

ARTICLE

Power Emergentism and the Collapse Problem

Elanor Taylor

Johns Hopkins University, Baltimore, Maryland, US
Email: etaylo42@jh.edu

(Received 24 August 2020; revised 01 December 2020; accepted 27 January 2021; first published online 12 January 2022)

Abstract

Strong emergentism is the position that certain higher-level properties display a kind of metaphysical autonomy from the lower-level properties in which they are grounded. The prospect of *collapse* is a problem for strong emergentism. According to those who press the *collapse problem* any purportedly strongly emergent feature inheres in the emergence base and so is not genuinely autonomous from that base. Umut Baysan and Jessica Wilson (2017) argue that power emergentism avoids the collapse problem. In this paper, I challenge the claim that power emergentism avoids the collapse problem and argue for explanatory emergentism in its place.

1. Introduction

When philosophers face apparently higher-level phenomena that are intriguing and hard to explain, such as free will, consciousness, or entanglement, they often reach for some version of *strong emergentism*. This is the position that certain higher-level properties display a kind of metaphysical autonomy from the lower-level properties in which they are grounded. Precisely what that autonomy amounts to, the nature of the distinction between higher and lower levels, and what features are emergent are the subject of accounts of strong emergence. Strong emergentism appears across different areas of philosophy, but fundamentally, debates over the viability of strong emergentism are about the metaphysical interpretation of science—about whether, and how, we can take scientific results to deliver metaphysical verdicts.¹

The prospect of *collapse* is a problem for strong emergentism.² According to those who press the *collapse problem*, any purportedly strongly emergent feature inheres in the

¹ This is reflected in the history of strong emergentism. As documented by McLaughlin (2008) the British Emergentists of the late 19th and early 20th centuries took themselves to be offering a metaphysical interpretation of the science of the time, only to have their views rendered implausible by scientific progress, including the development of quantum-mechanical explanations of chemical bonding.

² I discuss this problem, and coin its title, in Taylor (2015b). I will refer to “the” collapse problem, but there are several different formulations of it, discussed by authors including Broad (1925) Grelling as discussed by Hempel and Oppenheim (1948) Howell (2009) O’Connor (1994) Shoemaker (2007).

emergence base and so is not genuinely autonomous from that base. Attempts to draw a principled distinction between the features of the base and the purported emergent break down and with them the claim that there is strong emergence. I have argued that the correct response to the collapse problem is to abandon strong emergentism and embrace a metaphysically neutral, explanatory account of emergence in its place. Recently, Baysan and Wilson (2017) have defended a different approach, arguing that a powers-based version of strong emergentism can avoid the collapse problem, and so a genuinely metaphysical approach to emergence is viable.

In this paper I respond to Baysan and Wilson, raising objections to their claim that powers-based emergentism avoids the collapse problem. I begin by outlining the threat of the collapse problem for strong emergentism and the route from the collapse problem to explanatory emergentism. I then describe Baysan and Wilson's strong emergentism and respond to their four strategies against the collapse problem. I conclude by attempting to ameliorate the impact of abandoning strong emergentism, in describing some features of explanatory emergentism that may make it more appealing to defenders of strong emergentism. I suggest that, combined with a metaphysics of explanation, explanatory emergentism can play a useful role in interlevel metaphysics.

2. Strong emergentism and the collapse problem

There are many different forms of strong emergentism, defended by authors including Baysan (2020), Broad (1925), Chalmers (1996, 2006), Gillett (2016), Humphreys (2016), Morrison (2012), Merricks (2001), and Wilson (2010, 2013, 2015), among others. Similarly, there are many different versions of the collapse problem, pressed and discussed by authors including Grelling as discussed by Hempel and Oppenheim (1948), Howell (2009), O'Connor (1994), Shoemaker (2007), and myself (Taylor 2015b). In this discussion, I will focus on my own presentation:

...cases of emergence presuppose a distinction between micro-level and macro-level properties. For any purported case of emergence, there are properties that *prima facie* belong to the micro level, but if they are included in the micro level then the purported emergent fails to meet a necessary condition for emergent autonomy. I call these problematic properties collapse-inducing properties because when they are included in the micro level, the purported emergent effectively 'collapses', and yet it seems arbitrary to exclude them. Furthermore, this problem does not depend on the details of any particular account of emergence and so applies quite generally. (Taylor 2015b, 732)

To press this objection, I considered a range of accounts of emergence, arguing that despite differences in the detail of each of these accounts, the emergence in each case "collapses."

For a simple, historical example, consider a collapse objection to Broad's claim that certain features of chemical compounds, such as the water-solubility of sodium chloride, strongly emerge from the features of their composing elements (Broad 1925, 61–65; Taylor 2015b, 735–36). Broad held that a necessary condition for emergence and a mark of emergence is that an emergent feature of a compound is not deducible from full knowledge its composing elements in isolation. However, sodium has the property of giving rise to a compound that is soluble in water when combined with

chlorine, and given this, the water-solubility of sodium chloride does not meet the nondeducibility criterion for emergence. Chlorine has a similar property, from which a similar result follows. The collapse problem for strong emergentism in general follows this pattern: identifying a property that apparently legitimately belongs to the emergence base, such that its presence in the base makes the emergence collapse. Following this pattern, the collapse problem applies generally to any form of strong emergentism.

Although the collapse problem is deceptively simple, its implications are serious. Strong emergence is supposed to be a metaphysical phenomenon. Accordingly, the difference between what is emergent and what it emerges from must be a metaphysically significant difference, such that emergent features are genuinely metaphysically autonomous from the base that gives rise to them. The collapse problem, however, indicates that this is not the case. If the emergent feature is in the emergence base already, then there is no metaphysically significant difference between the emergent and the base, and the claim that there is any strong emergence at all is undermined. Furthermore, the collapse problem appears to indicate that emergence is *arbitrary*, in that it can be made to appear or disappear depending on the set of properties that we may decide to consider part of the emergence base.

