who reject physicalism hold views incompatible with forms of property dualism. They are committed to rejecting the notion that the state exhibiting the intentional content has a distinctive intrinsic nature, a something it is like, independent of its content. That content, with perhaps a few minor additions, exhausts the nature of conscious properties. (Here again, Chalmers (1996, chapter 6) creates headaches for taxonomers.)

Two qualifications are in order. First, because the external world is wildly diverse, even physicalist representationalists can reject *typal* identities. While they all "reduce" qualia to intentional contents, some are uncommitted on further physicalist reductions. However, on the basic differences with emergentism, the representationalists of concern are one with physicalism. Second, it should not be supposed that all semantic externalists are physicalists, or even naturalists. For example, Tyler Burge (1979; 2007b,c) distances himself from physicalism and representationalism by denying that the sociolinguistic information definitive of our phenomenal states is tantamount to intentional content.

A popular view, making up yet another branch on our tree of physicalist options, is psychofunctionalism (plain 'functionalism' here). On it, mental properties are *defined* in terms of their abstract functional roles. If M is a (type or token) mental property, it is delineated via its typical relations to (a) stimuli, such as perception or hearsay, which give rise to M; (b) interactions with other mental properties—say, M', M'', M''', etc.—by way of inference, causation, and other associations; and (c) behavior, including reasonings, made more probable by (a) and (b).

On this account, M has been exhaustively characterized in terms of its extrinsic features, its relations with the environment, other mental properties, and the behavior in which it issues; no mention is made of any intrinsic (non-relational) features it may have. Because of this, the most favorable candidates for functionalization are the properties most receptive to dispositional components, such as beliefs and desires.

For conscious properties, standard functionalism appears at first glance to ignore the something it is like of conscious experience. If that is the final draft, functionalism is among the class of the eliminativist theories set aside for reasons aired in section 1.7. (Although it is useful to distinguish forms of reductionism, which that sort of functionalism would exemplify, from eliminativism, reductionism differs from eliminativism chiefly in name. As Chalmers (1996, p. 165) has noted, it does no more than retain the title 'experience' for its explananda whereas eliminativism does not.) However, that assessment of functionalism may be premature. Pioneering functionalists regarded their views as doing no more than introducing "topic-neutral" translations of mental terms in an attempt to overcome initial resistance to locating what may seem to be extra-physical in a physical framework. After surmounting this obstacle, a physicalist identification or definition is a natural next step. Recently the procedure has gone as follows: following functionalization, we identify the functional (mental) property with whatever physical realizer is responsible for that property's behavior (Lewis 1983c; Kim 2005). That form of functionalism is squarely physicalist.

On the other hand, there are philosophers who believe that our labors terminate with the functional definitions. That view resists physicalism. Physicalists have recourse to the foregoing considerations: What matters are not the definitions obtained through functional analysis, but the identities with whatever physical aspects fill the causal roles specified in the analysis. Those realizers of mental aspects defined by their functional roles are invariably physical. That is the sort of functionalism that has a place in our consideration of physicalist options.

This brief review shows that the different takes on the physical bases of the mental present a crowded and messy field. Nonetheless, we may identify major lines of inquiry and clearly delineate views in that broad category as distinct from the variety of emergentism under consideration, or, for that matter, from any other variety of it with which I am familiar. Two qualifications will help to round off this initial characterization of physicalism.

First, I bypass the question of what counts as physical. Various suggestions, ranging from the spatially extended to the inanimate to whatever is non-mental, have been offered. As Noam Chomsky famously remarked, the notion of a physical explanation is guaranteed to cover *all* explanation that becomes accepted "for an uninteresting terminological reason, namely that the concept of 'physical explanation' will no doubt be extended to incorporate whatever is discovered in this domain, exactly as it was extended to

accommodate gravitational and electromagnetic force, massless particles, and numerous other entities and processes that would have offended the common sense of earlier generations" (1972, p. 98). Here I simply rely on the fact that none of the physicalist-leaning views examined in this work hinge on the sorts of borderline cases that would compel us to be more precise about inclusion in the physical. (But see section 5.2.) Physicalists, functionalists, and representationalists alike tend to regard their views as offering reductionist, objective (that is, third-person) identifications of the mental. Put otherwise, upon successful completion of the view we are entitled to claim that a given mental aspect is "nothing but a(n) ___" (the blank to be filled in as directed by one's favored theory), in contrast with its being "something over and above" the physical. That will suffice to contrast those theorists with emergentists, and to entitle them to full participation in the discussions that follow.

