Why Physicalism Fails to Exclude Strongly Emergent Properties

John Symons

University of Kansas

1. Introduction

Many recent anti-emergentist arguments in the philosophy of mind have assumed that physicalism provides a complete fundamental ontology. This completeness is thought to make strongly emergent properties impossible.[1] In response, proponents of emergence deny that a purely physical ontology has the resources to account for all metaphysically basic features of reality.[2] Physicalism has been criticized as question-begging, while emergentism has been criticized for countenancing mysterious or self-contradictory properties. Both sides contend that some feature of the other makes it metaphysically unsustainable. Progress in this debate depends largely on our understanding of the completeness of physicalism.

I will argue that the prospects for deriving anti-emergentist conclusions from the completeness of physicalism are bleak and that the overall balance of considerations tips in favor of emergentism. While there are non-metaphysical reasons to favor physicalism, the empirical considerations that can be brought to bear in its favor are insufficient to preclude the possibility of strongly emergent properties.

Physicalism has faced a number of important conceptual challenges since the middle of the 1990s. In his recent book, Daniel Stoljar (2010) presents the most important arguments against the view. While physicalism is unlikely to maintain its dominant position among philosophers, it continues to hold some appeal insofar as alternatives like emergentism seem even less viable and have a strong whiff of mystery about them. Right-thinking, scientifically-informed common sense tends to assume that some version of physicalism can serve as our default ontology while allowing that there might be some lingering difficulties at the level of detail. In recent decades, we have adopted the habit of thinking that the only way to stand with scientific rationality is to embrace physicalism. Very roughly, the epistemic virtues that we tend to associate with scientific rationality include an aversion to miracles and mysteries, and an attraction to the prospect of non-enthymatic, rationally satisfying explanations. However, these

---

[1] Jaegwon Kim provides the clearest and most influential example of the argument against strongly emergent properties. According to Kim, putatively higher-level properties are causally preempted by their underlying physical constituents. Given the causal closure of the physical world, this preemption admits no exceptions. (Kim 1998)

[2] Terrence Horgan (quoted in Tim Crane 2010) puts the difference as follows: "A physicalist position should surely assert, contrary to emergentism … that any metaphysically basic facts or laws -- any unexplained explainers, so to speak – are facts or laws within physics itself." (1993, 560)

virtues can all be preserved without accepting physicalist ontology. In fact, as I will argue in this paper, cleaving too fervently to physicalism risks falling prey to the vices that naturalists hope to avoid.

2. Sketch of the argument ahead

The argument of this paper begins from the observation that any cycle or set of cycles in a system of states can have as predecessors what are known in engineering and mathematics as transients. As we shall see, transients can precede any cycle or system built out of any set of rules. I argue that this general result applies to any proffered set of fundamental metaphysical rules in the same manner that it applies to cycles of states in the formal computational or engineering contexts. In what follows, I will refer to the rules governing the systems in question as *generative fundamentals*. Generative fundamentals are the set of total states of a system and the possible transformations on that set. For the purposes of the present paper, a transient will be defined simply as a sequence of states of a system that has a first member.[3] The possibility of transients suffices to guarantee that any generative fundamentals can result from other generative fundamentals. We will consider cases where the previous generative fundamentals are inaccessible from the perspective of agents in some system governed by new generative fundamentals. Many of the properties in these systems are such that, relative to the successor or predecessor system, they can be called emergent.

The kind of emergence exhibited by these systems can be called strongly emergent insofar as the novel system's generative fundamentals differ from the system that preceded it. Given the account of transients presented below, the generative fundamentals governing the later system are not sufficient to account for all the metaphysically basic features of reality. The purpose of the argument from transients is simply to note a limitation on attempts to use the completeness of some set of generative fundamentals as the basis for an argument against emergence.

After describing how we can characterize the role of ontological fundamentality in our metaphysical theory in general terms, I will describe the ideas of generative fundamentals and transients in detail. From there, the paper explains some of the kinds of scenarios that physicalists must exclude a priori in order to block the possibility of emergence. Specifically, we will consider cases where a complete physicalist ontology excludes interactions from which new generative fundamentals can emerge. I will argue that some of these candidates for exclusion are perfectly respectable from a

[3] This way of understanding transients is similar to the concept of transients as they appear in a Markov chain analysis. However, the present analysis is dramatically simplified and does not involve probability or any assumptions concerning randomness. In a Markov chain if there is some non-zero probability that the system will never return to a state, we say that this state is transient. If a state is not transient, it is recurrent. In this paper, a transient will be defined as a sequence of states, not a single state. For an overview of Markov chains see Booth (1967)

naturalistic perspective. Given that interactions of this kind admit of the kinds of explanation, investigation, manipulation, and intervention that naturalists hope for, I concludes that the price the physicalist pays for excluding emergence is unacceptably high.

3. Physicalism and Completeness

Philosophers in the late Twentieth Century found many good reasons to accept physicalism and it has played an especially prominent role in the philosophy of mind.[4] Physicalism has taken a variety of forms. (Stoljar 2010) Despite this diversity, we can identify a set of core commitments that most physicalists would endorse. Most obviously, physicalists provide the following criterion for distinguishing the real from the unreal: To be real is to be constituted or somehow determined by the fundamental physical constituents of the universe.

