Metaphysical Emergence: Weak and Strong*

Jessica Wilson†

Draft: July 23, 2012

Introduction

Why care about what emergence is, and whether there is any? To start, many complex entities of our acquaintance—tornados, plants, people and the like—appear to be composed of less complex entities, and to have features which depend, one way or another, on features of their composing entities. Yet such complex entities also appear to be to some extent autonomous, both ontologically and causally, from the entities upon which they depend. Moreover, and more specifically, many "higher-level" entities (particulars, systems, processes) treated by the special sciences appear to be broadly synchronically dependent on "lower-level" (and ultimately fundamental physical) entities.[1] Yet, as is suggested by the associated special science laws, many higher-level entities appear also to be ontologically and causally autonomous, in having features in virtue of which they are *distinct from* and *distinctively causally efficacious relative to* the lower-level entities upon which they depend, even taking into account that the latter stand in various aggregative relations. An account of emergence making sense of these appearances would vindicate and illuminate both our experience and the existence and tree-like structure of the special sciences, as treating distinctively real and efficacious higher-level entities and their features.

Reflecting these motivations, nearly all accounts of emergence take this to involve both

*Thanks to Benj Hellie for comments.

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

[1]Talk of "higher-level" and "lower-level" entities is relative, and reflects the pre-theoretic and theoretic appearances. Here I treat as at the same "level" both individual entities treated by a given science, and certain combinations of such entities, where the allowable modes of combination include aggregations of relations which may hold between individual entities, as well as mereological and certain boolean combinations of such individuals or relational entities. So, for example, both atoms and relational entities consisting of atoms standing in atomic relations are taken to be at the same level, as are mereological or disjunctive combinations of atoms or relational atomic entities.

broadly synchronic dependence[2] and (some measure of) ontological and causal autonomy.[3] Beyond this agreement, however, accounts of emergence diverge into a bewildering variety, reflecting that the core notions of dependence and autonomy have multiple, often incompatible interpretations.

In particular: candidate conceptions of (broadly) synchronic dependence include composition (Mill 1843/1973, Stephan 2002); supervenience/necessitation (Broad 1925, van Cleve 1990, Kim 2006); causation or causal constitution (Mill 1843/1973, Searle 1992, O'Connor and Wong 2005); and functional or other realization (Antony and Levine 1997, Yablo 1992, Shoemaker 2001, Gillett 2002*a*). Candidate conceptions of ontological and causal autonomy are even more various. Metaphysical accounts of autonomy include ontological irreducibility (Silberstein and McGeever 1999, Kim 2006); novelty (Anderson 1972, Humphreys 1996); fundamentality of properties, powers, forces, laws (the British Emergentists, Cunningham 2001, O'Connor 2002, Wilson 2002); non-additivity (the British Emergentists, Newman 1996, Bedau 1997, Silberstein and McGeever 1999); "downward" causal efficacy (Sperry 1986, Searle 1992, Klee 1984, Schroder 1998); and multiple realizability (Putnam 1967, Fodor 1974, Klee 1984, Shoemaker 2001, Wimsatt 1996). And epistemological accounts of autonomy include in-principle failure of deducibility or predictability (Broad 1925 and other British Emergentists, Klee 1984); predictability, but only by simulation (Newman 1996, Bedau 1997); and lack of conceptual or representational entailment (Chalmers 1996, Van Gulick 2001). No surprise, then, that many recent articles on emergence are devoted mainly to taxonomizing its many varieties (Klee 1984, Van Gulick 2001, Stephan 2002).

Though in general a thousand flowers may fruitfully bloom, this much diversity is unuseful for purposes of illuminating the structure of natural reality. Different accounts often disagree on whether an entity is emergent; and when they agree, there is often no clear basis for this agreement. Hence it is said that references to emergence "seem

[2]Some 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 (see §3.3.1). O'Connor and Wong (2005) take emergence to be diachronic on grounds that emergent features are caused by lower-level features, and causation is diachronic; but the essentials of their account are preserved if (the relevant) causation is synchronic (see §2.3). And Rueger (2001)) takes emergence to be diachronic since involving temporally extended processes; but the emergence of such processes is compatible with these "synchronically" depending on a temporally extended base (compare spatiotemporally global supervenience). Humphreys (1997) characterizes an irreducibly diachronic emergence, involving the exhaustive (non-mereological) "fusion" of lower-level entities into another lower-level entity; but such emergence is besides the point of accommodating the existence of higher-level entities, and so will be set aside here.

[3]These core components are occasionally explicitly flagged (see Bedau 1997), but more typically are encoded in specific accounts of dependence and autonomy, as when Kim (2006, p. 548) says "two [...] necessary components of any concept of emergence that is true to its historical origins [...] are supervenience and irreducibility". Here and 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; 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 that we will later consider. 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 are occasionally presented as accounts of "emergence" (see Chalmers 2006), they are not so in the sense at issue here.

to have no settled meaning" (Byrne 1994, p. 206), that accounts of emergence are "not obviously reconcilable with one another" (O'Connor 1994, p. 91), and that "those discussing emergence, even face to face, more often than not talk past each other" (Kim 2006, p. 548). Moreover, and importantly for the relevance of emergence to contemporary debate, different accounts often disagree over whether emergence is compatible with *Physicalism*, according to which all broadly scientific entities are "nothing over and above" physical entities. So, to take just one example, Kim 1999 takes physical realization to be incompatible with emergence, while Gillett 2002*b* takes such realization to be required.

I'll argue here that much of this apparent diversity is superficial. I'll start by showing, by attention to the available responses to the problem of higher-level causation, that there are two and only two schematic conceptions of higher-level metaphysical emergence of broadly scientific entities: *Strong* and *Weak* emergence, respectively (§1). The two schemas are similar in each imposing a condition on the powers of entities taken to be emergent, relative to the powers of their base entities. *For purposes of appreciating the generality of the schemas, it is of the first importance to register that the notion of "power" here is metaphysically almost entirely neutral, reflecting commitment just to the plausible thesis that what causes an entity may potentially bring about (perhaps only contingently) are associated with how the entity is—that is, with its features.*[4] As I'll discuss, even a categoricalist contingentist Humean could accept powers in the weak sense at issue in the schemas. Though similar in each involving a condition on powers, the schemas are also crucially different—a difference reflected in the fact that (given the physical acceptability of the lower-level entities) one schema is compatible with *Physicalism* and the other is not. (The results here generalize to distinguish two basic forms of higher-level emergence from lower-level entities, whether or not the latter are physically acceptable.) I will then consider the main accounts of emergent dependence (§2) and emergent autonomy (§3 and §4), and argue that all such accounts intended as characterizing metaphysical emergence are appropriately interpreted as targeting one or the other schema. The two schemas thus unify and clarify the many apparently diverse accounts of higher-level metaphysical emergence, while explaining controversy over whether emergence is compatible with *Physicalism*.

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; see, for example, Smart 1981, Bedau 1997, Chalmers 2006, and Clayton 2006. My treatment goes beyond these (typically gestural) treatments in explicitly cashing the distinction between *Weak* and *Strong* emergence in metaphysical rather than epistemological 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 emergent dependence and emergent autonomy. My treatment also goes beyond previous taxonomic 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 synchronically emerge from lower-level such entities, and—along a critical dimension—to show that certain accounts have more work to do if they are to ensure satisfaction of the condition on (metaphysically neutral) powers requisite to the intended schema.

[4]Here and elsewhere, nominalists are invited to interpret talk of features (properties, states) in their preferred terms

1 Two schemas for emergence

1.1 The target cases

Accounts of emergence tend to focus on emergence of features (e.g., either tokens or types of properties or states) from lower-level features, it being supposed (as per the background contrast with substance dualism; see §2.1) that emergence of entities (systems, processes, particulars) may be understood in terms of emergent features.[5] The lower-level features are typically taken to be physically acceptable relational features—that is, physically acceptable features of relational lower-level (and ultimately physical) entities.[6] So, for example, a discussion of emergence might target the seeming autonomous dependence of . . .

- ∼ . . . the higher-level property/state (of a complex system) of being in the basin of a strange attractor, on the (lower-level, relational) property/state (of a system of molecules) of having parts with certain positions and momenta.
- ∼ . . . the property/state (of a plant) of being phototropic on the (lower-level, relational) property (of the plant's cellular walls) of being such as to undergo certain cellular wall weakenings and cellular expansions.
- ∼ . . . a mental property/state (of a person) on a (lower-level, relational) neurophysiological property/state (of certain neurons standing in certain neuronal relations).

Emergence, as applying to such cases, is treated (multiple dependence or realizability to one side) as a one-one relation between higher-level and lower-level features. This treatment presupposes that certain relational lower-level entities exist and have features serving as a dependence base for the associated emergent features. The presupposition is useful, in encoding (as had by the posited lower-level relational entity) the sorts of features of complex entities that are assumed by all parties *not* to be emergent, in any interesting sense. Alternatively, one might dispense with the relational lower-level middleman and take the dependence base to consist in collections of comparatively non-relational lower-level features (say, features of individual molecules and pairwise relations between individual molecules), understood as combinable via certain ontologically "lightweight" compositional principles, including additive causal combination (see §3.3), and certain boolean or mereological operations (see fn. 1 and §3.5). In any case, it's clear that the "one-one" and "many-one" approaches target the same phenomena: the latter considers the nature of the dependence of a higher-level entity on comparatively non-relational lower-level entities given certain allowable combinatorial principles, whereas the former considers the nature of the dependence of a higher-level entity on relational lower-level entities having features allowed by the combinatorial principles. By default I'll take the one-one perspective, but as we'll see some accounts of emergent dependence and autonomy take the many-one perspective.

[5] Hence Bedau (2002, p. 6) says: "[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] "Physically acceptable" here refers to entities and features that are (taken to be, in some or other sense) "nothing over and above" physical entities and features, where physical entities and features are, roughly and commonly, the relatively non-complex, not-fundamentally-mental entities and features that are the proper subject matter of fundamental physics (see Wilson 2006).

1.2 The problem of higher-level causation

The primary challenge to the supposition that higher-level entities may be dependent on yet ontologically and causally autonomous from lower-level entities is posed by the problem of higher-level causation, which gets started with the question: how can special science entities cause effects, given their strong synchronic dependence on lower-level (ultimately physical) entities?

First, some setup. I assume that the efficacy of entities lies in their having efficacious features; talk of entities themselves is thus suppressed. Moreover, 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 as reference to spatiotemporally located tokens (e.g., property instances, states, events) potentially of a type (property, state type, event type).[7]

Four of the premises leading to the problem concern special science features:

1. Special science features depend on lower-level physically acceptable relational features (henceforth, "base features") in that, at a minimum, special science features (at least nomologically) require and are (at least nomologically) necessitated by base features.

2. Both special science features and their base features are real.

3. Special science features are causally efficacious.

4. Special science features are distinct from their base features.

And two concern causation:

5. Every lower-level physically acceptable effect has a purely lower-level physically acceptable cause.

6. In general (apart from "firing squad" cases), effects are not causally overdetermined.[8]

There are two cases to consider, in each of which a special science feature S depends, on a given occasion, on base feature P (1). First, suppose that S causes special science feature S^* on a given occasion (compatible with 3). S^* is dependent on some base feature P^* (1), such that P^* necessitates S^*, with at least nomological necessity. Moreover, P^* has a purely lower-level physically acceptable cause (5)—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 (2, 4), S^* is causally overdetermined in a way (given 1) not appropriately understood in 'firing squad' terms, contra (6).

Second, suppose that S causes some base feature P^* (compatible with 3). P^*, being a lower-level physically acceptable feature, has a purely physically acceptable cause (by 5)—plausibly, and without loss of generality, P. So, it appears, both P and S cause

[7] That said, I will sometimes gloss the type/token distinction—e.g., when discussing necessitation of features (involving instances of the type at issue), below.

[8] The qualifier 'in general' reflects that effects may sometimes be causally overdetermined, as in so-called 'firing squad' cases.

P^*, and given that P and S are both real and distinct (by 2 and 4), P^* is causally overdetermined, again (by 1) in non-firing squad fashion, contra (6).