The most natural response to the collapse problem is to restrict the emergence base so as to exclude these collapse-inducing properties, and authors have considered a range of restriction strategies in response to this problem. For example, Skiles has argued that an account of emergence based on a notion of *generic essence* offers a solution to the collapse problem, as we can use generic essence to specify an appropriately restricted emergence base (Skiles 2016, reply in Taylor 2017a). However, these strategies turn out to be surprisingly difficult to defend. The principle used to restrict the emergence base must be motivated independently because without some independent motivation for the restricted base, the emergence base is restricted to preserve particular cases of emergence, which makes emergence an arbitrary phenomenon. It is easy to “cook up” a solution to the collapse problem for any given case of emergence by finding a distinction that excludes that particular collapse-inducing property from the emergence base. For example, for the sodium chloride case above, stipulating that the emergence base cannot include dispositional properties will prevent collapse. But finding a *principled reason* to do this, which is not merely an ad hoc response to one particular collapse problem, and an attempt to gerrymander the emergence base to protect a favored case of emergence from collapse, is very difficult indeed.³ To continue with the previous example, the restriction to nondispositional properties heads off collapse objections that involve dispositional properties, but there are collapse objections oriented around nondispositional properties as well. A restriction that may protect one set of cases of emergence will permit other cases to collapse, and so an independently motivated, nonarbitrary case must be made

³ We should expect a theory to be informed by motivating cases. Furthermore, given that many strong emergentists believe that the phenomenon is fairly rare, it is not intrinsically objectionable for an account to cover only a small number of cases. However, my concern is that building an account of strong emergence entirely around preselected cases and constructing the account to protect those cases from the collapse problem will result in an objectionably arbitrary form of emergentism. Thanks to an anonymous referee for pressing this concern.

for privileging one set of cases of strong emergence over other, apparently equally legitimate sets.

After considering and rejecting a range of attempts to motivate a restriction strategy, I took a different approach. I argued that there is a conceptual connection between emergence and explanation and that we use the concept of emergence to track the unavailability of certain explanations (Taylor 2015a). On this view,

A macro-level property p is emergent iff there is no available explanation of the fact that the following regularity obtains of natural necessity: Whenever components $A, B, C \dots n$ are combined in relation r , the resulting whole instantiates property p . (Taylor 2015b, 746)

In light of this *explanation condition* for emergence, what it takes to make emergence disappear, as the collapse objection apparently threatens, is to *explain* the instantiation of the emergent property. But collapse-inducing properties do not explain or provide resources for explanations of the instantiation of the relevant properties. This is because the collapse-inducing property can only support an attempt at explanation that falls foul of the requirement that facts do not explain themselves, and so the mere presence of a collapse-inducing property in the base cannot explain the instantiation of the emergent. This accounts for the impression of illegitimacy around the collapse problem because collapse-inducing properties cannot make a property nonemergent, or show it to be nonemergent, given that the only way to show a property to be nonemergent is to provide an explanation. Accordingly, the explanatory view of emergence avoids the collapse problem. Furthermore, given the conceptual connection between emergence and explanation, this explanatory view provides a satisfying, unified explication of the concept of emergence (Taylor 2015a).

However, explanatory emergence is not metaphysical emergence. I have argued that some cases of explanatory emergence may obtain for metaphysical reasons, given a metaphysics of explanation on which certain explanations can be unavailable for metaphysical reasons, such as if the unavailability of a causal explanation of one phenomenon in terms of another indicates that there is no causal connection between them. However, establishing such connections between explanation and metaphysics does not straightforwardly follow from explanatory emergentism (Taylor 2015a, 2017b). Some property's merely being emergent, on an explanatory view, is not enough to establish metaphysical autonomy. Accordingly, my view and other responses to the collapse problem that involve metaphysically neutral conceptions of emergence do not deliver the kind of strong, metaphysical emergence that many argue is needed to make sense of phenomena like consciousness, free will, or the apparent autonomy of nonfundamental sciences. Baysan and Wilson (2017) argue that a powers-based view of emergence can avoid the collapse problem, and so we can face down the collapse problem without abandoning a metaphysics of emergence.

3. The powers view of emergence

Baysan and Wilson (2017) do not argue that strong emergentism is true. Instead, they argue that the collapse problem does not undermine strong emergentism, and so strong emergentism makes sense (2017, 52). They focus on a powers-based version of strong emergentism, arguing that this approach is well-motivated and effectively captures the work that emergence is put to in philosophy, particularly in debates about physicalism and the metaphysical interpretation of science. As they put it “... attention to powers provides a systematic, historically supported, and properly metaphysical basis for accommodating weak as well as strong metaphysical emergence...” (Baysan and Wilson 2017, 59)

According to Baysan and Wilson, features are associated with powers in so far as those features endow entities with the capacity to produce certain effects in particular circumstances. They take the following schemas to capture the powers-based approach to emergence, in weak and strong versions:

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

Strong Emergence (SE): Token apparently higher-level feature *S* is strongly emergent from token lower-level feature *P* on a given occasion just in case, on that occasion, (i) *S* broadly synchronically depends on *P*, and (ii) *S* has at least one token power not identical with any token power of *P*. (Baysan and Wilson 2017, 59; Wilson 2015)

Their definition of strong emergence (hereafter SE) is the target of this discussion.