Getting clear this criterion involves clarifying some underlying assumptions concerning metaphysical fundamentality. The first is the assumption that *the physical world is causally closed.* Another widely held assumption is that at the fundamental level *being real involves the possession of some unique causal power.* (Kim 1999) A third also concerns individuation and is equivalent to what has come to be known as *Hume's dictum*. Jessica Wilson explains Hume's dictum as the claim that "there are no metaphysically necessary connections between distinct, intrinsically typed, entities." (2010, 1) The converse of Hume's dictum is the claim that if there *are* metaphysically necessary connections between entities or/properties then they *are not* distinct.

I will argue below that physicalism does not exclude properties that are distinct of physicalist generative fundamentals. By being distinct from the generative fundamentals, I mean not derivable from the generative fundamentals and not included in the set of states stipulated by the generative fundamentals. More importantly, in attempting to prevent the existence of properties that are metaphysically distinct from the physical fundamentals, the physicalist is forced into an implausibly restrictive view of what can count as fundamental.

While the history of science might encourage us to lean towards physicalism, ultimately, the question of whether physicalist ontology is complete will not be settled by empirical considerations alone. There are strikingly few examples of philosophers providing arguments for the completeness claim. At a minimum, such an argument involves ensuring that one's metaphysical system is complete in the formal

[4] As Sven Walter and Heinz-Dieter Heckman note: "the hallmark of today's physicalism is […] the contention that mental properties are either identical to or at least somehow realized, determined or constituted by physical properties. If judged only by head counting, physicalism was undoubtedly the uncontested champion of 20th century philosophy of mind." (2003, v)

sense; namely that it captures all the truths that can be formulated in the language of the system. Asking after the formal completeness of physicalism involves determining whether there are physical facts that cannot be derived from some set of initial conditions and the laws of physics. If all such propositions are derivable, physics can be said to complete with respect to the physical facts. The prospect of being able to capture all the truths of physics without recourse to an ontology that includes, at a minimum, some mathematical entities famously eluded naturalists like Quine. (1981) For the sake of this paper we shall grant that physics could be complete in this sense, although the prospects of this actually being the case seem permanently beyond reach.

Unlike the sense of completeness familiar from the study of formal systems, the metaphysical completeness of a system is not only a matter of capturing all the facts that can be stated in the language of the system. It also involves arguing for the adequacy of the language of the system. It is necessary to argue that nothing real is being forgotten or excluded from the system's inventory. So, in addition to formal completeness, metaphysical systems should be evaluated relative to their power to represent the world adequately. Standard criticisms of physicalism in the philosophy of mind target the adequacy of physicalism in this way.[5] Physicalists need to explain how metaphysically significant explananda like qualitative experiences or numbers can be reconciled with or explained in terms of one's physicalist ontology. Thus, from the perspective of the metaphysician, it is not simply enough to claim that physicalist principles suffice to capture all the physical facts. One must also explain why it is that there are no non-physical facts, where a non-physical fact is a fact that is not identical to a physical fact or somehow determined by physical facts.

Contemporary physicalism is an ontological doctrine.[6] Physicalists assume that there is some sense in which physicalism can be understood to provide a complete inventory of all facts about the actual

[5] The knowledge argument, qualia inversion arguments (Lycan 1973, Shoemaker 1982) zombie arguments (See Chalmers 1996 especially Chapter 3) and other arguments involving the specialness of phenomenal judgment and qualitative experience are directed at the view that physicalism fails to account for central features of conscious experience.

[6] As Stoljar and others have noted early 20th century physicalism was not a straightforwardly ontological position. In the Vienna Circle, physicalism was a thesis about the language of science and it differs in important ways from 19th century materialism . For Carnap "the general thesis of physicalism" is that "physical language is a universal language, that is, a language into which every sentence may be translated" (1959/1932:165). On this view physicalism played a similar conceptual role to that played by phenomenalism. Carnap suggests that the choice of physicalism over phenomenalism as the basis for a universal language is purely pragmatic (1947:207 – 8). The sole advantage of physicalism over phenomenalism is that the physical language is intersubjective and more suitable for scientific inquiry (1959/1932:166). Neurath's position is a little more difficult to interpret. He does not see physicalism predicated on the assertion of the dominance of physics over the other sciences. Instead he regards it as a kind of background condition for the coordination of distinct kinds of scientific practice. Both Carnap and Neurath clearly regarded physicalism as a linguistic rather than a metaphysical doctrine. Stoljar and others have noted that Carnap sometimes claims that physicalism is the language of ordinary physical objects and their distinctive properties rather than being directly connected to physics per se, or even to the disciplinary language of physics.

state of the world.[7] While there are a large set of powerful considerations to the contrary, even if we assume that all current anti-physicalist arguments can be answered, I will argue it still couldn't block the possibility of strong emergence. Furthermore, even if some suitably supplemented or ideal version of generative fundamentals provided a complete explanation of all facts about present and future states of nature without recourse to emergence and even if these generative fundamentals seemed to provide a deterministic story about past states of the system, there remains the possibility that its description of the past could mistakenly exclude the possibility of emergence. Thus, appeals to the completeness of a metaphysical system - including arguments that depend on the causal closure of the physical and the maximal generality of physics – do not rule out the possibility of emergence.[8] In order to rule out that possibility the physicalist must take a further step. This step leaves physicalism vulnerable to the criticism of being ad-hoc or question-begging in its dispute with emergentism: Let's call it the "And that's all there is" step.[9]

4. Books of the world

One way to understand the behavior of a system is to specify the possible states it can occupy and to provide some account of how the system changes from state to state. With this in mind, we can reconsider the problem of emergence in terms of the relationship of the putatively emergent property to some specified set of states and transformations. Different kinds of fundamentality will result in different

---

[7] Fodor (19??) has claimed that physicalism is an a priori doctrine but gives no argument for the claim and as Stoljar has noted, it is difficult to imagine what a good argument for it would look like.