So goes the argument that real, distinct and efficacious higher-level features induce problematic overdetermination. Responses to the (valid) argument involve rejecting one of the six premises. The first four are as follows:

- ∼ *Substance dualism* or *Pan/proto-psychism*. Deny (1): avoid overdetermination by denying that S depends on physically acceptable P.
- ∼ *Eliminativism*. Deny (2): avoid overdetermination by denying that S and/or S^* is real.
- ∼ *Epiphenomenalism*. Deny (3): avoid overdetermination by denying that S is efficacious.
- ∼ *Reductive physicalism*. Deny (4): avoid overdetermination by denying that S is distinct from P.

None of these strategies makes sense of the seeming emergence of higher-level features: *Substance dualism* and *Pan/proto-psychism* fail to accommodate dependence; *Eliminativism* and *Reductive physicalism* fail to accommodate ontological autonomy; *Epiphenomenalism* and, some think, *Reductive physicalism*, fail to accommodate causal autonomy.

1.3 The two "emergentist" strategies

The remaining strategies do better by way of accommodating emergence. These are:

- ∼ *Robust emergentism*. Deny (5): avoid overdetermination by denying that every lower-level physically acceptable effect has a purely lower-level physically acceptable cause.
- ∼ *Non-reductive physicalism*. Deny (6): allow that there is overdetermination, but deny that it is of the "firing squad" variety that would be intuitively problematic as generally characterizing higher-level causation.

I'll now argue, for each of these strategies, that the strategy may be perspicuously understood as imposing one or another condition on the causal powers (henceforth, just 'powers') of a given special science feature, and that satisfaction of the associated condition provides a plausible principled basis for taking the feature to be emergent, in ways that proponents of each strategy would endorse.

1.3.1 A metaphysically neutral understanding of powers

Before getting started, let us ask: What are powers? 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 a particular 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.[9]

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 of a higher-level feature, with a token causal contribution associated with a token of a 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").[10]

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 S 1.3.3., 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 robust 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 in-

[9]For example, even a contingentist categoricalist Humean can accept powers in the neutral sense here: for such a Humean, to say that a (ultimately categorical) feature has a certain "power" would be to say that, were a token of the feature to occur in certain circumstances, a certain (contingent) regularity would be instanced. Of course, contemporary Humeans will implement more sophisticated variations on this theme.

[10]For example, suppose a contingentist categoricalist Humean wants to take a non-reductive physicalist approach to the problem of higher-level causation, and so aims (as I will expand on below) to identify every token power of a token higher-level feature with a token power of its lower-level base feature. As previously, such a Humean understands "powers" in terms of actual or potential instances of a (contingent) regularity. Where the aim is to avoid overdetermination, the Humean may suppose, to start, that the (relevant instances of the) regularities overlap, both with respect to the (single) effect, and with respect to the (single) circumstances in which the two token features occur. If the Humean is moreover a reductive physicalist, they may suppose that such overlap motivates identifying the token features at issue, and hence the associated powers. If the Humean is a non-reductive physicalist, they can reject this identification of features, on "difference-making" grounds (e.g., of the sort associated with Mill's methods). Such a Humean will suppose that attention to broader patterns of regularities can provide a basis for identifying token powers of token features, even when the token features are not themselves identical. Whether reductive or non-reductive, the contingentist categoricalist Humean can make sense of the claim that some, all, or none of the token powers of token features are identical. This case is like the case of New York: if we can make it (out) here, we can make it (out) anywhere.

dividuated 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. 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.

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 is transparent to such further metaphysical details, and so too should be—and are—our associated conceptions of emergence.

1.3.2 Robust emergentism

As above, the robust emergentist maintains that some special science features are real, distinct, and distinctively causally efficacious as compared to their (lower-level, physically acceptable) base features. Overdetermination is avoided by denying, contra all varieties of *Physicalism*, that every lower-level physically acceptable effect has a purely lower-level physically acceptable cause.

In terms of powers: the robust emergentist maintains either (a) that P does not have the power to cause E, or more plausibly, (b) that while P does have the power to cause E, this power is not identical with that had by S—it is manifested differently, or in different conditions (perhaps, e.g., P causes S, and S then more directly causes E). Implementing either (a) or (b) requires that the powers of the higher-level feature satisfy the following condition:

> *New powers condition*: Token higher-level feature S has, on the occasion in question, at least one token power not numerically identical with any token power of the token lower-level feature P on which S synchronically depends, on that occasion.

It is clear that satisfaction of this condition guarantees that S is both ontologically and causally autonomous from P: since S has a token power that P doesn't have, S is distinct from P (by Leibniz's law) and can do at least one thing that P can't do, or in any case cannot do in the same way as S.

1.3.3 Non-reductive physicalism

Like the robust emergentist, the non-reductive physicalist maintains that (some) special science features are real, distinct, and distinctively causally efficacious with respect to their base features. Problematic overdetermination is avoided by taking higher-level and base features to stand in a relation that, while not identity, is intimate enough to avoid overdetermination of the "firing squad" variety. 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, which plausibly avoids problematic overdetermination; these

include functional realization, the part/whole relation, and the determinable/determinate relation. This seeming diversity hides a deeper unity of strategy, however, which again can be put in terms of a certain condition on powers (see Wilson 1999 and forthcoming).

To start, the non-reductive physicalist maintains, as does the reductive physicalist, that every token power of S is identical to a token power of P. They moreover maintain that the token powers of S are a proper subset of the token powers of P, as per:

> *Subset condition on powers*: Token higher-level feature S has, on the occasion in question, a non-empty proper subset of the token powers of the token lower-level feature P on which S synchronically depends, on that occasion.

Satisfaction of the condition clearly blocks problematic overdetermination: when a power of S manifests in a given effect on a given occasion, there is only one causing (as between S and P), not two.

Satisfaction of this condition also guarantees conformity to *Physicalism*, compatible with both ontological and causal autonomy.

Let's start with conformity to *Physicalism*. To start, note that the recipe for avoiding overdetermination accommodates the core physicalist claim (premise 5, above) that every lower-level physically acceptable effect has a purely lower-level physically acceptable cause.

Moreover, imposition of the condition blocks all the usual routes to physical unacceptability. The main concern about physical acceptability turns on the possibility that a realized entity might be robustly emergent, such that, as above, either (a) S does not have the power to cause the effect E in question, or (b) that while P does have the power to cause this effect, this power is not identical with that had by S (it is manifested differently, or in different conditions). Satisfaction of the subset condition blocks both (a) and (b). Satisfaction of the condition also blocks the other live routes to physical unacceptability, associated with S's being non-natural (see Moore 1903) or supernatural: such designations plausibly require the having of non-natural or supernatural powers, which are ruled out by satisfaction of the subset condition (assuming, as we are, that the base feature P has no such powers).

Now, as it stands (and remaining broadly neutral on the metaphysics of features) satisfaction of the proper subset condition is compatible with S's having a non-causal aspect not had by P—say, a non-causal quiddity or an epiphenomenal quale. But, as discussed above, and as is reflected in the dispute between robust emergentists and physicalists, any non-causal aspects of S are irrelevant to broadly scientific goings-on: scientific truths do not in any way depend on or otherwise track whether scientific features have non-causal aspects (much less track how any such aspects are related). Hence that S has such aspects (whether or not shared by P) cannot undermine S's physical acceptability, given P's physical acceptability.

This point bears emphasizing, since many have supposed—following the usual suppositions of advocates of non-reductive realization as involving a proper subset condition on powers (e.g., Shoemaker 2001 and Clapp 2001)—that such an account of realization requires commitment to an account of features on which these are essentially or exhaustively individuated by their powers. Hence Melnyk (2006, pp. 141–143) suggests that unless features are identified with clusters of token powers, satisfaction of the proper subset condition will not guarantee conformity to *Physicalism*, since such satisfaction will not guarantee that physically realized entities are constituted by physical entities, or that

truths about physically realized entities are made true by physical goings-on. More specifically, Melnyk claims that if realized features have non-causal aspects, then even though an entity's having P entails that it has (bestowed upon it) the token powers associated with having S, it won't follow from satisfaction of the proper subset condition that the entity's having P constitutes its having S, or that the entity's having P (along with physical laws, etc.) makes S truly attributed to it. But this is incorrect: truths about physical constitution or truthmaking, being broadly scientific truths, are neutral as regards whatever non-causal aspects of features there might be; hence the grounds of such truths must also be neutral on whether properties have non-causal aspects. It follows that satisfaction of the proper subset condition suffices for conformity with *Physicalism* independent of whether states or features are exhaustively individuated by their associated powers.[11]

The general pattern, blocking any route to S's physical unacceptability, is as follows: if P is physically acceptable, and every power of S is identical with a power of P, then any causal aspects of S are guaranteed to be physically acceptable; non-causal aspects of S are irrelevant to M's physical acceptability; hence a realization relation satisfying the proper subset condition on powers guarantees S's physical acceptability, in conformity to *Physicalism*, independent of what account of properties one endorses.

Let's turn now to the question of autonomy. Satisfaction of the subset condition clearly accommodates ontological autonomy: if S has only a proper subset of P's powers, then S is distinct from P, by Leibniz's law. The strategy arguably also makes room for S's being causally autonomous, with the key idea being that causal autonomy does not require that S have a distinctive power. Rather, it is enough that S have a distinctive set (collection, plurality) of powers—that is, a *distinctive power profile*.

How might the having of a distinctive power profile suffice for causal autonomy? One case for this appeals to difference-making or other "proportionality" considerations, in cases where S (or S's type) is multiply realizable. Suppose S is a state of feeling thirsty, which causes an effect E—say, a reaching for a glass of water. Now suppose that S (or another instance of S's type, etc.) had been realized by P' rather than P. Would the (or a) reaching still have occurred? Intuitively, yes, because the "extra" powers possessed by P, in virtue of which it differs from P'—say, to produce a specific reading on a neuronal state detector—don't matter for the production of the (or a) reaching. Rather, all that matters for this are the powers associated with S. That S's distinctive power profile contains just the powers crucial for E provides a principled reason for taking S to be causally efficacious *vis-á-vis* E in a way that is distinctive from P's efficacy *vis-á-vis* E.[12]

Another case for taking distinctive power profiles to (at least sometimes) suffice for causal autonomy appeals to the connection between sets of powers and distinctive systems

[11]Similarly for Melnyk's other claim (138–140) that unless realized entities are identified with clusters of powers, the condition's satisfaction will not guarantee satisfaction of the "necessitation" condition, according to which a physically acceptable realized entity must (perhaps together with physical laws, etc.) metaphysically necessitate the realized entity: "Why should it? Why assume that along with possession of power-tokens of certain types there automatically comes possession of a property [...] that would have conferred them?" (140). Given that truths about broadly scientific entities are transparent to facts about non-causal aspects of entities, from an entity's possession of power-tokens of a type it does "automatically" follow that the entity has the feature, whether or not features have non-causal aspects.

[12]Note 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.

of laws (e.g., the special science laws governing entities of S's type). Plausibly, systems of laws track causal joints in nature. Correspondingly, S's distinctive power profile is indicative of a distinctive causal joint in nature. Causal joints may overlap—in particular, in respect of S's and P's power to cause effect E. Still, if the joints as a whole are different, this provides a principled basis for taking S to be distinctively efficacious *vis-á-vis* E, in that S produces E as part of a different system of laws than P.[13]

1.4 Strong and Weak emergence

The robust emergentist and non-reductive physicalist responses to the problem of higher-level causation are the only responses aiming to accommodate the metaphysical emergence—dependence with ontological and causal autonomy—of higher-level entities; and as just argued, there are cases to be made that satisfaction of either of the associated conditions on powers would fulfill this aim. Moreover, and independent of the specifics of the problem (in particular, independent of the shared assumption that the base entities are physically acceptable), attention to these conditions makes clear the relatively limited ways in which, most crucially, the causal (hence also ontological) autonomy of a higher-level feature *vis-á-vis* its base feature may be gained. To wit: the feature may (as per robust emergentism) have *more* powers than its base feature; or the feature may (as per non-reductive physicalism) have *fewer* powers that its base feature. Since complete coincidence of powers doesn't make room for causal autonomy, these routes to emergence exhaust the options.