Although the collapse problem is a general problem for all forms of strong emergentism, Baysan and Wilson focus on two versions which they argue are the most salient and threatening to SE. These are “base-level power inheritance,” the worry that due to lawlike connections between the base and the purported emergent any purported emergent will be inherited by the base, and “dispositional feature inheritance,” the worry that any purportedly emergent feature is actually the manifestation of a disposition had by the base, and so properly a feature of the base (Baysan and Wilson 2017, 63–77).

Below, I will consider and respond to Baysan and Wilson’s four strategies for defending SE against these versions of the collapse problem. But first, a note about the dialectic. I offer a recipe for generating collapse problems for different approaches to emergence and argue that this recipe generally applies. This makes the collapse problem a general problem for all forms of strong emergentism, not merely powers-based views like SE. Because of this, a defense of SE is only a limited solution to the collapse problem. However, this does not undermine Baysan and Wilson’s defense of SE. Indeed, they could argue that it illustrates what is at stake in the conversation because a defense of SE against the collapse problem where defenses of other versions of strong emergentism fail will show that SE is the *only* viable form of strong emergentism.

A related consideration about defending SE is that it involves very specific metaphysical details, including specific ontological commitments. The view itself requires such commitments, and each of the four defense strategies invokes a further set, including an appeal to both lightweight and substantial dispositions for strategy 2 and the claim that there are strongly emergent objects for strategy 4. Even if these details are well-motivated, they are a cost of the view in comparison to explanatory and other metaphysically neutral accounts of emergence. These issues about parsimony are not problems for SE in itself, but they are a cost of SE in comparison with other views, and I will return to this point while discussing the four strategies.

In what follows I will not show that SE fails. However, I will raise some problems for Baysan and Wilson's defenses of SE against the collapse problem and argue that these problems tilt the scale in favor of an explanatory view. This is especially the case given that, as I will suggest in section 5, explanatory emergentism may, in combination with a metaphysics of explanation, deliver some of the apparently desirable features of strong emergentism.

4. Defending SE

4.1 Direct and indirect powers

Baysan and Wilson's (2017) first response to the collapse problem targets the base-level power inheritance version, the objection that because there are nomological connections between higher- and lower-level properties the emergence base inherits any purportedly emergent power (2017, 78). Baysan and Wilson argue that we can defend SE against this problem because even if the emergence base inherits the powers associated with the emergent feature, it does so only in an *indirect* way. As a reminder, their formulation of SE is:

Strong Emergence (SE): Token apparently higher-level feature *S* is strongly emergent from token lower-level feature *P* on a given occasion just in case, on that occasion, (i) *S* broadly synchronically depends on *P*, and (ii) *S* has at least one token power not identical with any token power of *P*. (2017, 59)

As Baysan and Wilson put it, "The suggestion is that *S*'s novel power or powers are not had or manifested by lower-level features in the same direct or immediate way as they are had or manifested by *S*" (2017, 78). More specifically, they suggest that although *P* synchronically necessitates *S*, *P* is only a precondition for *S*, which is "the more direct locus of the power" (2017, 79). Accordingly, although *P* necessitates *S*, *P* is an indirect, rather than direct, source of *S*'s novel, emergent power.

Baysan and Wilson argue that this distinction between having powers directly and indirectly is independently motivated and that there are two ways to substantiate the intuition and strategy, through an analogy with temporally extended causal chains and through an analogy with sets and subsets (2017, 79–80). With temporally extended causal chains, even if each link is nomologically sufficient for the next link, we can still distinguish between more and less direct causes of the end result because temporally more proximal causes are more direct while temporally less proximal causes are less direct. We can use this comparison to motivate the idea that even

though the emergent feature is dependent on its base in so far as the base inherits the emergent power, the base may have the emergent power less directly than the relevant higher-level feature has the power.

On the second approach, the analogy with the set/subset distinction, Baysan and Wilson argue that we can distinguish between different sets of circumstances associated with a single temporal interval. As they put it:

... powers are individuated, in part, by the circumstances in which they manifest and contribute to the production of a given effect; but just as we can distinguish between a set and its subsets at a time, there seems to be no in-principle reason why we cannot distinguish different sets of circumstances associated even with a single temporal interval (instantaneous or extended). (2017, 80)

More concretely, they suggest that P has the power to contribute to the production of S in circumstances K that do not include S and so has an indirect power to contribute to causing anything that S can cause. S, on the other hand, has its novel/emergent power directly, and it manifests in set of circumstances K', which is distinct from set K because K' includes S. Because powers are at least partly individuated by the circumstances in which they manifest, and the emergent power is manifest under different circumstances by P and by S, this permits us to distinguish between the indirect sense in which P has the emergent power and the direct sense in which S has the emergent power. Just as we can distinguish between a set and a subset, so we can distinguish between a power associated with one set of circumstances and a power associated with a subset of those circumstances.

In order to protect SE from the collapse problem, the distinction between direct and indirect powers motivated by these comparisons must be nonarbitrary and metaphysically robust. That is to say, the distinction must be independently motivated, to avoid the concern that it has been constructed specifically to protect a favored case of emergence from collapse, and it must be a genuinely metaphysical distinction, rather than pragmatic or interest-relative, to preserve the idea that emergence itself is metaphysical rather than pragmatic or interest-relative. We can see why the distinction operative in each comparison case appears to satisfy this requirement. In the case of the causal chain, temporal proximity provides an independently motivated basis for distinctions between direct and indirect causes. In the set/subset case, distinctions between sets and subsets are nonarbitrary, commonplace, and well-motivated independent of any considerations about the collapse problem.