[8] Tim Crane describes the role of the generality of physics claim in current emergence debates as follows:

> The claim that all properties have physical properties (the denial of Cartesian dualism) I shall call the generality of physics… We believe that the laws of physics apply unrestricted lyacross the universe; there are no regions where these laws fail or break down…for the laws to have this generality, then all the objects to which they apply must have the kinds of properties which these laws concern: physical properties. Everything in space-time has (or has parts that have) these properties: for example, mass, temperature, electrical charge, and so on. (2010, 28)

[9] The most prominent recent example of the claim that one's some basic ontology can provide a complete inventory of all the facts is David Lewis' Humean supervenience. He famously describes Humean supervenience as:

> …the doctrine that all there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another. (But it is no part of the thesis that the local matters are mental). We have geometry: a system of external relations of spatiotemporal distances between points. Maybe points of space-time itself, maybe point-sized bits of matter or aether or fields, maybe both. And at those points we have local qualities: perfectly natural intrinsic properties which need nothing bigger than a point at which to be instantiated. For short: we have an arrangement of qualities. And that is all. There is no difference without difference in the arrangement of qualities. All else supervenes on that. (1986 b , ix)

Lewis's "And that is all" claim, is an example of the extra condition that is necessary for putatively metaphysically complete systems to avoid the threat of strong emergence.

sets of states and transformations. Let's call these sets of possible states and transformations, books of the world.[10]

The physicalist favors some books over others. In our time, physicalists will reject books whose narratives violate physical conservation principles. These principles serve as meta-transformations that govern physical reasoning about possible transformations from state to state. To accept these or any other meta-level constraint involves reducing the space of possible books of the world from, for example, the space of logically possible books to the space of what we might call the extended nomologically possible books of the world.[11] Restricting books via the conservation laws still includes worlds with alternative physical laws to the actual.

Conservation principles are simply features of our physics that are held to be invariant over time. There are a variety of ways that a book might exhibit this kind of invariance. For example, given some set of *n* possible states, we might be concerned with the set of possible cycles through those states which preserve information and include all states. Let's call these *1-cyle books*. So, for example, given a set of four possible states {a,b,c,d} and the constraint that our book contain a single circuit and conserve information, the following books of the world would be possible :

Book1	Book 2	Book 3	Book 4	Book 5	Book 6
a to b	a to c	a to d	a to d	a to d	a to b
b to c	c to b	d to b	d to b	d to c	b to d
c to d	b to d	b to c	b to c	c to b	d to c
d to a	d to a	c to a	c to a	b to a	c to a

These books tell a story in which an agent at any step in the successive unfolding of the system would, in principle, be able to determine the entire sequence of steps in the history of the system. Even if an agent did not know how many possible states there are in its world, given the knowledge that the universe behaves like a 1-cycle book, the agent could be sure that there are *n*-1! sets of possible books. Why would a metaphysician confine herself to 1-cycle books of the world? The advantage of these books is that given knowledge of a single state of the system at an instant, an agent who has access to the book would know everything about the system. Ordinarily, physicists are not that optimistic and neither, I suppose, are physicalists. They recognize, instead, that the conservation principles permit the existence of

[10] The set of states and transformations that a metaphysical theory provides is equivalent to what Peter Railton has called its complete explanatory text. (1981)

[11] Nomologically possible worlds are those which obey the same laws of physics as the actual world. The set of extended nomologically possible worlds are those which obey some metaprinciples of physics, like the conservation principles, but which do not necessarily contain the same physical laws as our own.

independent subsystems. What do we mean by subsystems here? Well, in addition to the six cases we described above, we could add cases like the following

Book 7	Book 8	Book 9	Book 10
a to b	a to c	a to a	a to a
b to a	c to a	b to c	b to b
c to d	b to d	c to d	c to c
d to c	d to b	d to b	d to d

and so on. If we allow subsystems that are independent of one another and that obey conservation principles, we will find $n!$ possible books (24 in this case). We could call this larger, more inclusive, set *the multi-cycle books*.

Alternatively, we could imagine not caring about the conservation laws, such that we would drop any restriction on possible transformations among states. In this case there would be n^n sets of books where n (again) is the number of possible states that the system can be in. In this case, we could say that our choice of book would be constrained only by logical possibility. Most of the stories included in this set of books would violate conservation principles of various kinds.

As we begin to read one of these forbidden books we quickly encounter a phenomenon known as transience. A transient state is simply one that is part of a sequence of states with a first member. If the state is part of a finite sequence of states which does not have a first member, then that state is in a cycle. Transients are anathema to physics insofar as they fail to be deterministic into the past. The laws of physics as we know them are symmetrical or time reversal invariant. Physicists say that laws are invariant under time translation. Transients are not. In a system with set of states {a,b,c,d} dropping any restrictions would allow for transformations that go:

a→a
b→a
c→a
d→a

or

a→b
b→a
c→a
d→c

Any system whose possible states and transformations allow for more than one arrow out or in, or whose states lack an in arrow would have transients. Theoretical physicists might not like transients, but in nearly all other scientific contexts, they are ubiquitous. For example, any science which uses computational models will have to find some way to cope with transients. Any book of the world that is generated by computational modeling is likely to exhibit behavior that does not comport with conservation principles of the usual sort. In cellular automata, [12] for example, transients will appear in most runs of the system and systems that exhibit reversibility or conservation will be a tiny minority.[13]

In systems like cellular automata, initial states follow transition rules such that they tend towards an oscillator. This means that after some finite sequence of steps, the cellular automaton will enter into a repeating cycle of states. Note that these systems do not preserve information concerning past states of the CA. Instead, they feature transients that are not accessible once the system has entered into an oscillating sequence of states. Clearly, the algorithms governing CA are different from the kinds of transformations we would usually associate with physical laws.