We may thus take the responses as exhaustive representative bases for two schematic conceptions of metaphysical emergence. The first schema is that associated with robust emergentism:

> *Strong emergence*: Token higher-level feature S is strongly emergent from token lower-level feature P just in case (i) S synchronically depends on P, and (ii) S has at least one token power not identical with any token power of P.

(For simplicity, in presenting the schemas I suppress reference to occasions.) The first condition minimally specifies synchronic dependence; the second (effectively, *New powers condition*, above) captures the comparatively strong sense in which an emergent feature is causally, hence ontologically, autonomous *vis-á-vis* its base feature.

The second schema is that associated with non-reductive physicalism:

> *Weak emergence*: Token higher-level feature S is weakly emergent from token lower-level feature P just in case (i) S synchronically depends on P, and (ii) S has a proper subset of the token powers had by P.

Again, the first condition minimally specifies synchronic dependence; the second (effectively, *Subset condition on powers*) captures the comparatively weak sense in which an emergent entity is causally, hence ontologically, autonomous of its base entity.

Each schema encodes a different way in which a higher-level feature might be dependent on, yet ontologically and causally autonomous from, a base feature; and thus is promising, so far as accommodating the motivations for emergence is concerned. And

[13] See Wilson in progress for a fuller defense of this claim.

again, attention to the available responses to the problem of higher-level causation, and the associated relations between powers that might serve as a basis for dependent causal autonomy, indicate that these schemas encode the *only* options for characterizing the metaphysical emergence of higher-level, broadly scientific entities (henceforth, typically, just 'emergence').

Let's now turn to seeing how specific accounts of emergent dependence and emergent autonomy, properly disambiguated and interpreted, aim to conform to one or other schema. In what follows, I'll usually leave off the qualifier "aim to", since my primary goal is not to assess the success of these accounts for purposes of characterizing emergence, but to make explicit their underlying theoretical intentions for doing so. That said, as prefigured, my discussion will track certain concerns about whether a given account presently satisfies its aim. One final remark before getting started: reflecting the role of emergence in the physicalism debates, accounts of emergent dependence and autonomy frequently presuppose that the base entities at issue are physically acceptable; the morals to be drawn, however, are broadly independent of this presupposition.

2 Emergent dependence

Four accounts of emergent dependence are on offer: material composition, modal covariation, causation or causal dependence, and functional or other realization. The first two are not exclusive of the last two: effectively all accounts of higher-level emergence take both material composition and modal covariation to be some part of emergent dependence. Where accounts primarily differ, as we will see, is in the assumed strength of modal covariation, and (relatedly) in whether broadly causal or rather realization-based dependence is (tacitly or explicitly) assumed. As I'll argue, accounts of emergent dependence differing in these respects conform to either *Strong* or *Weak* emergence, respectively.

2.1 Material composition

Accounts of emergence typically suppose that special science entities (systems, processes, particulars) depend on lower-level, ultimately physical entities at least in that the former are exhaustively composed of the latter:

> 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 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. (Mill 1843/1973, p. 243)

> 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 [...]. (Stephan 2002, p. 79)

Indeed, the assumption of compositional dependence reflects the intended contrast with dualist accounts on which higher-level features depend on the existence of physically unacceptable entities (e.g., souls, entelechies, conscious or proto-conscious fundamental particles). Compositional dependence is, however, compatible with either *Weak* or *Strong* emergence, and indeed, with the absence of emergence, since it is a further question, concerning any exhaustively physically composed particular, what features it has and whether any of these are emergent in either schematic sense.

2.2 Modal covariation

A further common baseline assumption is that emergent features depend on base features in standing in certain relations of (at least nomologically) necessary covariation, reflecting that emergent features both require (for their occurrence) and are upwardly necessitated by base features. So, for example, Broad (1925) maintains that emergent features of a compound are functionally dependent on features of the compound's parts (54-5), and that emergent features are "completely determined" by such lower-level features, 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); van Cleve (1990) concurs that "an emergent property of w is one that *depends on* and is *determined by* the properties of the parts of w" (222). The holding of both directions of necessary correlation may be expressed by (a version of) supervenience (see Kim 1990) that I'll call "minimally nomological supervenience", according to which an emergent feature (at least nomologically) requires some base feature, and a given base feature (at least nomologically) necessitates any associated emergent feature.[14]

Understood as an asymmetric relation (see Kim 1998, 11), minimally nomological supervenience distinguishes reductive from emergent dependence. Without further specification, however, such a conception is compatible with either *Strong* or *Weak* emergence. Broad and other robust emergentists typically maintain that emergent features minimally nomologically supervene on base features. And the schema for *Strong emergence* makes sense of such claims: laws of nature, after all, express what broadly scientific entities can do—that is, what powers they have; hence if an emergent feature has a power not had by its base feature per (as *Strong emergence*), it is plausible to suppose that the features stand in some sort of nomological connection (see §2.3). Minimally nomological supervenience is also compatible with *Weak emergence*, for some relations satisfying *Weak emergence* (e.g., the determinable/determinate relation) entail that the higher-level entities supervene with metaphysical, hence with nomological, necessity (see §2.4).

It remains to consider whether strengthening of the modal covariation relations—pertaining specifically to the strength of upward necessitation—distinguishes *Strong* from *Weak* emergence. Indeed, many accept as characteristic of physically unacceptable emergence that emergent features would supervene with *only* nomological necessity on base entities, in contrast with relations (like identity or the determinable-determinate relation) which plausibly preserve physical acceptability. So, for example, Chalmers (2006) says,

> [C]onsciousness still supervenes on the physical domain. But importantly, this

[14]The notion of upward necessitation may be stochastic (see Kim 2006, p. 550); emergent dependence need not be deterministic.

> supervenience holds only with the strength of laws of nature (in the philosophical jargon, it is natural or nomological supervenience). (p. X)

Van Cleve (1990) similarly characterizes emergence of the sort intended to contrast with *Physicalism*:

> 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. (p. 222)

Though common, the supposition that *Strong* and *Weak* emergence contrast with respect to modal strength of dependence relation is problematic, for two reasons. First, a physically acceptable feature might supervene with only nomological necessity on a physically acceptable base feature. For example, the subset condition in *Weak* emergence could be satisfied even if features are essentially individuated by non-causal quiddities and only contingently associated with their actual powers. Second, a physically unacceptable feature might supervene with metaphysical necessity on a physically acceptable base feature (see Wilson 2005). This would be the case if, for example, a consistent Malbranchean God brought about certain higher-level features upon the occasion of certain lower-level features in every possible world; or if features are essentially constituted by (all) the laws of nature into which they directly or indirectly enter; or if some robustly emergent features are grounded in non-physical interactions, and all the fundamental interactions are unified.

These considerations lead to a dilemma for anyone aiming to distinguish physically acceptable from unacceptable emergence by appeal to modal correlations alone. Those characterizing robust emergence in terms of mere nomological supervenience sometimes reject counter-cases whereby *Strong* emergent features supervene with metaphysical necessity on base properties, as violating Hume's Dictum, according to which there are no metaphysically necessary connections between (wholly) distinct entities. As it happens, *Strong* emergent features need not be wholly distinct from base features (see Stoljar 2007), and in any case post-Humean reasons for believing Hume's Dictum are in short supply (see Wilson 2010*b*). But suppose that Hume's Dictum is accepted, and grant that it ensures that *Stong* emergent dependence holds with only nomological necessity. It remains, as per the first counter-case, that physically acceptable features might supervene on base features with only nomological necessity—if, as above, features are essentially individuated by non-causal quiddities, not powers. To block this case, non-causal quiddities must be rejected as individuated of powers. But—here's the dilemma—proponents of Hume's Dictum arguably must (and typically do) accept non-causal quiddities as essentially individuating features, since after all (as per their denial that there are no metaphysically necessary causal connections) they cannot take features to be essentially individuated by powers. The means of blocking the two counter-cases are thus incompatible with each other, in which case a modal characterization of the distinction between physically acceptable and unacceptable emergence cannot be maintained, even if one is willing to commit to certain controversial metaphysical theses.

Moving forward, it's worth noting that, though covariation accounts officially aim to characterize emergent dependence in purely correlation terms, they rely for their plausibility on the underlying contrast between certain nomological relations (e.g., causation) and

certain metaphysical relations (e.g., the determinable/determinate relation). The next two proposals each cash out emergent dependence by explicit appeal to such relations, so as to both plausibly and determinately target either *Strong* or *Weak* emergence.

2.3 Causation or causal dependence

Yablo (1992) notes "a subtle interpretive question about supervenience", according to which

> On the emergence interpretation, a thing's physical properties are metaphysically prior to its mental properties and bring them into being. To caricature emergentism just slightly, supervenience is a kind of "supercausation" which improves on the original in that supercauses act immediately and metaphysically guarantee their supereffects [...]. (pp. 256-7)

The suggestion that emergent dependence is in some sense causal rages back to Mill Mill (1843/1973), the father of British Emergentism. Here it is important to be clear concerning how emergent features are considered causally dependent on base features. Mill's discussion initially focuses on a distinction between "homopathic" and "heteropathic" effects of a composite entity, where the former but not the latter effects are broadly additive combinations of effects of the sort that would have been produced were the component entities acting separately. Such a conception of emergence aims ultimately to characterize emergent autonomy in terms of a failure of additivity of causal influences, where such failure, in turn, is criterial of the composite entity's having a new power (to produce the heteropathic effect); see McLaughlin 1992. Hence it is ultimately not (heteropathic) effects, but rather features of complex entities having powers to produce such effects, which are emergent by Mill's lights. (See §3.3.1.)

That said, the question remains whether emergence of such features might itself be a causal phenomenon. Indeed, there are two ways in which emergent features might be causally dependent on base features. First, base features might act as a synchronic nomologically necessary precondition for the operation, or coming into play, of certain nomological features—i.e., fundamental forces or interactions—associated in turn with new powers; even if the relation here is not causation as traditionally understood, it nonetheless involves broadly causal lawful dependence (see Wilson 2002). This is the sort of causal dependence that is generally operative in British Emergentist accounts. Second, base features might more straightforwardly cause emergent features, as some contemporary emergentists (O'Connor 1994, O'Connor and Wong 2005) suppose. The two approaches (causal dependence vs. causation) are close variants, with the primary difference being that, if one supposes that causation is diachronic, one might further suppose that emergence is diachronic (as do O'Connor and Wong); but the diachronicity of causation is controversial, and nothing important hangs on this issue.

In particular, independent of whether emergent dependence is synchronic or diachronic, a conception in terms of causation or causal dependence will make good sense of *Strong emergence*. Either way an emergent feature has powers different from its base features: if caused, because effects typically have powers different from those of their causes; if causally dependent, because the operation of new fundamental forces or interactions serves as a (perhaps partial) ground for the having of new powers. The precise nature of the ground

for the new powers varies depending on the preferred account of causal autonomy (see §1.2, §1.3, and §1.4).

Seeing how causation and causal dependence make sense of *Strong emergence* sheds light on Kim's (2006, p. 558) claim that "the emergence relation from $[P]$ to S cannot properly be viewed as causal". Kim asks, rhetorically, "How can there be a causal chain from [e.g.] pain to the hand motion that is separate and independent from the physical causal chain from the neural state to the motion of the hand?" (fn. 7). This would indeed be strange against the assumption of *Physicalism*, and the associated closure claim that every lower-level physically acceptable effect has a purely lower-level physically acceptable cause; however, the robust emergentist's strategy encoded in *Strong emergence* just is to deny the closure claim, rather maintaining that the production of some physically acceptable effect requires (the manifestation of) powers not had by any lower-level physically acceptable feature. That said, Kim is clearly right that causation and causal dependence cannot characterize physically acceptable emergence, since such a nomologically "generative" connection does not ensure that the powers of emergent and lower-level features stand in the proper subset relation requisite for *Weak emergence*.