However, although each analogy provides a clear sense of what we *need* from a distinction between direct and indirect powers that will protect SE from this version of the collapse problem, it is unclear that the powers framework meets this need.

In the case of the causal chain, time mediates the distinction between more and less direct causes. If we ask, "why is cause A a more direct cause of effect E than cause B?," the answer is, "cause A is closer in time to effect E than cause B." In the emergence case, the powers framework does not seem to offer anything that plays this role, a further basis for this distinction between direct and indirect powers. When we ask, "why does P have the emergent power only indirectly, while S has it directly?," there is no answer beyond "P has the power indirectly, and S has the power directly." The powers framework permits us to draw the distinction and to

describe it, but does not appear to give us any further reason why it obtains in this case. Perhaps one could argue that the status of being a mere precondition for a power is doing this work. But to make sense of this we need a further story about what it is for one power to be had by a feature in such a way as to act as a precondition for the instantiation of the same power by another feature. Perhaps the powers in question are *not* the same, but in that case, the distinction at work in this strategy is not between having a power directly or indirectly, so much as it is a distinction between two entirely distinct powers. At this stage, the powers framework is not delivering a clear, straightforward basis for this distinction between direct and indirect powers.

However, perhaps this is too quick. Given that powers are at least partly individuated by circumstances, the comparison to the set/subset distinction that appeals to this aspect of powers might provide a justification and basis for the claim that P has the relevant power indirectly while S has it directly.

In the analogy with the set–subset distinction, Baysan and Wilson (2017) claim that just as we can distinguish between a set and subsets, so we can find a distinction between direct and indirect powers on the basis of the direct power being associated with one set of circumstances, and the indirect power being associated with a subset of those circumstances. This seems like a promising resource for a deeper basis for the distinction between direct and indirect powers. In particular, P manifests the emergent power in circumstances that do not include S and hence has the power indirectly, while S manifests the emergent power in circumstances that do include S and so has the power directly. However, if we look at any situation in which P manifests the emergent power, P will manifest that power *through* the instantiation of S (given the background presumption that the base is nomologically sufficient for S). Although we can distinguish between a set of circumstances that includes P and a set of circumstances that does not, along the lines of the set/subset analogy, this distinction is fairly flimsy because any circumstances in which P manifests the emergent power will necessarily include S. Accordingly, the idea that the basis of the distinction between direct and indirect powers is that P manifests the power in circumstances that *do not* include S seems too weak to protect SE from the collapse problem.

The broader powers framework is rich and detailed, and these considerations do not end this conversation. For instance, a proponent of SE could perhaps reject the idea that we need a basis for and an explanation of the distinction between direct and indirect causes, as I have claimed here. But the concerns I have raised generate some doubt about whether these two analogies, using causal chains and sets to make sense of the distinction between direct and indirect powers, really help the proponent of SE to identify a distinction between powers that will protect SE from the collapse problem.

4.2 Lightweight and substantial dispositions

The second strategy is a variation on the first, adapted to reply to the dispositional feature inheritance version of the collapse problem, the objection that the purportedly emergent feature is the manifestation of a disposition had by the base level. In

response, Baysan and Wilson (2017) argue that the relevant disposition inheres in the base level in only a *lightweight* rather than a *substantial* manner. As they put it:

... the intended sense in which the physical base features have dispositions to bring about strongly emergent features here is lightweight, signifying just that the base features are preconditions for the occurrence of the fundamentally novel strongly emergent feature, contra physicalism. (2017, 81)

The similarity with the direct/indirect power strategy is evident. Much as the indirect power had by lower-level feature P was a precondition for the instantiation of the direct power by higher-level feature S, the lightweight disposition associated with the lower level is a mere precondition for the manifestation of the emergent feature, and hence the emergent power, at the higher level.

The threat of the collapse problem is that some property in the emergence base crosses the barrier between higher and lower level and so makes a purported emergent no longer meet a necessary condition for emergence. Accordingly, the collapse problem is always relativized to some criterion for emergence. In the case of SE, that criterion is that the higher-level feature has a power that is genuinely novel, in that it is not had by any feature of the base.

This appeal to a distinction between lightweight and substantial dispositions is a promising response to the collapse problem because standard ways to individuate dispositions provide an independently motivated basis for distinguishing between the dispositions in this case. Take Broad on water-solubility again as an illustration (1925, 61–65). For Broad, the water-solubility of sodium chloride was an emergent property of the compound sodium chloride. In formulating the collapse objection to this case, I noted that sodium has the disposition to form a compound that is soluble in water when combined with chlorine and that this is sufficient to collapse this case relative to Broad's deducibility criterion for emergence (Taylor 2015b, 735–36). However, in the case of SE, this is not so straightforward. Because dispositions are individuated by their stimulus and manifestation conditions, the disposition that sodium has is different from the disposition that sodium chloride has. The stimulus conditions differ, in that sodium must be combined with chlorine in order to manifest the water-solubility, whereas sodium chloride does not. So, the higher-level disposition is distinct from the lower-level disposition, which appears to protect SE from this version of the collapse problem and vindicate the idea that the lower level has a disposition to manifest solubility in only an indirect way, as opposed to the direct disposition had by the higher level.

However, the manifestation of each disposition is the same, in that it amounts to the dissolving of sodium chloride in water. Even if there is a distinction between lightweight and substantial dispositions, in having this disposition, the lower level has a power to manifest water-solubility, which is the purportedly emergent power in this case. SE requires that there be some higher-level power that the lower level *does not have*. But the lower level *does* have the power to manifest water-solubility. Even if the lower level has the power in a *different* way from the way that the higher level has that power, the lower level still has it, and so this appeal to lightweight and substantial dispositions does not properly protect SE from this version of the collapse problem.