Earlier we saw how one could constrain one's choice of book in accordance with conservation principles. Notice now that an agent whose universe is the result of fundamentals organized according to the algorithm of a cellular automaton might be in a position to propose a book that accords with the sequence of states in the oscillator. We could imagine an agent at some state in the oscillator, believing that she had accounted for the entire past, present and future of her universe while ignoring the possibility that her present cycle is compatible with another book in which the rules could have generated transient states prior to the initiation of the first run of the oscillator that she now inhabits. We will return to this point in greater detail below.

5. The Book of Φ

Returning to systems that are constrained by the kinds of conservation principles that interest physicists we recognized that books obeying conservation principles include what we called multi-cycle books. As we saw above if we include isolated loops and subsystems in a system of n states, the number of transformations that preserve invariance is $n!$. So, to summarize, this $n!$ set of books would include transformations with discrete subsystems but it would exclude books with transients.

---

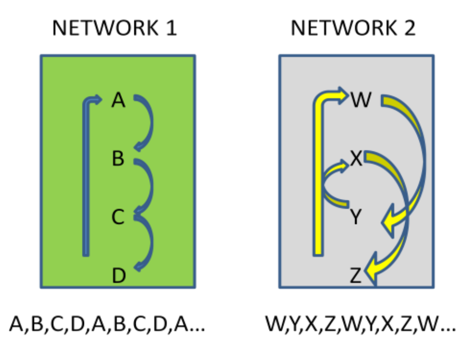
[12] Cellular automata are abstract objects which can be characterized in terms of a quintuple set: {Cells, Cell Space, Cell State, Neighborhoods, Rules}. Where *cells* are the basic objects or elements of the CA each having some individual state depending on the rules of the CA. *Cell space* is defined as the set of all cells and their values at some time. *Neighbors* are the set of cells surrounding some any center cell and *rules* are the transition functions of cell states, mapping cell spaces to cell spaces. (Hu Richa. and Xiaogang Ru 2003, 1047) The rules of the CA are defined as being maximally general with respect to the cells in the model and the application of rules updates each cell synchronically.

[13] Steve Wolfram showed that among the 256 elementary cellular automata with two colors and nearest neighbor rules for transformations, only 6 exhibit a reversible output. (2002, 436) Irreversibility is a characteristic of most cellular automata. It's worth noting too that the elementary cellular automata constitute a small subset of the set of algorithms which we might wish to consider.

Given that there are n! systems of fundamentals, some of which can have distinct cycles or subsystems, let's consider an example of their behavior. When we think about these possible, physically law abiding subsystems, it can be fruitful to think of them as networks with their own specific sets of laws. At this stage we can consider both networks as written in a book which we will call Φ.

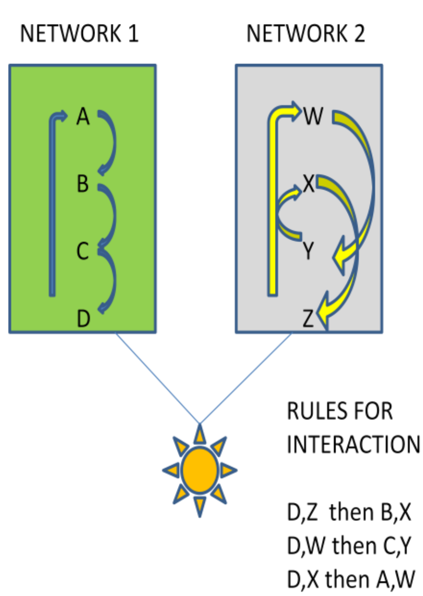
Networks 1 and 2, pictured below, are a representation of the story told by Φ. At this stage one might ask whether it is legitimate to understand these as isolated, law governed systems or as consequences of the laws governing some system that is composed of networks 1 and 2. In this case, let's simply stipulate that Φ describes two distinct oscillators or networks and that Φ is also consistent with conservation principles. Φ does not specify any interaction between the two networks.

Now, imagine, for the sake of the present argument that one is interested in possible ways that these two networks *could* interact. While no such rules for interaction are included in Φ, one could ask whether Φ excludes interaction or simply leaves such interaction undetermined? From the physicalist perspective, the only books worth having are complete books and therefore we should assume that interaction of Networks 1 and 2 is excluded by Φ.