2.4 Non-reductive realization

The second metaphysically robust notion of emergent dependence is in terms of realization. While there are many accounts of this notion, all have in common the aim of characterizing a realized entity as "nothing over and above" its realizing entity (or entities), compatible (given the physical acceptability of base entities) with *Physicalism*. Some physicalists moreover think that such nothing over-and-aboveness is compatible with a realized feature's being emergent. Hence Gillett (2002*b*) sees the project of establishing the possibility of emergence as "deeply interwoven with the project of vindicating non-reductive physicalism as a viable position" (p. 102).

A realization-based conception of emergent dependence is indeed well-suited for physicalist purposes, in that the standard accounts of realization each have understandings on which their holding guarantees satisfaction of the conditions of *Weak emergence*. Here I consider a representative sample.

First, consider a "functionalizability" account, according to which realized features are second-order features, having causal roles played by the lower-level features that realize them on a given occasion (see Putnam 1967, Fodor 1974, Papineau 1993, Antony and Levine 1997, Melnyk 2003, and others). Now, to be associated with a distinctive causal role is just to be associated with a distinctive set of powers; hence if the distinctive causal role of a realized feature is, on a given occasion, played by a lower-level realizing feature, every token power of the higher-level feature, on that occasion, will be numerically identical with a token power of the feature upon which it synchronically depends, on that occasion. This much suffices, as previously argued, for the physical acceptability of a functionally realized feature, as per *Physicalism*. Still, one might think that functional realization is incompatible with *Weak emergence*, on grounds that a functionally realized feature inherits *all* of the token powers of its realizing feature:

> A functional reduction of pain has the following causal and ontological implications: Each occurrence of pain has the causal powers of its neural realizer; thus if pain occurs by being realized by N, this occurrence of pain has the

> causal powers of N. [...] 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, p. 554).

Where a functional role may be played by multiple realizers, however there is a case to be made that a functionally realized feature has, on a given occasion, only a proper subset of the token powers of the feature realizing it on that occasion. To see this, recall the analogy initially motivating functionism (see, e.g., Putnam 1967), to cases where multiple hardware systems may implement the instructions associated with a given piece of software. Here the realizing systems are similar in each having whatever powers are needed to implement the software, but are different in having other powers associated with their distinctive hardware bases. More generally, in cases where a type of functionally characterized higher-level feature may be multiply realized, it is plausible that each of its realizing types will have all of the powers associated with its functional role, and more besides.[15] Correspondingly, a proper subset relation will hold between the powers of the realized type and those of any of its realizing types. This relation between powers will hold on any occasion of realization involving tokens of the types; hence an account of emergent dependence in terms of functional realization will conform to *Weak emergence*.

Second, consider powers-based accounts of realization (see Wilson 1999, Shoemaker 2001, Clapp 2001). On Shoemaker's account,

> 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).

Shoemaker moreover claims:

> 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 (pp. 78-9).

His motivations here parallel those used to motivate the same claim for functionally realized properties. In brief, higher-level features are associated with distinctive sets of powers; if such a feature is multiply realized, then its realizing types will share the powers of the realized type, but will differ in respect of further powers. This relation will plausibly hold on any occasion of realization of tokens of the types; hence an account of emergent dependence in terms of powers-based realization (and the associated "part-whole" accounts of type or token features that Shoemaker and Clapp associate with their accounts) will conform to *Weak* emergence.[16]

Finally, consider accounts of non-reductive realization in terms of the determinable/determinate relation (see Yablo 1992, Wilson 2009), the relation of increased specificity paradigmatically holding between colors and their shades. Plausibly, as is reflected in Yablo's claim

[15] See, e.g., the discussion in Antony and Levine (1997).

[16] Note that nothing in the preceding line of thought requires acceptance of any particular account of the metaphysics of properties. As previously, the physical acceptability of a higher-level feature hinges solely on the relations between its token powers and those of its base feature on a given occasion; as such, issues of physical realization are independent of whether features have non-causal quiddities; and one may correspondingly also maintain that issues of physical realization are independent of whether the actual powers of a given feature are essentially or exhaustively individuative of it.

that determinables and determinates "are not causal rivals" (p. 259) and as is developed in Wilson 1999 and 2009, every token power of a determinable feature is identical with a token power of the more determinate feature(s) upon which it synchronically depends. Again, one might be concerned that such an account of realization is incompatible with *Weak emergence*, on grounds that instances of determinables and associated determinates are token-identical (see MacDonald and MacDonald 1986 and Ehring 1996); for in that case a higher-level feature will inherit *all* of the token powers of the feature that realizes it on that occasion. But here too, there is a case to be made that instances of determinables have only a proper subset of the token powers of the features that determine them on a given occasion. Plausibly, a given determinable will be associated with a distinctive set of powers; moreover, this determinable will typically be "multiply determined" by associated determinates; distinct determinates of the determinable will share the powers of the determinable, but will differ in respect of other of their powers. Moreover, insofar as determinables are distinctively unspecific, this characteristic should be preserved in their instances; but if a determinable token is identical with a determinate token on a given occasion, the former have all the token powers of the latter, and this distinctive lack of specificity will be lost. This provides another reason to suppose that a determinable token will have only a proper subset of the powers of their associated determinate token on any given occasion, in conformity with *Weak emergence*.

2.5 Results

We have arrived at the following results concerning accounts of emergent dependence:

- ∼ Conceptions of emergent depenent in terms of material composition are compatible with either *Weak* or *Strong* emergence, as well as with ontological reduction.
- ∼ Conceptions in terms of asymmetrical minimally nomological supervenience rules out ontological reduction, and is compatible with either *Weak* or *Strong* emergence.
- ∼ Conceptions in terms of mere nomological supervenience aim to conform only to *Strong emergence*, and conceptions in terms of metaphysical supervenience aims to conform only to *Weak emergence*; however, there are cases to be made that either strength of modal correlation is compatible with either schema; blocking all the cases require endorsing controversial theses (the rejection of quiddities, Hume's dictum) which moreover appear to be incompatible.
- ∼ Conceptions in terms of causation and causal dependence aim to conform to *Strong emergence*.
- ∼ Conceptions in terms of realization aim to conform to *Weak emergence*.

3 Emergent autonomy: metaphysical conceptions

I turn now to considering metaphysical accounts of emergent autonomy in light of the two schemas for emergence.

3.1 Ontological and causal autonomy

Causal autonomy (distinctive causal efficacy) guarantees ontological autonomy (distinctness), by Leibniz's law. But for reasons previously noted, ontological autonomy does not guarantee causal autonomy. Causal autonomy is necessary, however, for vindicating the ontological *and* causal autonomy of special science entities, and relatedly, for solving the problem of higher-level causation in a way preserving both the dependence and the distinctive causal efficacy of higher-level entities. Hence an account of metaphysical emergence aiming to accomplish these goals must do so in virtue of causal differences between higher-level and base features, rather than in virtue of any "bare" ontological differences there may be between these features.

This observation is crucial in appropriately interpreting accounts of emergent autonomy. Consider, for example, the conception of emergent entities as being "new" or "genuinely novel" with respect to their base entities:

> [Emergence involves] a new kind of relatedness (Morgan 1923, p. 19)
>
> [Emergence involves] a new quality [...] distinctive of the higher-complex (Alexander 1920, p. 45)
>
> [A]t each new level of complexity entirely new properties appear (Anderson 1972, p. 393)
>
> What seems to be central to our conception of emergent phenomena is the idea that something genuinely novel is present in the emergent entity that is not present in entities that are prior to it (Humphreys 1996, p. X)

All such conceptions need to make explicit that the novelty/difference at issue has causal as well as ontological implications. Note that mere adherence to Alexander's Dictum (a.k.a. the Eleatic Principle)—that real (broadly scientific) properties have powers—will not in itself establish that a novel/different feature has the desired causal autonomy. A stronger conception of emergent autonomy is needed, establishing that novel/different emergent features have either powers or power profiles different from those had by their base features.

Here I will consider five common ways in which emergentists fill in the notions of novelty/difference so as to gain causal as well as ontological autonomy, by appeal to: (1) fundamental powers, forces, laws; (2) non-additivity of effects; (3) downward causal efficacy; (4) imposition of lower-level constraints; and (5) multiple realizability and its variants. As we will see, individual variants on these strategies aim to characterize emergent autonomy as involving either *fundamental* or *non-fundamental* novelty/difference, along lines encoded in *Strong* and *Weak* emergence, respectively.

3.2 Fundamental powers, forces, laws

The notions of ontological novelty or difference are sometimes supplemented by appeal to fundamentality (ontological basicness):

> A fundamental property is an ontologically basic property of a basic entity [...] An ontologically-emergent property is an ontologically basic property of a complex entity. (Cunningham 2001, p. S67)

> [Emergence involves] a fundamentally new kind of feature. (O'Connor and Wong 2005, p. 665)

An appeal to fundamentality is in the right direction, but still does not make the requisite causal implications explicit, since a feature might be fundamentally new in having a fundamentally new non-causal quiddity.

It is appropriate, then, that accounts of emergent autonomy as involving fundamental novelty/difference more specifically take this to involve fundamentally new powers, forces, or laws. Such conceptions are characteristic of British Emergentism 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, p. 52. As McLaughlin goes on to note, these powers were typically taken to be powers to "generate fundamental forces not generated by any pairs of elementary particles" (p. 71). Relatedly, British Emergentists commonly took emergent features to be governed by fundamental laws (tracking or otherwise associated with the having of new powers to produce fundamental forces, etc.). Hence 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. (Broad, 64-5)

Appeal to fundamentally new powers, forces, or laws is similarly a theme in contemporary accounts of emergent autonomy. So, for example, Silberstein and McGeever (1999) understand emergent features as having irreducible causal capacities (that is, fundamentally new powers):

> 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. (p. 186)

O'Connor and Wong (2005) similarly make explicit that emergent features are "fundamentally new" specifically in having 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. (p. 665)

And reflecting that powers are plausibly grounded in fundamental forces/interactions, Wilson (2002) offers a fundamental interaction-relative account of emergence, according to which (in present terms) a dependent higher-level feature S is emergent from its base feature P, relative to a set of fundamental interactions F, just in case S has a token power different from any token powers of P grounded only in forces/interactions in F.

Accounts on which emergent autonomy involves fundamentally new powers, forces or laws all conform to *Strong*, and not *Weak* emergence. Accounts on which emergent features have fundamentally new powers explicitly do so, and the other accounts implicitly do so, since these accounts entail that emergent features will have new powers to generate fundamental forces/interactions, and in virtue of which they will enter into fundamental laws.

3.2.1 The flip side: failure of realizability

Under the rubric of emergent autonomy as involving fundamental powers, forces, or laws we may also place negative conceptions of emergent autonomy as involving a failure of realizability. So, for example, Kim (2006) identifies "irreducibility of emergents" as a necessary condition of emergence, where this is understood in terms of failure of functional realizability:

> Property M is emergent from a set of properties $N_1, \ldots, N_n$ only if M is not functionally reducible with the set of the Ns as its realizer. (p. 555)[17]

As above (§2.4), standard accounts of (non-reductive) realization all guarantee satisfaction of the condition, in *Weak emergence*, that the token powers of emergent and base features stand in the proper subset relation. Putting aside epiphenomenalism, then, an account of emergent autonomy as involving failure of (any such account of) realization will entail that an emergent entity has a new power, as per *Strong emergence*.

3.3 Non-additivity

Mill characterized emergent autonomy in terms of a failure of causal additivity. As we'll shortly see (§3.3.1), in the British Emergentist tradition such appeals are aimed at providing a (negative) metaphysical criterion for fundamental powers (and associated forces or laws); such conceptions of emergent autonomy thus conform to *Strong emergence*. As we'll also see, however, certain contemporary understandings of non-additivity, as grounded in non-linearity associated with, e.g., chaotic dynamical systems (§3.3.2), or in powers that latently exist at the microphysical level (§3.3.3), have been associated with *Weak emergence*. I'll address each of these approaches, in turn.