There are further places this conversation could go. For instance, a proponent of SE could argue that the emergent power is not a power to manifest water-solubility, but is instead a power to manifest water-solubility *without* certain other conditions being in place. But at this stage these attempts to protect the strategy run into the danger I mentioned before of “cooking up” an emergence base to protect a particular case of emergence. This leads to broader, interesting questions about arbitrariness and ad hocness, but for now I take these considerations to shed doubt on the capacity of this appeal to lightweight and substantial dispositions to protect SE from the collapse problem.

4.3 Powers relativized to fundamental interactions

The third strategy is very different from the first two and is based on the idea that the powers an entity has are grounded in particular or particular sets of fundamental forces or interactions. As Baysan and Wilson put it, “. . . powers are relativized to sets of fundamental interactions, making room for higher-level features to have powers that are in some sense new, as SE requires . . . (2017, 84). For example, Baysan and Wilson discuss the example of the power of being able to bond with an electron, which is specifically grounded in the electromagnetic (or electroweak) interaction.

In Wilson’s recent work on this topic she defends a view of fundamental interactions as second-order multitrack dispositions, such that the powers of ordinary objects or their substantial components are local manifestations of more fundamental powers of one or more fundamental fields (2021). On this view, physicalists think that physical interactions are the *only* fundamental interactions, while the strong emergentist holds that there are nonphysical, configurational interactions. This idea funds a reformulation of SE, as follows:

Interaction-relative Strong Emergence (Interaction-relative SE): Feature S is strongly emergent from feature P relative to the set {F} of fundamental physical interactions, just in case (i) S broadly synchronically depends on P, and (ii) S has at least one power that is not identical with any power of P that is grounded only in the fundamental interactions in {F}. (Baysan and Wilson 2017, 86)

SE requires genuinely new powers to emerge at a higher level. On Interaction-relative SE that newness is secured by *an interaction* that is not among the set of physical interactions.

Baysan and Wilson argue that there are a number of advantages to this reading of SE. It is in the spirit of the original British Emergentists because it involves a version of what they called *configurational forces* (2017, 86–87). Interaction-relative SE clearly distinguishes between dependent features that are over and above their base features in a way that is compatible with physicalism, from those which are over and above their base features in a way that is not compatible with physicalism (2017, 87). On this version of SE, emergentism need not be associated with explanatory gaps and may not follow from even an insuperable explanatory gap, which Baysan and Wilson argue is a benefit of the view given the difficulties involved in interpreting explanatory gaps (2017, 87). Finally, Baysan and Wilson argue that this version of SE provides resources for a response to the collapse problem (2017, 87).

This response is based on the idea that, even if features of the composing system inherit all the powers of the features they give rise to, the composite features may be associated with powers that are new in that they are not grounded *only* in the set of fundamental physical interactions. Relativizing powers to fundamental interactions provides a principled basis for distinguishing between dispositions that are *mere preconditions* for the occurrence of strongly emergent features from those that *actually have* the novel power in question. This permits the emergentist to maintain that the novel powers are not in the base because the physical powers are grounded only in fundamental physical interactions, and the novel powers are grounded in a different set of interactions. This is a point at which Howell's specific framing of the collapse problem is helpful and illustrative (2009). Howell discusses the idea that an emergence base could be "polluted" by the emergent property, which undermines the claim that the emergent is properly autonomous from the base. In Interaction-relative SE, the emergentist can articulate what it is for the emergence base to be polluted or not polluted by the emergent because this can be relativized to base-level or higher-level interactions. A power grounded in a nonphysical interaction has no place in a physical base and so pollutes the base.

This is a rich and detailed response to the collapse problem, and so objections can only be sketched here. But, my first concern is about how to individuate powers and interactions in the robustly metaphysical manner required to defend Interaction-relative SE from collapse. Consider Baysan and Wilson's example of the power of being able to fall when poised above the Earth's surface, which is grounded in the gravitational force (2017, 84). Because this power is grounded in the gravitational force, it is a physical power. However, consider the comparison between the power to fall and related powers such as the power to *dive*. The power to dive seems to be distinct from the power to fall and also paradigmatically nonphysical, in that it is not a concern of physical theory.⁴ However, the power to dive is very close to the power to fall and presumably also grounded in the gravitational force. So, how are we to individuate these powers? Is the power to fall *part* of the power to dive? In that case, does that mean that the power to dive is a hybrid higher- and lower-level power, higher because it is not a standard subject of physical theory and lower because it is composed of powers grounded in physical interactions? Or are they the same power, described differently? To defend Interaction-relative SE we need to be able to clearly distinguish between powers and also to be sure that this distinction is metaphysical, rather than merely pragmatic, but this starts to look difficult even with the paradigmatically lower-level case of the power to fall.⁵ Such worries do not show that

⁴ Baysan and Wilson take being a proper subject of physics as a rough criterion for an operative notion of the physical (2017, 85)

⁵ One could reply to this worry by pointing out that the power to dive is complicated by the role of intention. If diving falling with intention, then intention is responsible for the apparently higher-level aspect of the power. If physicalism about the mental is false, then the power to dive will be emergent, whereas if physicalism is true, then it will be a physical power. In response, I take this detail to illustrate the challenge posed by this case. The power to dive is almost physically identical to the power to fall. However, the role of intention and other aspects such as not being the apparent proper subject of physical theory appear to make it a distinct, higher-level power. Given that, we face a challenge in simply attempting to individuate these powers before we can move on to examine the interactions and forces in which they are grounded and thereby work out whether they are emergent by the standards of Interaction-relative SE. Thanks to an anonymous referee for pressing this point.