Next, whereas any physicalist doctrine can limit its scope to a mere selection of mental aspects, our interest is in a comprehensive physicalism. The inability to include some conscious aspects in its identities or reductive definitions counts as a failure of the view. Such partial views are to be regarded as unsuccessful attempts to disguise counterexamples.

With the foregoing qualification in mind, we still need a representative version of physicalism to play off against emergentism. For that purpose I shall adopt old-school physicalism, which identifies kinds of mental phenomena with bodily states, etc., almost always spiking activity in the nervous system.[20] David Lewis (1999, p. 291) writes: "I am a realist and reductive materialist about mind. I hold that mental states are contingently identical to physical—and in particular, neural—states." Roughly, this is a view to which physicalists across the board might subscribe, although old-school physicalists may differ over whether the thesis is contingently or necessarily true. Moreover, old-school physicalists who pursue central-state identities or necessities may divide into chauvinists, who hold that only creatures whose anatomy resembles ours in certain respects have consciousness, or pluralists, who grant consciousness, but only in a different sense, to creatures anatomically or materially different from us. Details of that distinction are provided in chapter 6.

20. This expositional convenience is not intended to rule out externalism. Perhaps conscious states with contents have identity conditions referring to the subject's external environment or language community.

1.10 Non-Reductive Physicalism Contrasted with Emergentism

Thus far, the salient realist options still in view are emergentism, representationalism, and old-school physicalism. The last is a natural ally of forms of reductionism. However, it should also be noted that it is not mandatory in general to discard whatever gets reduced (see, e.g., Sober 1999). For example, we continue to determine amounts of heat and pressure with the Boyle-Charles law despite the absorption of thermodynamics into statistical mechanics. However, when we are discussing the reduction of mental to physical properties in philosophy, what stands out is not the ability to streamline our equations, but that the reduced class takes an ontological back seat. Thus, for the cases of current interest, an identity for a target class of Xs with Ys is intended to enable the claim that Xs are nothing but (or over and above) Ys. Some reductive physicalists hold, even more radically, that the identities are an important lemma in a demonstration that the relationship is not like that of thermodynamics to statistical mechanics, but rather like that of phlogiston to oxygen (viz., of alchemy to chemistry). Physical distinctions may then be viewed as "wip[ing] out familiar distinctions [between mental aspects] as spurious" (E. Nagel 1961 p. 340).

A point to bear in mind is that only identities between the generic or the abstract, such as laws or types of properties or entities, lend themselves to reductive analyses. And not even all of those hold out prospects for reduction—consider baby buggies and perambulators. Contemplating reduction is more prevalent where laws are directly involved, less so when what is identified are different names for the same general types. Thus, distinct names for a single species, such as 'The Grizzly Bear' and 'Ursus Horribilus' or 'Brontosaurus' and 'Apatosaurus', simplify one's ontology without thereby reducing either type to the other. Nevertheless, it appears that every identification that is fodder for reductionism is with a type or with the respective membership lists of types.

On the other hand, many physicalists are not committed to such type-level accounts. These are non-reductive or token physicalists. They reject identities between specific types of mental and physical phenomena, but accept a physicalist account of some sort at the level of tokens. Non-reductive physicalists—e.g., Davidson (1980b)—typically hold that there are token identities between mental and physical aspects, but some—e.g., Horgan (1993), Pereboom (2002), Wilson (2002), and Antony (2007)—appear

willing to settle for non-identity where the mental token can be robustly explained by a physical one.