3.3.1 Non-additivity as a criterion for fundamentality

As previously discussed, Mill ("On the Composition of Causes", 1843/1973) distinguishes two types of effects of joint or composite causes. "homopathic" effects 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 ended up, had the forces operated sequentially. "Heteropathic" effects violate the principle in not being mere sums in the previous sense, and are therefore indicative of the operation of new laws. Mill says:

> 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. (p. 408–409)

[17] Note that Kim here, somewhat uncharacteristically, takes the 'one-many' perspective on emergence.

And he offers chemical compounds and living bodies as entities that are capable of producing heteropathic effects.

Mill did not use the term "emergence" (evidently Lewes 1875 first did so), but his notion of heteropathic effects serves as a basis for characterizing *Strong emergence*. To start: given the reciprocal connection between powers and effects, to say that an effect of a feature of a composite entity is non-additive, relative to effects of features of the parts acting separately, is just to say that the higher-level feature has a power not had by its lower-level base features when in additive combination (taking the many-one perspective) or, equivalently, that the higher-level feature has a power not had by its relational lower-level base feature (taking the one-one perspective). 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 (p. 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 non-additivity 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" (p. X). All this is in conformity with *Strong*, and not *Weak*, emergence.

Most other British Emergentists followed Mill in characterizing emergent autonomy 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. An interesting exception to this rule is found in Lewes' (1875) characterization of emergent autonomy as involving any failure of "general mathematizability", with emergence being correspondingly harder to come by. As in Mill's case, and following the standard British Emergentist conception of emergent autonomy as involving fundamental powers, forces, or laws, these appeals to non-additivity are best seen as attempts to provide a substantive metaphysical criterion of fundamentality, in conformity with *Strong*, and not *Weak*, emergence.

3.3.2 Non-additivity and non-linearity

Though British Emergentists saw non-additivity as characteristic of *Strong emergence*, some contemporary accounts of emergent autonomy (see Newman 1996 and Bedau 1997) take non-additivity of the sort associated with non-linear features of complex systems (e.g., being in the basin of a strange attractor) as motivating a conception of emergence compatible with *Physicalism*. What accounts for this discrepancy in the status as physically acceptable, or not, of non-additive higher-level features?

We should start by noting that certain motivations for taking non-linear phenomena to be physically acceptable do not establish this claim. Newman (1996), for example, cites the fact that complex systems are "strictly deterministic" in support; but strict determinism of non-linear systems does not rule out such systems as being *Strongly* emergent, for in the first instance such determination is a matter of nomological necessity, and as previously, all emergentists agree that emergent features (and associated powers to produce systemic behaviors) are (at least) nomologically necessitated by base features. Relatedly, that macro-states of non-linear systems are "derivable" from non-linear equations and initial (more generally, external) conditions does not establish physical acceptability, since it remains to consider the metaphysical basis for non-linearity (and associated equations). Bedau (1997) claims that features of non-linear systems are physically acceptable because "structural" (effectively: are features of relational lower-level entities); but given that non-linear phenomena do not consist solely in additive combinations of micro-level goings-on, the claim that such features are merely structural needs to be established, not assumed. What is needed to warrant taking non-linear phenomena to be physically acceptable is specific attention to the metaphysical basis for the non-linearity, and some argument to the effect that this basis does not involve new fundamental powers (or associated forces/interactions or laws).

Along these lines, it is worth noting that some accounts of the metaphysical basis for non-linearity are compatible with *Strong emergence*, contra *Physicalism*. Consider, for example, cases where the non-linear phenomena involves feedback between the micro-entities constituting the base, associated with strange attractors and other dynamic phenomena. As Silberstein and McGeever (1999) note, one metaphysical account of non-linearity (again compatible with "strict determinism") appeals to 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 non-linearity found in chaotic systems. [...] But why is non-linearity so central? [...] Non-linear relations may be an example of what Teller calls 'relational holism' [...]. (p. 197)

As above (§3.2), Silberstein and McGeever take the associated holism as indicative of emergent features' possessing fundamentally new powers ("irreducible causal capacities"). Such an account of the metaphysical basis of non-linear emergence is again in line only with *Strong emergence*.

Proponents of non-linearity as characteristic of *Weak* emergence have a different interpretation in mind, typically illustrated by attention to one or more specific examples (often involving cellular autonomata). The general moral to be drawn from these examples is that (*pace* traditional appeals to failures of additivity) a metaphysical account of non-linearity need not involve fundamental higher-level powers or laws, but rather only micro-level goings-on (notwithstanding that the aggregative result of such micro-interactions can be very surprising), compatible with *Physicalism*.

Granting this moral, a remaining, underappreciated, and more serious problem for taking non-linearity as characteristic of *Weak emergence* concerns whether the higher-level features at issue are plausibly understood as being ontologically and causally autonomous from their base entities, in having only a proper subset of the powers of their base entities. Indeed, both Newman and Bedau maintain that non-linear features are in-principle

reducible to micro-level phenomena, though Bedau attempts to ground a measure of higher-level autonomy in certain broadly metaphysical constraints on the predictability of non-linear and other phenomena supposed to instance weak emergence. We'll consider Bedau's account of such autonomy down the line (§4.2). Here I want to focus on another aspect of non-linear phenomena, also noted by Bedau:

> [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. (p. 395)

That non-linear phenomena associated with complex dynamical systems give rise to "simple, general macro-level patterns" may indeed provide a basis for the ontological and causal autonomy of the associated higher-level features, compatible with *Physicalism*, quite apart from how such patterns may be discovered.

Here we are motivated to attend to a second way in which higher-level phenomena may be ontologically novel or different—namely, as being *non-fundamentally* novel or different. And given that this form of difference must have causal implications, if it is to be characteristic of emergence, the strategy for establishing that features entering into higher-level patterns have the desired form of autonomy is clear: one must establish, first, that the macro-level patterns are different from (in being, plausibly, more "general" or less "specific" than) those at the micro-level, and second, that the correct account of this difference entails that the target (token) higher-level features have only proper subsets of the powers of their (token) base features, as per *Weak emergence*. One strategy for establishing that the requisite proper subset relation is in place might appeal to the higher-level features' being functionally or otherwise multiply realizable, and so having causal roles that are indeed more general than those of their realizers, in being associated with fewer of the latter's powers. Another strategy, which I will discuss in §3.5, may be implemented even if a given non-linear feature is only singly realizable.

In any case, proponents of non-additivity as a basis for physically acceptable emergence need to establish that the requisite autonomy is in place, and, it seems clear, should dispense with claims of in-principle ontological and causal reducibility. Such claims of reducibility may be motivated by thinking that in-principle ontological reducibility is required for *Physicalism*; but this motivation is suspect, given the seeming viability of the non-reductive physicalist's strategy for resolving the problem of higher-level causation, encoded in the schema for *Weak emergence*.

3.3.3 Non-additivity and micro-latency

Yet another understanding of the source of non-additivity is as involving the manifestation of powers that are existent, but latent, at the micro-physical level. For example, Shoemaker (2002) distinguishes between "micro-manifest" and "micro-latent" powers of lower-level entities, and suggests that emergent features have ("Type-2") powers that are latent at the micro-physical level:

> When micro-entities are combined in an emergence engendering way, the resulting object will apparently have two sorts of micro-structural properties. One sort, call these provisionally Type-1 micro-structural properties, will consist of properties that 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. [...] The other sort, which I will provisionally call Type-2 micro-structral properties, will be properties that are specified in terms of *all* of the powers, micro-latent and micro-manifest, of the constituent micro-entities. [...] Type-2 micro-structural properties, although they are micro-structural, will be emergent properties. [...] 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. (p. 55)

The underlying suggestion here is that, while emergent features may be non-additive (have powers to produce non-additive effects) relative to micro-manifest powers, this need not impugn their physical acceptability; Gillett (2002*b*) offers a similar account as "vindicating non-reductive physicalism as a viable position" (p. 102). Interestingly, Shoemaker traces the suggestion to Broad (1925), who seems to have taken the view that the powers of emergent features are micro-latent as a variant on the view that emergence involves violation of composition laws and associated coming into play of 'trans-physical' laws, as per *Strong*, and not *Weak*, emergence. So here again the question arises whether emergent autonomy as involving non-additivity is or is not compatible with *Physicalism*.

In answering this question, we should first note that the mere existence of micro-latent powers does not suffice to render emergent features physically acceptable, for proponents of *Strong emergence* will generally agree that in some broad sense physical entities have latent powers to bring about emergent features:

> [I]t is true in an emergentist scenario that everything that occurs rests on the complete dispositional profile of the physical properties prior to the onset of emergent features. For the later occurrence of any emergent properties are contained (to some probabilistic measure) within that profile, and so the effects of the emergent features are indirectly a consequence of the physical properties, too. [...] The difference that emergence makes is that what happens transcends the immediate [...] interactions of the microphysics. (O'Connor and Wong 2005, p. 669)

Such a weak dispositional understanding of micro-latent powers is compatible with micro-goings-on' being preconditions for the occurrence of new fundamental powers, forces/interactions, or laws at the higher-level, contra *Physicalism*. Indeed, Broad's assumption that emergence had anti-materialist implications indicates that he had such a weak dispositional sense in mind, in allowing that micro-physical entities have latent powers that become manifest when in "emergence-engendering" combinations. Physicalist proponents of micro-latent powers as a metaphysical basis for failures in additivity thus need to identify a more substantive understanding of micro-latency, which blocks a *Strong* emergent reading of apparent failures of additivity.

The prospects for their doing this are unclear, however. To start, it isn't enough to specify, as Shoemaker does, that the effects of micro-level dispositions also be micro-level, since this is compatible with the conditions of manifestation of the micro-level disposition involving physically unacceptable goings-on. Gillett (2002*b*) more explicitly recognizes the concern, and attempts to block it, as follows:

> In our broached scenario [...] the fundamental micro-physical properties have such conditional powers which they contribute conditionally upon instantiating certain realized properties. In such a case, a realized property instance thus determines that one of its realizer properties contributes a certain power that it would not otherwise contribute. It is important to mark the non-causal nature of the determination exerted by the realized property in such a scenario, for this suggests that there will likely be no new ontologically fundamental forces (or other properties). The relevant realized property instance, ''*H*', is not causing a microphysical property instance, '*P*', to contribute certain powers. Causal relations typically are mediated by forces and/or the transfer of energy—thus if *H* causally determined *P*'s contribution of powers then there might well be a new force. But in the scenario, *H* is exerting a non-causal determinative influence.

One problem here is that, even on the *Strong* emergentist interpretation of non-additivity, the weak dispositional micro-latent powers will not be "caused" by the higher-level feature: new fundamental interactions do not cause, but rather enter into constituting, the new powers for which they serve as a (perhaps partial) ground. The deeper problem, however, is that Gillett is stipulating that, rather than explaining how or why, the powers occurrent in "emergence-engendering" combinations might not involve (either causally or constitutively) any fundamental higher-level interactions or the like. The question remains: how are we to make sense of the claim that such powers are compatible with *Physicalism*, given that these powers do not make an appearance in the laws of fundamental physics (which deals only with micro-entities in relatively non-complex combination) and given that they cannot be understood as additive combinations of powers which do make such an appearance?

What the proponent of micro-latency needs to do in order to establish that non-additivity is compatible with *Physicalism* is to make a case that fundamental physical laws *might themselves entail violations in higher-level composition laws* when micro-entities enter into emergence-engendering combinations. It is unclear how this can be established however, since composition laws (incorporating, e.g., scalar and vector addition, along with other "ontologically lightweight"—boolean, mereological—modes of combination plausibly preserving physical acceptability) appear to exhaustively encode the broadly additive ways in which micro-manifest entities might combine while preserving physical acceptability. At the very least, at present it remains unclear how (metaphysical) non-additivity might fit with the usual understanding of *Physicalism* as the thesis that all broadly scientific goings-on are nothing over and above the goings-on explicitly (and not just "latently") at issue in fundamental physics.

Relatedly, there is reason to avoid characterizing physically acceptable emergence in terms of micro-latent features. Traditionally, the dispute between physicalists (of all stripes) and robust emergentists has turned on whether or not all broadly scientific goings-on are "nothing over and above" goings-on that are *manifest* at the micro-level, when

micro-entities are not in "emergence-engendering" combinations. From this perspective, a characterization of *Weak emergence* as involving micro-latent powers is not in the spirit of *Physicalism*. As Clarke (1999) notes, if higher-level features have token powers not identical with those of their base features ...