Interaction-relative SE fails. But they do show that what we need for this strategy to work is a full metaphysics of interactions and powers, to fund an unambiguously metaphysical individuation of powers and of their levels.

Another concern about this strategy does not target the strategy in itself, but instead targets the strategy as a defence of Interaction-relative SE against explanatory emergentism and other metaphysically neutral forms of emergentism. Part, or even most, of the job of an account of strong emergence is to help us to identify metaphysical distinctions between levels. If an account of strong emergence is to do anything, it is to articulate what it is for some higher-level feature to be metaphysically novel and thereby autonomous from the base-level features that give rise to it. One worry about explanatory emergentism is that it does not deliver these metaphysical results. This appears to be a *prima facie* count against explanatory emergentism and in favour of strong emergentism for those who are interested in interlevel metaphysics—explanatory emergentism does not deliver a metaphysics of levels while strong emergentism does.

However, Interaction-relative SE does not appear to do this work. Because the metaphysical autonomy of the higher-level features resides in differences between the forces to which the powers are relativized, the metaphysical autonomy of emergence amounts to the differences between physical and nonphysical forces. Once we recognize that there are fundamental physical forces and fundamental nonphysical forces, we have established a metaphysical distinction between levels, and so the metaphysical work in distinguishing between levels is not performed by the account of strong emergence. If Interaction-relative SE is not doing the work of identifying a distinction between levels, then it offers no extra metaphysical resources over those offered by an explanatory view.

An attractive feature of the powers approach is that it appears to preserve genuinely metaphysical emergence in the face of the collapse problem, but this defence of it gives up on the idea that emergence itself is doing the metaphysical work. Accordingly, this purported benefit of the powers view over the explanatory view is undermined, and we are left to consider the differences between them on other grounds, such as the extent to which each view unifies scientific and philosophical discourse on emergence or the extent to which each view is parsimonious. This does not show that Interaction-relative SE is false, but it undermines an apparent advantage of Interaction-relative SE over explanatory emergentism.

4.4 Strongly emergent objects

The fourth strategy is based on the idea that a strongly emergent feature must be instantiated in an *emergent object*, distinct from the lower-level object bearing the lower-level base features. This distinction between the different objects instantiating the base and emergent properties delivers the kind of metaphysically nonarbitrary distinction between groups of properties that is required to overcome the collapse problem. Because the emergent property is borne by an entirely different object from the object that bears the nonemergent properties, there is a genuine metaphysical barrier between the strongly emergent properties and the base properties, which would appear to protect the strongly emergent property from collapse. This idea has been defended in different forms by a range of authors, and Baysan and

Wilson discuss four versions (2017, 89), from O'Connor and Jacobs (2003), Nida-Rümelin (2006), Heil (2012), and Baysan (2016).

O'Connor and Jacobs use this approach to defend the emergence of certain features of conscious entities, arguing that conscious subjects require a "thisness" or "particularity" that their parts cannot have (2003). Accordingly, emergent properties of conscious entities emerge from an object, a subject, distinct from the parts of that entity. Similarly, Nida-Rümelin is motivated by the apparent need for a distinct subject of conscious experience (2006). Heil offers a different motivation, arguing that emergent properties emerge from new objects because any bearer of a property must be a simple substance, and so the bearer of an emergent property must be an emergent simple substance (2012). Baysan argues that although we typically attribute powers to properties, it is more appropriate to attribute them to the objects that bear the properties. Accordingly, we should expect a strongly emergent power to be instantiated in a new, emergent object (2016). A response to the collapse problem requires a principled distinction between emergent properties and base properties, and being instantiated by entirely different objects certainly offers a basis for that distinction. However, this strategy of appealing to emergent objects faces some problems.

First, this is the least parsimonious of all the strategies considered so far. This view relies on not only a powers ontology and SE as an account of emergence, but also the view that emergent properties are borne by a class of higher-level objects. Furthermore, the view relies on our capacity to distinguish between emergent and nonemergent objects in a way that permits us to make metaphysical claims and to protect claims about emergence from collapse. This is a significant group of commitments and the most detailed and metaphysically unparsimonious of all four strategies.

A second concern about this strategy is that it is oriented around specific cases of emergence. This is most striking in the versions defended by O'Connor and Jacobs and by Nida-Rümelin, who base their views around favored cases of emergence, focusing particularly on consciousness. We should expect some back-and-forth between the cases an account is intended to capture and the distinctions used to capture them. However, in these cases that back-and-forth has tilted very far in one direction, toward building the metaphysics *entirely* around chosen cases. The authors have decided which phenomena are strongly emergent and have oriented the theory around those phenomena, rather than beginning with criteria for metaphysically significant distinctions between levels and then using those criteria to *discover* what is metaphysically emergent from what.

Third, this strategy faces a similar problem in comparison to explanatory views as came up for strategy 3, that on this view the metaphysical work in drawing distinctions between levels inheres in this association between emergence and new objects and so is not performed by the account of emergence. As before, this is not a problem for this defense of SE in itself. Instead, this is a problem for this defense of SE against explanatory emergentism. The metaphysical work that an account of strong emergence is supposed to do and which is an advantage of strong emergentism over explanatory emergentism is passed off onto another distinction, which leaves this version of strong emergentism at no advantage over an explanatory view.

Finally, it seems that some of the most intriguing and interesting cases of apparent emergence that this emergent objects strategy accommodates involve an explanatory gap, as in Nida-Rümelin's discussion of consciousness (2006). Given that explanatory considerations drive these cases, they can be accommodated by the explanatory approach to emergence, which not only makes room for such cases, but also offers strategies for identifying and contextualizing such explanatory gaps and working out when it may be legitimate to draw metaphysical conclusions from the unavailability of explanations (Taylor 2017b).