Identities between individuals—say, between Charles Lutwidge Dodgson and Lewis Carroll, between Constantinople and Istanbul, or between The Crimean War and The Eastern War—do not lend themselves to reductions. They may commit their holder to only one thing where she might have previously taken there to be two, but the items in question provide no simplification or streamlining of one's ideology. If there are *reductions* at the level of particulars (say, 'That's light reflected on the stream, not a fish'), it isn't clear that they are identities. Moreover, if they were identities, they would not be those of interest for the notion of reduction in science and philosophy. Thus, the acceptance of individual identities between mental and physical instantiated properties, although incompatible with emergentism, doesn't by itself advance reductivist aspirations.

Like emergentism, token physicalism rejects identities between various mental and material property types and may hold that the mental is dependent on the material in ways to be spelled out by supervenience or realization. Indeed, token physicalists are sometimes erroneously taken, especially by critics, to be emergentists. Whereas token physicalists form a mixed collection, as a group they are distinguished from emergentists by being committed to one or both of the following:

(i) *Token identity*. Each token mental property is identical with some token physical property.

(ii) *Explanatory access*. Each mental property is explicable in terms of some physical property *in a way that goes beyond the fact* that the mental property supervenes on or is realized by that physical property.

A parallel disjunction may be constructed in terms of states. And some—e.g., Loewer (2007) and J. Wilson (2002)—may prefer to frame our second disjunct in terms of the physical's *necessitating* rather than *explaining* the mental.

In its commonest incarnation, token physicalism opts for (i) above (that is, token identity) or for both (i) and (ii). But, as (ii) demonstrates, one can be a token physicalist and reject all such identities. Earlier I mentioned problems for type physicalists who sought to replace identities with explanations or necessitations. Those particular difficulties need not vex token physicalism's exclusive appeals to explanatory relations; it bears only on views with reductionist ambitions. However, a token physicalist taking this route must hold not only that mental aspects supervene (or are otherwise

dependent on) physical ones, but also that we are able, or should eventually be able, to explain *why* the mental supervenes on the physical. That may engender a different set of problems.

The emergentist now under discussion (see (1)–(3)) rejects both (i) and (ii). Another way of putting the rejection of (ii) is to state that the relation of, say, conscious properties with their physical dependency base is brute or primitive. Samuel Alexander writes of emergent qualities as "under the compulsion of brute physical fact" (1920, volume I, p. 46). Of course, there are always further things to say about anything. But the emergentist holds that there is no further minimal explanation of the fact that mental state *m* depends on (or supervenes on, or is realized by) its physical base *p*.

These features make it easier to put in relief the differences between emergentism and non-reductive physicalism. Consider again my tripartite statement of emergentism for property E:

(1) E is *dependent* on different sorts of a non-emergent base in a way made manifest by E's *supervenience* (or *realization*) on those same properties.

(2) There is *no* further (minimal) *explanation* of why E is supervenient on (or dependent on, or realized by) that non-emergent base; viz., their relationship is brute.

(3) E is a cause (of both mental and physical aspects) in ways in which there is no sufficient cause in context at the levels of E's non-emergent base(s).

Both parties can accept (1), subject to the proviso that token physicalists may demur at regarding token identity as a kind of dependence. It is clear that emergentism parts ways with non-reductive physicalism at (2). Whereas (3) is an issue for emergentism, token-identity materialists have no need, or less of a need, to respond to criticisms of mental causation. If they accept (i), a mental cause is always identical with one or another physical cause. If they reject (i), explanations via (ii) in terms of physical causes can still be substituted for those in terms of mental causes, providing at least partial relief from the difficulties raised for mental causation.

1.11 Conclusion

I began with a brief history of emergentism, exploring both its ontological aspirations and the reasons for its general repudiation and subsequent

neglect. This classical emergentism was then contrasted with the restricted emergentism to be developed in these pages. Finally, I produced a quick sketch of the major theories that compete with emergentism, several of which will be explored in greater detail in part II. Thus, we now have before us a roughly sketched map of the issues. With that in hand, I turn to the task of elaborating emergentism and the beginnings of an explanation of why I take it to be a defensible view of mental life.