> ...emergent causal powers would be due to (bestowed by) some macro-level, structural properties possessed by the complex object [...] It matters little whether the macro-level properties that are acknowledged to carry emergent powers are said to be physical properties or whether the emergent laws are said to be physical laws; if there are emergent powers, then the kind of micro-explanation that is the ambition of most physicalists, an explanation of the behavior of all objects in terms of micro-level properties and relations and micro-level laws, will be impossible. (p. 309)

As such, it is no surprise that Broad did not feel the need to rule out the micro-latent interpretation in taking apparent violations of composition laws to have anti-materialist implications.

3.4 Downward causal efficacy

Many accounts of emergent causal autonomy require that such autonomy be specifically with respect to lower-level goings-on. Hence Morgan (1923) says:

> But when some new kind of relatedness is supervenient (say, at the level of life), the way in which the physical events which are involved run their course is different in virtue of its presence—different from what it would have been if life had been absent. (p. 15)

In a series of papers, Sperry (1969, 1986, 1976) suggests that conscious mental phenomena are emergent in causally affecting underlying neurophysical states, as does Searle (1992). More generally, as Kim (2006) observes, "downward causation is of paramount importance to the emergentists. For they want to claim that the emergence of consciousness and rational thought has made a fundamental difference to the world at the physical level" (p. 558).[18]

Unclarity over whether downward causation is compatible with *Physicalism* is a main source of unclarity over whether emergence is compatible with *Physicalism*. There is, perhaps, a *prima facie* appearance of incompatibility:

> Of all the marks of emergence [downward causation] is the one which presents the clearest and most direct challenge to micro-determinism. (Klee 1984, p. 58)

On the other hand, commentators disagree—sometimes over a single account—over whether downward causation is so incompatible.

Sperry's account is a nice case in point. On the one hand, Sperry (1976) speaks of downward influence as involving higher-level powers:

[18] That said, Kim thinks that the supposition of downward causation is problematic, for reasons we will consider in §3.4.1.

> The conscious subjective properties in our present view are interpreted to have causal potency in regulating the course of brain events; that is, the mental forces or properties exert a regulative control influence in brain physiology. (1976, 44)

McLaughlin (1992) interprets such talk as committing Sperry to an account involving fundamental configurational forces, hence as incompatible with *Physicalism*. On the other hand, Sperry (1969) describes downward influence as analogous to that involved when the atoms in a wheel must go where the wheel goes:

> The subjective mental phenomena are conceived to influence and to govern the flow of nerve traffic by virtue of their encompassing emergent properties. Individual nerve impulses and other excitatory components of a cerebral activity pattern are simply carried along or shunted this way and that by the prevailing overall dynamics of the whole active process (in principle—just as drops of water are carried along by a local eddy in a stream or the way the molecules and atoms of a wheel are carried along when it rolls downhill . . . " (p. 532)

Schroder (1998) (following Klee's 1984 suggestion) interprets this analogy as suggesting that downward causation involves not new powers, but lower-level constraints: "we can see what is wrong with a critique of emergentism that castigates it for assuming 'configurational forces' [. . .]. Emergentists who adopt downwards causation as a criterion for emergent properties need assume no such force. [. . .] In order to produce live and mindful beings, what is needed is not special laws but special structures that constrain the sequence of possible events in special ways" (p. 449). Searle's (1992) account of "radical" emergence, which is supposed both to involve new powers and to be no more physically problematic than, say, liquidity, has produced a similar degree of interpretive confusion.

A plausible diagnosis of this confusion reflects that there are two ways for a higher-level feature to be downwardly causally efficacious: one conforming to *Weak emergence* and one conforming to *Strong emergence*. Confusion concerning Sperry's and Searle's accounts is then plausibly located in these authors' failing to sufficiently disambiguate which form of downward causation (hence of emergence) they have in mind, as with Sperry's remarks, above.

That downward causation may be interpreted in line with *Strong* emergence is clear: one simply additionally requires that the new power associated with a *Strongly* emergent feature be associated with the production of lower-level effects. Similarly for a version of emergence discussed by Chalmers (2006) involving "a sort of incompleteness of physical laws even in characterizing the systematic evolution of low-level processes" and which he thinks is best understood "as involving a sort of downward causation". Here the appeal to downward causation may be seen as providing an account of the specific way in which *Strongly* emergent features are fundamentally novel or distinct.

Alternatively, downward causation may be interpreted along *Weak* emergent lines, as involving the holding of certain physically acceptable constraints on lower-level entities; here the appeal to downward may be seen as providing an account of the specific way in which *Weakly* emergent features are non-fundamentally novel or distinct. That said, as with appeals to non-linearity it is not obvious that such downwardly efficacious features

have the requisite ontological or causal autonomy, even granting that they inherit the physical acceptability of their base features. To the *prima facie* contrary: might not a given (token) feature of the wheel in virtue of which it rolls, on a given occasion, be identified with a (token) feature of the relational lower-level entity (consisting of atoms standing in atomic relation) constituting the wheel on that occasion? We will revisit this issue when considering conceptions of emergent autonomy that more directly appeal to the imposition of constraints (§3.5); there I will sketch a strategy that may work to gain autonomy in at least some cases of constraint-based downward causation. The broader moral at present is that additional work needs to be done to establish that features associated with lower-level constraints are non-fundamentally autonomous in the way required for physically acceptable emergence.

3.4.1 Kim's concerns about downward causation

The latter issue is key to Kim's concerns about downward causation and his associated exclusion argument, which aim to establish that non-reductive physicalism must collapse either into reductive physicalism or expand into robust emergentism (see, e.g., his 1989, 1993*a*, and 1998). As Kim correctly notes, blocking the robust emergentist's understanding of higher-level features requires accepting what he calls the "Causal inheritance principle", according to which every token power of a realized property instanced on a given occasion is numerically identical with a token power of the property instance realizing it on that occasion; and the challenge he has offered to the non-reductive physicalist is to show how, if token higher-level features have no powers not already had by token realizers, the former may be ontologically and causally (in particular, downward causally) efficacious. Certainly it is hard to see how such autonomy might be gained if higher-level instances inherit *all* the powers of their realizing instances. However, the explicit identification of *Weak emergence* as encoding the non-reductive physicalist's distinctive approach to higher-level causal autonomy, at least makes clear what the non-reductive physicalist needs to do in order to address Kim's concerns about downward causation. First, the non-reductive physicalist must establish that it suffices for causal autonomy that a higher-level feature have a distinctive power profile, as per the proper subset condition in *Weak emergence*; second, they must establish that at least some higher-level features in fact have distinctive power profiles. As noted, non-reductive physicalists do have resources along these lines, but whether these strategies succeed is the subject of ongoing debate.

3.5 The imposition of constraints

Closely related to physicalist conceptions of emergent autonomy in terms of downward causation are conceptions on which such autonomy is taken to reflect the imposition of lower-level constraints (see Klee 1984, Schroder 1998, and Wilson 2010*a*).

To repeat, granting that features associated with the imposition of lower-level constraints conform to *Physicalism*, it is not obvious that such features have the requisite ontological or causal autonomy; hence additional argument is needed to show that this conception conforms to *Weak emergence*. I provide a detailed such argument, for a special class of features associated with lower-level constraints, in Wilson 2010*a*. Here I sketch, very briefly, the strategy of that argument.

To start, I consider the notion of a degree of freedom (DOF)—roughly, one of a minimal set of independent parameters needed to characterize the states upon which the law-governed features of a (token of a given type of) entity (including systems) functionally depends.[19] Attention to DOF is useful in the present context, because the imposition of constraints at a lower-level generally affects, one way or another, the DOF needed to characterize the higher-level entities whose existence is to some extent determined by the holding of the constraints. Some sorts of changes in DOF resulting from the imposition of lower-level constraints may not be indicative even of *Weak emergence*—for example, cases where the DOF needed to characterize a higher-level entity (e.g., a rigid body, or a molecule) are identical to those needed to characterize lower-level relational entities realizing such higher-level entities, when the latter DOF can take on only constant or a restricted range of values. However, sometimes the imposition of lower-level constraints involves not just reductions or restrictions of (values of) lower-level DOF, but moreover *eliminates* certain lower-level DOF from those needed to characterize the associated higher-level entity. This is the case, for example, with certain features of quantum, statistical-mechanical, and complex dynamical entities or systems. (Note that the present strategy, supposing it works, would vindicate accounts of *Weak* emergence appealing to non-linearity.)

In such cases of elimination of DOF, I argue, there are reasons to think that the associated higher-level feature satisfies *Weak emergence*, in having only a proper subset of the token powers of the relational lower-level feature upon which it depends. As above, the usual strategy for showing this appeals to S's (functional or other) multiple realizability. Attention to DOF suggests a means of establishing satisfaction of *Weak emergence* even if S is only singly realized. Suppose S is singly realized by a (relational lower-level) base feature P (that is, a feature of a lower-level relational entity consisting in certain lower-level entities standing in lower-level relations). Now, again, 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 S are thus those expressed by the laws in the theory treating (constrained) entity S, while the powers of P are those expressed by the laws in the more fundamental theory treating the (relatively unconstrained) lower-level constituents of P—that is, the constituents of P as existing both inside and outside the constraints associated with S. Consequently, the laws of the theory treating S express what happens when certain lower-level entities stand in relations associated with certain lower-level constraints, and the laws treating P express what happens when certain lower-level entities stand both in these relations and in other relations not associated with the constraints. Hence the relational base feature P has more powers than S, and the proper subset relation between powers in *Weak emergence* is appropriately taken to be in place.[20]

[19] So, for example, specifying the configuration state for a free point particle requires 3 independent parameters (e.g., x, y, and z; or r, rho, and theta); hence a free point particle has 3 configuration DOF, and a system of N free point particles has 3N configuration DOF. And specifying the kinematic state for a free point particle requires 6 independent 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.

[20] For example, suppose that P is a quantum relational entity, and S is a classical entity singly realized by P. Then the causal powers of S include all those powers to produce, either directly or indirectly, effects that can occur in the macroscopic limit. The realizing entity P has all these causal powers, and in addition

Of course, this is only a sketch of how higher-level autonomy may be gained via the imposition of lower-level constraints (see my paper for details). The larger point for present purposes is that this or some other work needs to be done if such constraints are to serve as the basis for *Weak* metaphysical emergence.

3.6 Multiple realizability and its variants

As previously discussed (§2.4), non-reductive physicalists commonly appeal to multiple realizability in service of establishing the ontological and causal autonomy of higher-level entities. Related conceptions are in terms of "dynamical autonomy", where micro-level changes do not make a causal difference at the level of a system's dynamics (Wimsatt 1996), and "compositional variance", where the base entities of a given higher-level system exhibit "a much greater degree of variance and fluctuation from moment to moment than does the level of organization where [the higher-level entity] occurs" (Klee 1984, p. 48).

Why should multiple realizability, dynamical autonomy, or compositional variance support ontological and causal autonomy? Making the case for autonomy is crucial, since a now-standard reductionist strategy for accommodating multiple realizability and its variants proceeds by identifying multiply realized types with the disjunctions of their realizing types (see Clapp 2001). Plausibly, instances of a disjunctive type, on a given occasion, are identical with instances of whatever disjunct is instanced on that occasion; hence disjunctive identification blocks conformity to either *Strong* or *Weak* emergence.

That said, as above the reductive strategy for accommodating multiple realizability and its variants may be resisted, in service of establishing that some higher-level realized features are non-fundamentally novel or distinct, in a way having appropriate implications for their causal autonomy. In particular, non-reductionists may understand multiple realizability, and its dynamical and compositional variants, as tracking the higher-level feature's association with a distinctive causal role—that is, with a distinctive set of powers. Each lower-level realizer will have these powers (else it would not be a realizer), and some others besides, reflecting lower-level causal potentialities which differ between it and other lower-level realizers. Hence one may reasonably maintain that any instance of a multiply realizable feature has only a proper subset of the token powers of the base feature realizing it on that occasion, as *Weak emergence* requires. Correspondingly, conceptions of emergent autonomy appealing to multiple realizability, dynamical autonomy and compositional variability are best understood as providing a plausible basis for establishing that the proper subset condition in *Weak emergence* is met.