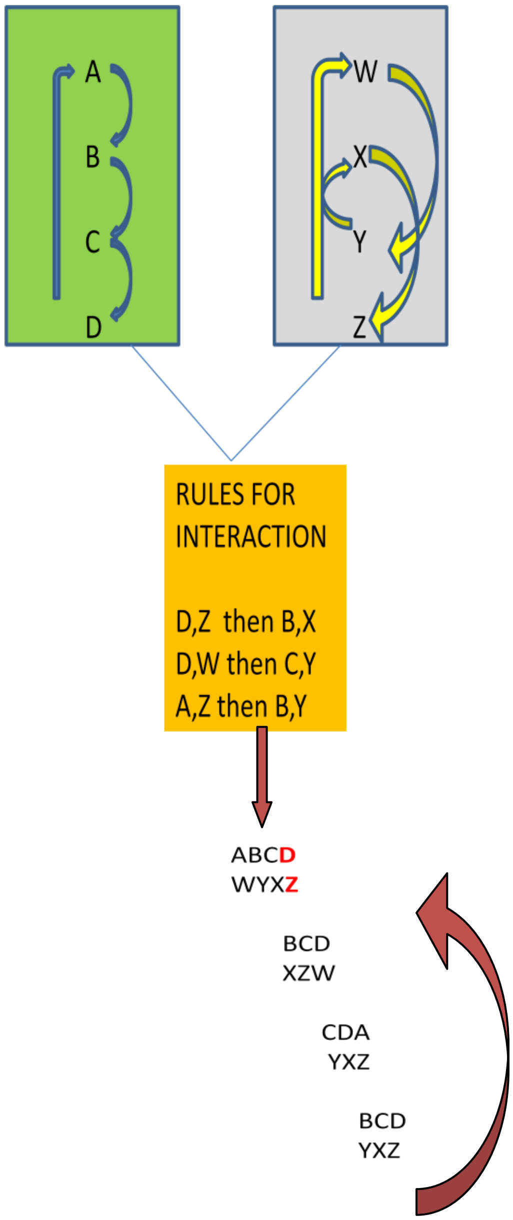


The methodological problem with the physicalist's view of fundamentals is becoming apparent. When building a set of fundamental metaphysical principles, surely one cannot simply claim completeness by specifying the fundamentals and then adding the condition that there are no events other than the events determined by the fundamentals. This would be almost as bad as claiming the completeness of *Principia Mathematica* by simply stipulating that it captures all truths of arithmetic; that no truths of arithmetic are unprovable in the system of PM. So, when we define Φ for any sufficiently complex or interesting cases, the anti-emergentist cannot simply forbid the interaction of sub-systems by fiat without begging the question against emergentism. I will provide a more extensive defense of this claim below.

For now, and solely for the sake of argument, let's assume that there are possible interactions between the two networks that are contingent with respect to Φ. This would mean that Φ was an incomplete book of the world. For the purposes of the present argument, the source of the interaction rule is not important. So, for example, given our two networks described above, let's arbitrarily introduce some new rule, call it the interaction rule, which gives us a new book Φ*.



Φ*, pictured here, is a new book which results from adding interaction rules to Φ. So, beginning with A in Network 1 and W in Network 2, let's consider the way that the systems unfolds in light of the interaction rules specified. Notice that here the interaction rules are taking as their inputs, the results of the action in the distinct networks. Let's assume that the system begins with A in Network 1 and W in Network 2. If the Networks give as outputs the antecedents of the interaction rules then the activation rule determines the consequent. Twelve steps after beginning at AW we notice that the system returns to DZ but that the combination of A and W drops out and will not reappear in the future runs of the system. This is an example of what we earlier called 'transient'. The steps are illustrated in the figure below.


A
B
C
D
W
X
Y
Z
RULES FOR INTERACTION
D,Z then B,X
D,W then C,Y
A,Z then B,Y
ABCD
WYXZ
BCD
XZW
CDA
YXZ
BCD
YXZ

After an initial run and with the rules of interaction in place we can see Φ* as a new system governed by a new set of functions in which the combination AW no longer appears. The sequence of states that characterize Φ* are as follows:

DBCDCDABCD...
ZXZWYXZYXZ…

and the set of functions that characterize the behavior of the cycle that Φ* produces (given initial starting point at AW) are as follows:

If A then B
If B then C
If C then D
If DBCD then C
If ABCD then B
If W then Y
If X then Z
If Y then X
If ZXZ then W
If YXZ then Y

Interaction rules can introduce transient states (like A,W) and can generate new patterns that are subject to generalizations that do not hold, or are even in violation of the rules given in the networks in isolation. So, for example, Φ* generates a sequence in Network 1 such that: ABCDBCDCDABCD… In this sequence, DBCDC is clearly a violation of the ABCD rule. However once we can only observe the behavior of the cycle that Φ* produces after the transient state has passed, the notion that there really are relevant Network 1 rules to violate becomes difficult to see.

The starting points or the initial conditions are important to the role of the interaction rule we introduced to generate Φ*. So let's consider some alternatives. Alternative initial conditions result in alternative sets of transformations governing the networks. For example, beginning with DX also results in a transient state, but in this case, it leaves the original rules governing Network 1 unchanged. By contrast, the initial conditions plus the interaction rules modify the behavior of Network 2.

D A B C D B C D C D A…
X Z Y X Z X Z W Y X Z…

(D,X) is left transient, Network 1 is unchanged, Network 2 has appears to be governed by new rules.

We could canvass other possible sets of initial conditions. For example, starting Φ* at BZ does nothing. It never allows the activation of the interaction rules. In this case, it would not be possible to distinguish between the behavior of Φ* or Φ. The interaction rules would be forever dormant and would never have had the occasion to manifest. [14]

8. Blocking Interaction

So, what can we conclude from toy cases like these? In the way I set up the case here, initial conditions are arbitrary with respect to the rules of the networks as are the rules for interaction. Once we include the rules for interaction, depending on the initial conditions, the resulting system Φ* can exhibit transient states. What should be striking is that once it enters into its cycle, the resulting system will show no evidence of these transient states and may instead exhibit a new set of rules governing both networks.
The transients aren't conserved in the new arrangement Φ*.