5. Metaphysically interpreting explanatory emergence

Baysan and Wilson (2017) have argued that their powers-based version of strong emergentism, SE, avoids the collapse problem. I have raised some problems for each of their strategies in defense of SE against the collapse problem. Even if the broader powers framework remains viable, these considerations cast doubt on SE's viability in the face of the collapse problem.

At this point, I want to return to explanatory emergentism. Overall, strong emergentism should deliver the result that emergence marks a genuine metaphysical distinction between levels, and as such strong emergentism should provide resources to articulate various positions in interlevel metaphysics. The collapse problem undermines strong emergentism. But I will suggest that we can still access some of these desirable features of strong emergentism from an explanatory view of emergence. A full defense and articulation of this position cannot be given here, but I will sketch some suggestions.

On an explanatory view of emergence, emergence may obtain for metaphysical reasons if there are connections between explanation and metaphysics. If it is the case that sometimes explanations are unavailable for metaphysical reasons, such that an unavailable explanation of a certain kind might indicate a metaphysical distinction between levels, then explanatory emergence might obtain for metaphysical reasons. Accordingly, combined with a metaphysics of explanation, an explanatory approach to emergence can be used to articulate the view that there are metaphysical distinctions between levels and can be used to identify such distinctions.

To illustrate, consider this more detailed version of my "explanation condition" for emergence:

Given components A, B, C . . . n arranged in relation r into a whole, and an observer O, property x of the whole is emergent for O iff there is no scientific explanation available to O of the fact that the following regularity obtains of natural necessity: Whenever components A, B, C . . . n are combined in relation r, the resulting whole instantiates property x. (Taylor 2015a, 659)

On this view, claims about emergence are relativized to a number of different factors, including a distinction between whole and parts, a standard for explanation and a standard for unavailability. Although on this formulation emergence is relativized to an observer, observers may converge on cases of emergence, and this detail accommodates historical cases of emergence and emergence relativized to particular communities, as well as cases that obtain for all possible observers. For example,

when discussing historical claims about emergence, we can identify the explanations un/available for a given community given the state of knowledge at the time and contrast this with what explanations are un/available for contemporary observers. Similarly, we can compare cases of emergence relativized to different definitions of explanation and different standards for unavailability. I have argued that because of these features, this view offers a useful framework for clarifying historical and contemporary debates about emergence, including debates about the explanatory gap in philosophy of mind, and historical debates between mechanists and vitalists (Taylor 2017b). It can also be used in combination with a metaphysics of explanation to identify cases of emergence that obtain for metaphysical reasons.

To sketch this process, consider the following description of some dye being dropped in water, which List and Pivato suggest is a case in which higher-level description is indispensable:

... imagine that we pour a few drops of blue dye into one part of a water tank, undisturbed by any movement. How will the dye diffuse? If viewed through a microscope, each of the trillions of jostling, jiggling blue dye particles would exhibit Brownian motion and wander along some convoluted, labyrinthine path through the tank, which, in turn, is the result of a deterministic kinetic-molecular process. All this is extremely hard to model. At a macroscopic level, however, the system admits a very simple and informative description: if we write down a function describing the three-dimensional density distribution of the dye in the water at time zero, then this function evolves predictably under a partial differential equation called the heat equation, which is often amenable to a relatively easy computational solution at all future times. (List and Pivato 2015, 143–44)

Imagine that a person takes up this claim and argues that because of the indispensability of the higher-level description, the pattern in the diffusion of the dye is strongly emergent from the properties of the particles. Explanatory emergentism provides a schema into which we can slot the details of the case and then evaluate the claim that a higher-level description is genuinely indispensable and decide whether the metaphysical implication really does follow. The relevant group of observers is the contemporary scientific community. The standard for availability is an epistemic standard about what we can and cannot model. The explanation unavailable with respect to this standard is a microscopic mechanistic explanation. The lower-level properties are the properties of the particles, and the higher-level property is the distribution pattern of the blue dye. Once we have filled out the schema, we can see which parts of the claim might be contentious, whether for scientific or metaphysical reasons. For instance, we might wonder how we can be sure that we will not develop more effective higher-level descriptions and whether the unavailability of a microscopic mechanistic explanation really does indicate that this is a case of metaphysical emergence. This is the point at which a metaphysics of explanation comes in, telling us about which unavailable explanations are metaphysically significant, and why.

Explanatory emergentism does not deliver a metaphysics of levels. But if there are connections between emergence and explanation, then, as illustrated in this case, the account can be used to hone in on metaphysically significant cases of emergence, which mark metaphysical distinctions between levels. Establishing that the

availability or unavailability of explanations is metaphysically significant requires a metaphysics of explanation, the development of which is necessary for reasons beyond these particular debates about emergentism. But if there are such connections, then explanatory emergence can be used to identify cases in which emergence obtains for metaphysical reasons and so can deliver some of the desirable features of strong emergentism by assisting the metaphysical interpretation of facts about the availability and unavailability of explanations.