3.7 Results

We have arrived at the following results concerning metaphysical accounts of emergent autonomy:

- $\sim$ Conceptions of autonomy in terms of mere ontological novelty/difference or fundamental novelty/difference guarantee ontological autonomy (distinctness) but not causal autonomy.

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 where no macro-entities can exist.

$\sim$ Conceptions in terms of fundamentality of powers, forces, laws (and relatedly, conceptions in terms of failure of realization) aim to conform to *Strong* emergence.

$\sim$ Conceptions in terms of non-additivity of effects aim to conform to either *Strong* or *Weak* emergence, depending on whether the source of the non-additivity (non-linearity) involves new powers. A pressing need here is for those taking non-linearity as a basis for physically acceptable emergence to establish that higher-level non-linear features are ontologically and causally autonomous from their base features, in satisfying the proper subset condition on powers in *Weak emergence*.

$\sim$ Conceptions in terms of downward causal efficacy aim to conform to either *Strong* or *Weak* emergence, depending on whether the source of the downward efficacy involves new powers, or rather merely involves the imposition of lower-level constraints. Here too, it remains for those characterizing physically acceptable emergence in terms of downward efficacy to establish that the requisite ontological and causal autonomy is in place.

$\sim$ Conceptions in terms of the imposition of lower-level constraints aim to conform to *Weak* emergence. Here too, it remains for those characterizing physically acceptable emergence in terms of lower-level constraints to establish that the requisite ontological and causal autonomy is in place (though see Wilson 2010*a*).

$\sim$ Conceptions in terms of multiple realizability, dynamical autonomy and compositional variance aim to conform to *Weak* emergence.

4 Emergent autonomy: cognitive conceptions

Many historical and contemporary accounts of emergent autonomy involve appeals to the failure to hold of certain epistemological, representational, or conceptual connections, including unpredictability (Popper and Eccles 1977), in-principle failure of predictability or deducibility (Broad 1925), predictability, but only by simulation (Bedau 1997), and failure of representational or conceptual entailment (Smart 1981, Chalmers 1996, Van Gulick 2001). Such accounts are broadly cognitive in that they appeal to one or other failure on the part of creatures like us (or suitably idealized versions of us) to recognize certain connections as holding between certain higher-level and base features. For convenience, then, I will speak broadly of such conceptions as "cognitive" conceptions.

With few exceptions, cognitive conceptions of emergent autonomy aim to characterize metaphysical emergence. Typically, the relevant failures of cognitive connections are supposed to be concomitants of novelty or ontological irreducibility (or both). This is characteristic of, for example, Alexander's (1920) understanding of emergent phenomena as admitting "no explanation" because involving "brute empirical fact"; Kekes' (1966) understanding of emergence as involving a priori unpredictability of (claims about) higher-level features from (claims about) lower-level structure, due to novelty of higher-level property; and Kim's (1999) characterization of emergence as involving the joint failure of explanatory, predictive, and ontological reduction. Such conceptions may fall under the rubrics of *Weak* or *Strong* emergence, respectively, depending on what ontological aspect is at issue (as per §3).

Here I want to focus attention on accounts of emergence that are primarily or in any case officially cashed in cognitive terms. Along the way, we will confirm both that those endorsing cognitive conceptions typically aim to characterize metaphysical autonomy, and that they take themselves to have reason to think this can be done in epistemological or other cognitive terms. This is not true across the board, however; and I'll close (§4.4) with discussion of certain accounts of "non-reductive" physicalism which are explicitly cashed in terms of failure of conceptual connection, and which are better seen as ontologically reductive physicalist accounts aiming to make sense of our seeming inabilities to bridge certain explanatory gaps.

4.1 Failure of in-principle deducibility

Broad's official formulation of emergence (1925) is as follows:

> 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 $R(A, B, C)$.[21] (p. 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.

Broad begins his discussion of emergence by observing a distinction between two kinds of inter-level ("trans-ordinal") laws, which distinction is also presented in seemingly epistemological terms. First are trans-ordinal laws holding "between physical properties and properties at higher levels of the hierarchy which, while deducible in principle from a theory of the physical properties alone, are not deducible in fact". Second are trans-ordinal laws that are moreover "trans-physical", holding "between physical properties and properties at higher levels which are not deducible, even in principle, from a theory of the physical properties alone". Broad's official formulation of emergence thus obliquely characterizes the holding of trans-physical laws, by reference to the associated in-principle failure of deducibility that he assumes attaches to such laws.

In turn, for Broad, the existence of trans-physical laws has clear metaphysical consequences. That Broad supposes that trans-physical laws are at odds with a "mechanistic" (materialistic, physicalistic) view is some indication of this. Yet more telling are Broad's previously cited remarks to the effect that such laws are "unique and ultimate" (pp. 64-5)—that is, fundamental. That Broad understands trans-physical laws as indicative of metaphysical emergence is confirmed in passages such as the following:

> On the emergent theory we have to reconcile ourselves to much less unity in the external world and a much less intimate connexion between the various sciences. At best the external world and the various sciences that deal with it will form a kind of hierarchy. (p. 78)

[21] Note that Broad takes a "many-one" perspective on the relata of emergence here, with an uncharacteristically flexible understanding of what features may enter into the deduction, as going beyond the holding of pairwise (or other relatively non-complex) relations between the composing entities, to include relations between lower-level relata in any other (possibly complex) situations besides that at issue.

Emergence has implications for the unity of "the external world" and for the unity of the sciences "that deal with" the external world. These are clearly claims about metaphysical emergence; no failures of cognitive connection are ultimately at issue.

Similar remarks apply to other British Emergentists (e.g., Alexander), who, like Broad, sometimes characterized emergence as involving a failure of predictability. More generally, as McLaughlin (1992, p. 73) notes, "Emergentists often speak of emergent properties and laws as unpredictable from what they emerge from. But [...] 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" (p. 73).

4.1.1 Why (failure of) deducibility?

Since Broad's concern is clearly metaphysical emergence and more specifically *Strong emergence* (as involving fundamental laws and associated powers and forces), why does he characterize emergence in epistemological terms?

I speculate that this reflects a felt need to clarify the notion of fundamentality at issue, since certain ways of understanding this notion will not make sense of the characteristic dependence of emergent phenomena. In particular, we cannot here understand "fundamental" as "basic", "independent", or "axiomatic". Relatedly, Broad may have wanted to provide a substantive criterion of fundamentality, for purposes of applying his account. Insofar as it will plausibly be the case that goings-on governed (in part) by fundamental "trans-physical" laws will not be deducible from goings-on governed by physical laws, it would be natural to look to deducibility as a means of clarifying the distinctively dependent sort of fundamentality in *Strong emergence*. And Broad might reasonably have thought that the immediate concern with characterizing metaphysical emergence in epistemological terms—namely, that creatures as limited as we are might not be cognitively situated to recognize metaphysical connections that in fact exist—could be overcome by additionally qualifying the failure of deducibility as being "in-principle".

That said, the concern remains that even an ideal reasoner might fail to recognize metaphysical connections that in fact exist, in which case the criterion will produce false negatives. The procedure might also produce false positives, if certain uncontroversially physically acceptable phenomena (say, complex dynamical phenomena, of which Broad wasn't aware) are in-principle as well as in-practice unpredictable (perhaps because the sensitivity of such systems to initial conditions would require in-principle unavailable resources for predictability into the indefinite future). Supposing so, then Broad's criterion will inappropriately deem some physically acceptable features of complex phenomena *Strongly* emergent, hence physically unacceptable. In-principle failure of deducibility is thus best seen as a good though not infallible epistemological guide to the metaphysical features (involving fundamental powers and laws) characterizing *Strong emergence*.[22]

4.2 Failure of in-practice deducibility

Failure of deducibility or predictability also enters into some accounts of emergent autonomy aiming to characterize physically acceptable emergence (see Newman 1996, Bedau 1997, Rueger 2001); I'll focus on Bedau's work as representative in what follows. Bedau's

[22]That said, we will shortly consider whether in-principle failure of the broader notion of *a priori* entailment might do better along these lines.

(1997) account applies under conditions where a system S is composed of micro-level entities having associated micro-states, and where a "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. (p. 378)

Derivation of a system's macrostate "by simulation" involves iterating the system's microdynamic, taking initial and any relevant external conditions as input. A broadly equivalent conception takes emergent autonomy to involve "explanatory incompressibility", where there is no "short-cut" explanation of macro-features of a system with emergent features (see Bedau 2008). In being derivable by simulation from a micro-physical dynamic, associated macrostates are understood to be physically acceptable; as Bedau (1997) says, such systems indicate "that emergence is consistent with reasonable forms of materialism" (p. 376).

Though Bedau sometimes speaks of such systems as "epistemologically weakly emergent", he is explicit that the emergence involved is also metaphysical. He signals that "the modal terms in this definition are metaphysical, not epistemological" (1997, p. 379); he states his aim of capturing a form of "metaphysical autonomy" (2002, p. 11); he emphasizes that "weak emergence is not just in the mind; it is real and objective in nature" (2008, p. X). Such claims would seem to be in tension with Bedau's taking it to be characteristic of physically acceptable emergence that "the macro is ontologically and causally reducible to the micro in principle" (2008, p. 445); but Bedau thinks this implication can be resisted:

> [W]eak emergence exhibits a kind of macro autonomy because of the incompressibility of the micro-causal generative explanation of the macro structure. Because the explanation is incompressible, it is useless in practice (except in so far as it serves as the basis for a good simulation of the system). (2008, p. 449)

But it is unclear how usefulness in practice of explanations appealing to complex microphenomena might be relevant to establishing the ontological and/or causal autonomy of higher-level features, even granting that there is a metaphysical fact of the matter about when a feature has or does not have a "compressible" explanation. Effectively, such facts, though perfectly objective, are not of the right sort to ground the requisite ontological and causal autonomy. There is a parallel here to the failure of mere ontological distinctness to successfully capture emergent autonomy: what is needed for such autonomy is not just some or other metaphysical distinction between the higher-level and base features, but moreover one which plausibly serves as a basis for the causal as well as ontological autonomy of the former.

There are, however, resources at least potentially available for making sense of genuine autonomy in the cases Bedau aims to characterize, to which Bedau himself sometimes gestures. We saw previously how Bedau's (1997) observation that non-linear phenomena may enter into "simple, general, macro-level patterns" might serve as a basis for establishing genuine emergent autonomy of a physically acceptable variety: if, more generally, explanatorily incompressible phenomena enter into different, higher-level systems of laws,

this might serve to support an understanding of the associated features as having only a proper subset of the token powers of their lower-level base features, and hence as genuinely metaphysically emergent. Relatedly, Bedau (1997) observes: "Interesting macrostates [of sort at issue in weak emergence] typically average over microstates and so compress microstate information" (p. 377). If such compression of information involves an elimination in degrees of freedom (see §3.5), this would provide another route to ontological and causal autonomy. Alternatively, one might argue that compression of information is indicative of multiple realizability and/or "difference-making" considerations, of the sort that, as we have seen, plausibly motivate taking the requisite proper subset relation to be in place (here Bedau's 2002, p. 25 remarks concerning glider streams and their variable constituents are evocative). Hence it may be that, while Bedau's broadly epistemological account of emergent autonomy does not itself serve to characterize metaphysical emergence, an account based on the relevant metaphysical features of "interesting" cases of explanatory incompressibility may do so.