Φ* can, given some initial conditions, comprise a new set of functions on the original set of states. Note that in our example, in order for the new arrangement to be in place, the old functions are required to govern the sub-network. In this sense, the original laws are subordinated to the rules of interaction. From the perspective of the physicist living in the new cycle with its new fundamental laws, the transients are simply irrelevant; they aren't conserved. As we saw, the initial conditions are also crucial in the specification of the consequences of interaction. Given some sets of initial conditions we can have a new total system with a new set of transformations in Φ* or we can see no difference at all between Φ* and Φ.

Obviously, the new arrangement only comes into place if there is the possibility of interaction. If there is no interaction, then there will be no question of the interaction laws trumping the local or sub-network laws.

The physicalist must make sure that there are no interactions of this kind in the future. One way to block the possibility of interaction between sub-networks in a system, would be to make all sub-networks fall into a single cycle such that there would be no genuine subsystems. This would require us to rule out books with subsystems, even though they are in compliance with the conservation principles and form a large part of the extended nomologically possible set of books described above.

As we can see, in order to block the possibility that distinct subsystems might interact, the anti-emergentist must limit consideration of the extended nomologically possible, to those systems which do not permit emergence. More precisely, these would be systems in which apparently distinct subsystems are in fact coordinated. This coordination of the subsystems would serve to preclude the possibility of interaction effects of the kind discussed here. However, this blocking strategy for the fundamentalist

[14] Elsewhere, (Boschetti and Symons (2011)) the mechanism for emergence that is sketched below is explored in greater detail.

comes at the price of constraining consideration of the set of books in an apparently question begging manner. Not only are we restricting the books of the world to those that comply with conservation principles, now we are ruling out books that have subsystems that obey conservation principles, just in case they might interact in the future. Since they don't say they won't, then let's stipulate that they won't in order to block emergence.

That price might be worth paying if it really blocked emergence and if it provided completeness for physicalist fundamentalism. Unfortunately for the physicalist it fails to do so. It is logically possible that we could live in a world that is accurately accounted for by a finished physics with a set of functions that determine the behavior of the system and are deterministic in the forward and reverse time directions. It is also possible, however, that these physical laws are the product of a process like Φ^*. If so, then there may have been a history over the course of which there was at least one transient state. As we have seen, it is possible therefore that there is no way to recover this history from exhaustive knowledge of the present laws of nature. These laws of nature may have emerged from different laws via a simple process whose character cannot necessarily be gleaned from the current laws themselves. As such, the laws or fundamentals in this case are not complete.

To summarize, as we have seen, even in cases where there is only one system or law which does not admit of any genuine novelty, it's always possible that this law-like system is itself an emergent consequence of a natural history which includes transient states. Even a closed, cyclical system governed by laws preserving backwards and forwards determinism could itself be an emergent feature of some prior or more fundamental metaphysical principles. Therefore, the fact that a book's laws govern the system in a way that is backwards deterministic might be the result of past processes that are not! This odd result stems from the simple fact that because our current state is part of a cycle we should not be led to the false conclusion that all prior states must always have been part of the same cycle.

While this result might seem to pose a merely skeptical challenge to physicalism, it does have a constructive element. The present account is constructive insofar as it shows how we might unpack the kinds of processes that led to some cycle or system. For example, it is possible that by isolating subsystems in the appropriate way, in our case by isolating the behavior of individual networks from the rules of interaction, it might be possible to reconstruct the mechanism by which the new system Φ^* emerged. In such a scenario, if we assumed that the Φ^* laws we observed at a later state were fundamental, then our set of natural laws would be the emergent product of interactions or mechanisms that would remain forever hidden from us. Notice that by isolating a sub-system we may be able to recover the laws governing that Network in isolation. So, if we know how to control interactions we have the prospect of rebuilding the process that led to the resulting system.

According to the view presented here, we should expect the possibility of nomologically contingent interaction of sub-systems. In such cases we could find an instance where the joint action of the two systems is such that the result is emergent with respect to those systems. In order to block this contingency we might posit some third more fundamental principle or system. If this third law subsumes the generalizations of the initial two subsystems, we revert back to a one-rule scenario. However, as we have also seen, even if the one-rule scenario accurately represents our current situation, it fails to block the possibility of emergence.

6. Conclusion

As David Papineau points out, the dominance of physicalism among philosophers of mind is largely due to a set of empirical considerations. Physicalists notice that over the course of the recent history of science, appeals to non-physical factors in scientific explanation have been associated with degenerate research programs and dead ends. By contrast, reductive methodological strategies and physicalist ontological assumptions have been associated with highly fruitful and progressive research programs in the natural sciences. (2001,7) Daniel Stoljar makes a similar point about the empirical basis for accepting physicalism. He notes, *pace* Fodor, that the denial of physicalism does not involve any obvious logical contradiction or conceptual error and compares its status to that of the theory of evolution or of continental drift. Denying physicalism, he argues, is not philosophically absurd, but it does put one in conflict with science and scientifically informed common sense. (2010,13) Papineau argues for the stronger empirical claim that physical science now has hegemony over other kinds of scientific inquiry and that this fact supports physicalist ontological views. (2001, 3) From the physicalist perspective, to reject the hegemony of physics and to advocate the ontological autonomy of the special sciences is to be on the wrong side of the history of science.