Although explanatory emergence can accommodate a metaphysics of levels, in this view the metaphysical work is done by connections between explanation and metaphysics, rather than in the account of emergence itself. Earlier, I criticized Baysan and Wilson's defenses of the powers view against the collapse problem for displaying just this feature, so one might wonder whether this is a problem for explanatory emergentism as well. However, this objection came up in a particular dialectical context. Baysan and Wilson were arguing that we can defend strong emergence in the face of the collapse problem and that part of the benefit of an account of strong emergence is that it offers an unambiguously metaphysical approach to emergence. Once the metaphysical work is done outside of the account, as in distinctions between levels or connections between explanation and metaphysics, then the account of emergence is not doing that metaphysical work. This is not a problem for the powers view, but for the defense of the powers view as being a properly metaphysical approach to emergence, in contrast to more metaphysically neutral approaches. Once we acknowledge that it is not a properly metaphysical view of emergence, then it does not have this advantage over the explanatory view, and then we must evaluate the differences between the accounts on different grounds, such as how well they unify discourse. This objection does not come up for the explanatory account because it has never been defended as delivering a strong, metaphysical view of emergence.

Overall, if there are the right kinds of connections between explanation and metaphysics, then explanatory emergentism can play a central role in inquiry into interlevel metaphysics. However, that work requires the development of a full metaphysics of explanation.

6. Conclusion

The threat of *collapse* is a problem for strong emergentism. In response to this threat, some have argued that we should embrace a metaphysically neutral, explanatory view of emergence. Baysan and Wilson (2017) have responded, arguing that a powers-based view of emergence can head off the collapse problem, and deliver a genuinely metaphysical form of emergentism. I have raised some problems for their defense of the powers view and with it their defense of strong emergentism against the collapse problem. To ameliorate this conclusion, I have sketched the suggestion that explanatory emergentism can deliver some of what seemed desirable about strong emergentism in that, combined with a metaphysics of explanation, it can play an illuminating role in interlevel metaphysics.⁶

⁶ Thanks to Patrick Connolly and two insightful and generous anonymous referees for discussion and feedback. Thanks also to Michaela McSweeney and Elizabeth Miller for help and encouragement while working on this project.

References

- Baysan, Umut. 2016. "An Argument for Power Inheritance." *The Philosophical Quarterly* 66:383–90.
- Baysan, Umut. 2020. "Causal Emergence and Epiphenomenal Emergence." *Erkenntnis* 85:891–904.
- Baysan, Umut, and Jessica Wilson. 2017. "Must Strong Emergence Collapse?" *Philosophica* 91:49–104.
- Broad, Charlie Dunbar. 1925. *The Mind and Its Place in Nature*. New York: Harcourt, Brace & Company.
- Chalmers, David. 1996. *The Conscious Mind*. New York: Oxford University Press.
- Chalmers, David. 2006. "Strong and Weak Emergence." In *The Re-emergence of Emergence*, ed. Philip Clayton and Paul Davies, 244–55. New York: Oxford University Press.
- Gillett, Carl. 2016. *Reduction and Emergence in Science and Philosophy*. Cambridge: Cambridge University Press.
- Heil, John. 2012. *The Universe as We Find it*. Oxford: Oxford University Press.
- Hempel, Carl, and Paul Oppenheim. 1948. "Studies in the Logic of Explanation." *Philosophy of Science* 15 (2):135–75.
- Howell, Robert. 2009. "Emergentism and Supervenience Physicalism." *Australasian Journal of Philosophy* 87:83–98.
- Humphreys, Paul. 2016. *Emergence: A Philosophical Account*. New York: Oxford University Press.
- List, Christian, and Marcus Pivato. 2015. "Emergent Chance." *Philosophical Review* 124 (1):119–52.
- McLaughlin, Brian. 2008. "The Rise and Fall of British Emergentism." In *Emergence: Contemporary Readings in Philosophy and Science*, ed. Mark Bedau and Paul Humphreys. Cambridge, MA: MIT Press.
- Merricks, Trenton. 2001. *Objects and Persons*. New York: Oxford University Press.
- Morrison, Margaret. 2012. "Emergent Physics and Micro-Ontology." *Philosophy of Science* 79 (1):141–66.
- Nida-Rümelin, Martine. 2006. "Dualist Emergentism." In *Contemporary Debates in Philosophy of Mind*, ed. Brian McLaughlin and Jonathan Cohen. Malden, MA: Blackwell.
- O'Connor, Timothy. 1994. "Emergent Properties." *American Philosophical Quarterly* 31 (2):91–104.
- O'Connor, Timothy, and Jonathan Jacobs. 2003. "Emergent Individuals." *Philosophical Quarterly* 53:540–55.
- Shoemaker, Sydney. 2007. *Physical Realization*. Oxford: Clarendon Press.
- Skiles, Alexander. 2016. "Emergence Reinflated." *Philosophical Quarterly* 66 (265):833–42.
- Taylor, Elanor. 2015a. "An Explication of Emergence" *Philosophical Studies* 172:653–69.
- Taylor, Elanor. 2015b. "Collapsing Emergence." *Philosophical Quarterly* 65:732–53.
- Taylor, Elanor. 2017a. "Only Explanation can Reinflate Emergence." *Philosophical Quarterly* 68 (271):385–94.
- Taylor, Elanor. 2017b. "Explanatory Emergence as a Guide to Metaphysical Structure." *Philosophica* 91:15–48.
- Wilson, Jessica. 2010. "Non-reductive Physicalism and Degrees of Freedom." *British Journal for the Philosophy of Science* 61:279–311.
- Wilson, Jessica. 2013. "Non-linearity and Metaphysical Emergence." In *Metaphysics and Science*, ed. Steven Mumford and Matthew Tugby, 201–229. Oxford: Oxford University Press.
- Wilson, Jessica. 2015. "Metaphysical Emergence: Weak and Strong." In *Metaphysics in Contemporary Physics: Poznan Studies in the Philosophy of the Sciences and the Humanities*, ed. Tomasz Bigaj and Christian Wuthrich, 251–306. Leiden: Brill.
- Wilson, Jessica. 2021. *Metaphysical Emergence*. Oxford: Oxford University Press.