4.3 Failure of conceptual entailment

Next, consider Chalmer's notion of emergence in terms of a failure of *a priori* or conceptual entailment. Chalmers (1996) characterized (physically unacceptable) emergence in terms of a failure of broadly logical (conceptual) supervenience; in recent work (see Chalmers 1999 and Chalmers and Jackson 2001) he has developed the suggestion that one aspect of meaning is appropriately seen as tracking *a priori* connections. The notion of *a priori* entailment here goes beyond deducibility or any other syntactic notion, rather being linked to ideal conceivability and associated judgements of what is true in situations that are fully described along a certain (i.e., fundamental physical) dimension. So, for example, Chalmers argues that, upon contemplation of a scenario in which exists a creature functionally and physically identical to ours, an ideal reasoner would positively conceive that such a creature might not be conscious; Chalmers moreover argues that such ideal conceivability suffices for establishing the metaphysical possibility in question. The precise nature of the possibility that is established by so-called "zombie" arguments is subject to different broadly dualist interpretations (including substance dualism, robust emergence, and pan- or proto-psychism); but perhaps in combination with possibilities established by other ideal conceivings (namely, that there could be no conscious entities that were not dependently embodied, some way or other, at least in worlds relevantly like ours) one might so aim to establish the truth of *Strong emergence*.

It remains controversial whether conceivability, even of the highly idealized and nuanced variety, suffices for establishing the truth of various possibilities (see, e.g., Block and Stalnaker 1999); and the additional concern remains that such an idealized account is unuseful for or irrelevant to our gaining insight into the structure of natural reality (see, e.g. Melnyk 2008). Here I want just to call attention to two points. First, Chalmers, like previous proponents of cognitive conceptions of emergent autonomy, does so in service of establishing the holding (or failure to hold) of a metaphysical dependence relation. Second, supposing the strategy works and the appropriate conceivings are in place, the conception conforms to *Strong*, and not *Weak*, emergence. *Strong emergence* involves fundamental powers (forces/interactions, laws), and such fundamentality makes room for and sense of the failures of conceptual entailment present even to idealized conceivers. By way of contrast, there is no clear reason why an idealized conceiver could not identify

the connections between *Weakly* emergent features and their base features, given world enough and time.

4.4 Mere failures in cognitive connection

Though most epistemological accounts of emergent autononomy are aimed at characterizing metaphysical emergence (either *Weak* or *Strong*), this is not uniformly the case. In particular, a not-uncommon way of formulating (a version of what is sometimes called) non-reductive physicalism is as combining both in-principle ontological reduction with failure of one or other variety of cognitive connection (see, e.g., Smart 1981 and Van Gulick 2001). On such accounts, the "non-reduction" at issue is understood in purely epistemological terms, having no metaphysical implications; on the contrary, ontological reduction is assumed. Metaphysically speaking, such accounts are best understood as versions of ontologically reductive, not ontologically non-reductive, *Physicalism*, which aim to makes sense of the presence and seeming intractability of various explanatory gaps, and show that these do not pose a threat to such reductionism (see, e.g., Perry 2001). Such accounts may still be seen as addressing the initial motivations for attending to emergence, of understanding and accounting for the appearances of dependence and ontological and causal autonomy of higher-level entities, and associated hierarchical relations between (entities and features treated by) special and more fundamental sciences. But the account they offer will be importantly deflationary, from a metaphysical point of view, in denying that the appearance of autonomy is genuine (which is not to say that the appearances themselves are not grounded in objective facts). Given the desirability of providing a metaphysical ground for the ontological and causal autonomy of higher-level entities, however, proponents of reductive accounts would do well to consider whether the epistemological failures in question might, as with Bedau's understanding of emergence as involving in-practice failure of deducibility, be at least sometimes understood in terms compatible with *Weak*, if not *Strong*, emergence.

4.5 Results

We have arrived at the following results concerning epistemological accounts of emergent autonomy:

- ∼ Conceptions of emergent autonomy in terms of failure of cognitive connection typically aim to conform to metaphysical emergence.
- ∼ Conceptions in terms of in-principle failure of deducibility aim to conform to *Strong emergence*.
- ∼ Conceptions in terms of in-practice failure of deducibility (due to, e.g., explanatory incompressibility) aim to conform to *Weak emergence*, and may do so if the assumption of in-principle ontological and causal reducibility is dropped and the requisite ontological and causal autonomy established.
- ∼ Conceptions in terms of failure of ideal conceivability aim to conform to *Strong* emergence.

~ Accounts of *Physicalism* characterized in terms of one or other failure of cognitive connection coupled with ontological and causal reducibility fail to characterize any variety of emergence; these are best seen as versions of ontologically reductive *Physicalism*.

5 Concluding remarks

The problem of higher-level causation acts as a crucial constraint on feasible accounts of synchronically dependent higher-level features; and though the problem is typically presented in service of motivating one or other stance on the question of physicalism, consideration of the spectrum of available responses to the problem provides, more generally, a convenient way of seeing what our options are, so far as making sense of the metaphysical emergence of higher-level features is concerned.

There are only two responses to the problem making sense of higher-level features as both appropriately dependent on, and ontologically and causally autonomous from, lower-level features. Correspondingly, there are only two schemas for metaphysical emergence, which like the responses to the problem turn on the two available ways in which dependent higher-level features may be causally autonomous *vis-á-vis* their base features: either by having more powers, as per *Strong emergence*, or by having fewer powers, as per *Weak emergence*. Again, the notion of power here is almost entirely metaphysically neutral, requiring nothing much more than the bare association of (perhaps nomologically contingent) powers with features. There are no other options for gaining the causal autonomy of dependent higher-level features; hence these two schemas exhaust the available options for the metaphysical emergence for such entities.

Flexibility remains in filling in the schemas, however, via suitable accounts of emergent dependence and emergent autonomy. As I have argued, the many seemingly diverse accounts of these notions, when properly understood, each aim to conform to one or the other schema. And though my task here was not to assess the success of these aims, I have pointed out where more work needs to be done if certain accounts of emergent dependence or autonomy are to satisfy the conditions of the intended schema. Perhaps most crucially, it remains largely to establish that accounts of *Weakly* emergent autonomy in terms of non-linearity, lower-level constraints, and/or explanatory compressibility characterize higher-level features as having the ontological and causal autonomy requisite for genuine metaphysical emergence. That proponents have not realized that this work needs to be done likely reflects, I submit, that the powers-based conditions on (broadly synchronic, higher-level) metaphysical emergence have not previously been made fully explicit.

Hence it is, I hope, that the two schemas do more than systematize and unify the seeming diversity of accounts while explaining the variance *vis-á-vis Physicalism*. Additionally, and perhaps more importantly, with the schemas on the table we are in better position to consider and assess the available ways of filling them in, in ultimate service of better understanding the potentially diverse—but after all, not all *that* diverse—structure of natural reality.

References

Alexander, Samuel, 1920. *Space, Time, and Deity*. London: Macmillan.

Anderson, P. W., 1972. "More is Different". *Science*, 177:393–396.

Antony, Louise M. and Joseph M. Levine, 1997. "Reduction with Autonomy". *Philosophical Perspectives*, 11:83–105.

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.

Block, Ned and Robert Stalnaker, 1999. "Conceptual analysis, dualism, and the explanatory gap". *Philosophical Review*, 108:1–46.

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.

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 and Frank Jackson, 2001. "Conceptual Analysis and Reductive Explanation". *The Philosophical Review*, 110:315–60.

Chalmers, David J., 2006. "Strong and weak emergence". In *The Re-Emergence of Emergence*. Oxford University Press.

Clapp, Lenny, 2001. "Disjunctive Properties: Multiple Realizations". *Journal of Philosophy*, 98:111–136.

Clarke, Randolph, 1999. "Nonreductive Physicalism and the Causal Powers of the Mental". *Synthese*, 51:295–322.

Clayton, Philip, 2006. "Conceptual foundations of emergence theory". In *The Re-Emergence of Emergence*. Oxford University Press.

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.

Ehring, Douglas, 1996. "Mental Causation, Determinables, and Property Instances". *Nous*, 30:461–480.

Fodor, Jerry, 1974. "Special Sciences (Or, The Disunity of Science as a Working Hypothesis)". *Synthese*, 28:77–115.

Gillett, Carl, 2002*a*. "The Dimensions of Realization: A Critique of the Standard View". *Analysis*, 62:316–323.

Gillett, Carl, 2002*b*. "The varieties of emergence: Their purposes, obligations and importance". *Grazer Philosophische Studien*, 65:95–121.

Humphreys, Paul, 1996. "Aspects of Emergence". *Philosophical Topics*, 24:53–70.

Humphreys, Paul, 1997. "How Properties Emerge". *Philosophy of Science*, 64:1–17.

Kekes, John, 1966. "Physicalism, the identity theory, and the concept of emergence". *Philosophy of Science*, 33:360–75.

Kim, Jaegwon, 1989. "The Myth of Nonreductive Materialism". *Proceedings and Addresses of the American Philosophical Association*, 63:31–47. Reprinted in Kim 1993*b*.

Kim, Jaegwon, 1990. "Supervenience as a Philosophical Concept". *Metaphilosophy*, 21:1–27. Reprinted in Kim 1993*b*.

Kim, Jaegwon, 1993*a*. "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 1993*b*.

Kim, Jaegwon, 1993*b*. *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.

Klee, Robert, 1984. "Micro-Determinism and Concepts of Emergence". *Philosophy of Science*, 51:44–63.

Lewes, G. H., 1875. *Problems of Life and Mind*. London: Kegan Paul, Trench, Turbner & Co.

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.

McLaughlin, Brian, 1992. "The Rise and Fall of British Emergentism". In Ansgar Beckerman, Hans Flohr, and Jaegwon Kim, editors, *Emergence or Reduction? Essays on the Prospects of Non-reductive Physicalism*, 49–93. Berlin: De Gruyter.

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.

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*.

Moore, G. E., 1903. *Principia Ethica* . Cambridge: Cambridge University Press.

Morgan, C. Lloyd, 1923. *Emergent Evolution*. London: Williams & Norgate.

Newman, David, 1996. "Emergence and Strange Attractors". *Philosophy of Science*, 63:245–261.

O'Connor, Timothy, 1994. "Emergent properties". *American Philosophical Quarterly*, 31:91–104.

O'Connor, Timothy, 2002. "Free Will". *Stanford Online Encyclopedia of Philosophy*.

O'Connor, Timothy and Hong Yu Wong, 2005. "The Metaphysics of Emergence". *Nous*, 39:658–678.

Papineau, David, 1993. *Philosophical Naturalism*. Oxford: Basil Blackwell.

Perry, John, 2001. *Knowledge, Possibility, and Consciousness*. Cambridge, MA: The MIT Press, second edition.

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.

Rueger, Alexander, 2001. "Physical emergence, diachronic and synchronic". *Synthese*, 124:297–322.

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.

Shoemaker, Sydney, 2001. "Realization and Mental Causation". In *Physicalism and its Discontents*, 74–98. Cambridge: Cambridge University Press.

Shoemaker, Sydney, 2002. "Kim on emergence". *Philosophical Studies*, 58:53–63.

Silberstein, Michael and J. McGeever, 1999. "The search for ontological emergence". *Philosophical Quarterly*, 50:182–200.

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.

Sperry, Roger W., 1969. "A modified concept of consciousness". *Psychological Review*, 76:532–36.

Stephan, Achim, 2002. "Emergentism, irreducibility, and downward causation". *Grazer Philosophische Studien*, 65:77–93.

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.

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.

Wilson, Jessica, 1999. "How Superduper Does a Physicalist Supervenience need to Be?" *The Philosophical Quarterly*, 49:33–52.

Wilson, Jessica, 2002. "Causal Powers, Forces, and Superdupervenience". *Grazer Philosophische-Studien*, 63:53–78.

Wilson, Jessica, 2006. "On Characterizing the Physical". *Philosophical Studies*, 131:61–99.

Wilson, Jessica, 2009. "Determination, Realization, and Mental Causation". *Philosophical Studies*, 145:149–169.

Wilson, Jessica, 2010*a*. "Non-reductive Physicalism and Degrees of Freedom". *British Journal for the Philosophy of Science*, 61:279–311.

Wilson, Jessica, 2010*b*. "What is Hume's Dictum, and Why Believe It?" *Philosophy and Phenomenological Research*, 80:595–637.

Wimsatt, William, 1996. "Aggregativity: Reductive Heuristics for Finding Emergence". *Philosophy of Science*, 64:372–384.

Yablo, Stephen, 1992. "Mental Causation". *The Philosophical Review*, 101:245–280.