There are many prominent criticisms of the inference from the empirical successes of reductionist research strategies in Twentieth Century science to the conclusion that physicalism provides a complete ontology.[15] Most famously, Carl Hempel noted that while physicalists regard physics as their guide to determining the correct ontology, they surely do not mean to commit to the ontology given by the physics of the past and can be nearly certain that contemporary physics is not the end of the story. The choice then is between committing to the ontology of a false or probably false theory on the one hand or of

[15] The view that physicalism can serve as an adequate ontological framework has been subject to increasingly critical scrutiny. In his recent book, *Physicalism*, Daniel Stoljar outlines the most significant problems with the view, concluding that "physicalism has no formulation on which it is both true and deserving of the name." (2010, 9) Other prominent criticisms of physicalism include van Fraasen (1996), Hempel (1969, 1980), and Daly (1998). Robert Koons and George Bealer (eds.) (2010) contains a set of papers criticizing materialism from a variety of perspectives. That volume provides a good representative sampling of the kinds of arguments that anti-materialists have presented in recent years.

committing to the ontology of an ideal finished physics. The trouble with the ontology of the finished physics is that, for all we know, the finished physics might take, for example, qualitative mental states to be fundamental. (Hempel 1968) Relying on the history of science as evidence for the completeness of physical ontology is not a strategy that guarantees success for the physicalist.

In contemporary philosophy, the focus of most arguments concerning emergence and fundamentality has been the characterization of the scope and limits of physicalism. Specifically, these debates have involved disagreements concerning how much of the higher-level action in our metaphysics is packed into the micro-level goings on. In this paper, we have seen that the physicalist's view of fundamentality leads to an ad hoc rejection of naturalistically respectable books of nature. So, unlike previous criticisms of physicalism, the current argument is not concerned with the question of how much is packed into the fundamentals, but rather, with the ways physicalists exclude the possibility of interaction and emergence from consideration from the outset.

Traditionally, philosophers have responded to physicalism in various ways, arguing for instance that the explanatory role of high-level generalizations is such that they cannot be reduced to some lower-level set of facts without loss of explanatory power. Alternatively some philosophers point out that there are features of some concepts which require that their bearers have the appropriate historical features. Insofar as they connect with traditional criticisms of physicalism, the arguments I have presented here contribute to this second line of argument.[16] The physicalist's response to specific examples of non-physically determinable differences is either to provide reasons to believe that these differences are unreal, or to claim that the truth of sentences mentioning these differences can be explained via a supervenience relation on the real action at the micro-level.

Physicalist completeness claims are directed to present and future states such that it cannot exclude the kinds of transients discussed here, nor can it do justice to the kinds of historical properties that Louise Antony and others have noted. (Antony 1999) As such, physicalism cannot exclude the

[16] Consider the following worry about historical features: If malevolent aliens introduced some creature who was able to successfully breed with horses and who gave every appearance of being a horse, we would, intuitively deny that it is a horse by virtue of its not having the right kind of natural history. Alternatively, following a version of Louise Antony's example, (1999) one can imagine my microphysical duplicate appearing at the front door where he is summarily shot by one of my enemies. Perhaps the duplicate is sent by benevolent aliens. If physicalists believe that there are moral facts, then they must believe that they supervene on physical facts somehow and since he is a duplicate of me, for a physicalist, he should have the same moral standing as I. Yet it would be perfectly reasonable, we would think, for my friend to prefer that the microphysical duplicate be killed instead of me. My friend would value one of us more highly in virtue of his having had the right kind of history. Only one of us duplicates actually took care of my children, signed the papers for the mortgage, is the son of my parents, and so on. So, even though he would serve all my functions just as well, my microphysical duplicate is not actually the freind (in spite of what he might remember or believe), the co-owner of our house, an employee of my university etc. My moral properties and the truth value of my beliefs has something do with my history.

possibility that the physical fundamentals themselves are emergent features of some more fundamental metaphysical principles. As we have seen, it is possible to grant completeness with respect to present and future states and still allow for the possibility of emergence in the past. So, the argument presented in this paper provides a plausible, and naturalistically respectable way of making sense of the possibility of historical properties for which physicalist fundamentalism cannot account.

Even if we dismiss the possibility of transients as an exotic, skeptical worry about the past, as we have seen, the physicalist will have to find ways to exclude contingent interactions in the future. Insofar as it is complete with respect to future facts physicalism must exclude nomologically contingent possibilities of interaction whose results lead to new generative fundamentals. In other words, only events that are a necessary consequence of the fundamentals can take place. However, there is a price to be paid for completeness with respect to the present and future.

As we saw, one way of blocking nomologically contingent future interactions is to claim that possibilities for interaction are already somehow accounted for in the fundamental physical properties. On this view, we simply assume that the fundamentals determine the outcome of every possible future interaction such there are no nomologically contingent interactions. This effectively eliminates the possibility that there are systems with subsystems that obey conservation principles. This restriction on the space of possible books of nature is motivated solely by the threat of possible interactions.

Common sense tells us that interactions can change the character of their participants. For example, it is common for emergentists to argue that participation in structural wholes can change the nature of the parts. Examples are discussed in Sandra Mitchell's recent book (2009) and elsewhere. (Symons, 2002) In order to rule out this kind of novelty and especially the possibility that structures might exhibit synchronic downward causal powers, the fundamentalist must include all possible fates of the constituents in the fundamentals ahead of time.[17] In effect, each part contains within itself some feature that makes it determinable how it will be in all possible states of affairs and interactions. On this view, physical properties cannot be permitted any interaction not already determinable given the fundamentals alone.

So, while it appears that there could be worlds with distinct susbsystems governed by conservation principles, the physicalist regards them as only apparent possibilities. In fact, such worlds only appear to have distinct subsystems. In fact, they are coordinated by some more basic single-cycle book. The trouble with physicalism is that it excludes too many naturalistically respectable kinds of worlds solely for the purpose of blocking the possibility of emergent properties.

[17] This is presumably what Shoemaker attempts to do in his discussions of emergent properties. (Shomaker 2002)

References

Louise Antony (1999) “Making Room for the Mental Comments on Kim’s “Making Sense of Emergence*””Philosophical Studies* **95:** 37–44.

Batterman, R. (2001) *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction and Emergence*. Oxford: Oxford University Press

Bedau, M. (1997), “Weak Emergence**”** *Noûs*, Vol. 31, Supplement: Philosophical Perspectives, 11, Mind, Causation, and World 375-399

Booth, T. (1967). *Sequential Machines and Automata Theory* (1st ed.). New York: John Wiley and Sons, Inc

Boschetti, F., Symons, J. (2011) **“**Novel properties generated by interacting computational systems: A minimal model” *Complex Systems* 20, 151-164, 2011

Burks, A.W., ed. (1966) *Theory of Self-Reproducing Automata* [by] John von Neumann. Urbana: University of Illinois Press.

Chalmers, D. 1996. *The Conscious Mind: In search of a Fundamental Theory* Oxford: Oxford University Press.

Crane, T. (2010) “Cosmic Hermeneutics vs. Emergence” in Macdonald, C. and G. Macdonald (eds.) *Emergence in Mind.* Oxford: Oxford University Press. Pp. 22-34.

Daly, C. (1998) “What are Physical Properties?” *Pacific Philosophical Quarterly* 79: 196-217

Davidson, D. (1970) “Mental Events”, in Foster, L. and Swanson, J.W. (eds.), Experience and Theory. Amherst: University of Massachusetts Press, 1970. Reprinted in Davidson 1980.

Davidson D. (1980) *Essays on Actions and Events*. Oxford: Clarendon Press.

Gardner, M. and J. Barnes (1942) *Transients in Linear Systems: Studied by the Laplace Transformation*. New York: John Wiley & Sons, Inc.

Hempel, C. (1969). Reduction: ontological and linguistic facets. In *Philosophy, Science and Method: Essays in Honor of Ernest Nagel*, ed. S. Morgenbesser et al. New York: St. Martin's Press.

Hempel, C. (1980) *Comments* on Goodman's Ways of Worldmaking. *Synthese* 45: 193-99

Holland, J. (1998), *Emergence*. From Chaos to Order. Redwood City, California: Addison Wesley

Hu Richa. and Xiaogang Ru. (2003) “Differential Equation and Cellular Automata Models.” IEEE *Proceedings of the International Conference on Robotics, Intelligent Systems and Signal Processing.* 1047-1051

Humphreys, P. (1997) "How Properties Emerge," *Philosophy of Science* 64 1-17

Kim, J. (1998) *Mind in a Physical World*. Cambridge MA: MIT Press

Kim, J. (1999) “Making sense of emergence.” *Philosophical Studies* (95) 3-36.

Kim, J. (2006) “Emergence: Core Ideas and Issues” *Synthese* (151) 547-559

Koons, R. and G. Bealer (eds) (2010) *The Waning of Materialism*. Oxford: Oxford University Press.
Levine, J. (1983) ‘Materialism and Qualia: The Explanatory Gap’, *Pacific Philosophical Quarterly* 64: 4.

Lewis, D. (1986a) *On the Plurality of Worlds* Malden: Blackwell

Lewis D (1986b) *Philosophical papers*: *Volume II*. Oxford University Press, New York

Lycan, W. (1973) 'Inverted Spectrum'. *Ratio* 15:315-319

Macdonald, C. and G. Macdonald (eds.) (2010) *Emergence in Mind.* Oxford: Oxford University Press.

Machamer, P., L. Darden and C. Craver, (2000) "Thinking about Mechanisms" *Philosophy of Science* (67) 1 1-25.

Mitchell, S. (2009) *Unsimple Truths: Science Complexity and Policy*. Chicago: University of Chicago Press.

Quine, W.V. (1981) *Theories and Things*, Cambridge, MA: Harvard University Press.

Shoemaker, S. (1982) The Inverted Spectrum. *Journal of Philosophy* 79: 357-381

Shoemaker, S. (2002) "Kim on Emergence" *Philosophical Studies* (58) 53-63

Sider, T. (2011) *Writing the Book of the World*. OUP Oxford, 2011.

Stoljar, D. (2010) *Physicalism* New York: Routledge

Symons, J. (2002) "Emergence and Reflexive Downward Causation" *Principia* 6: 14-39

Symons, J. (2008) Computational Models of Emergence *Minds and Machines* (18) 475-491,

Symons, J. (2008b) "Supervenience" in U. Windhorst, M. Binder, and H. Hirokawa (eds.), *Encyclopedic Reference of Neuroscience*. Berlin: Springer.

Symons, J., F. Boschetti (2012) "How Computational Models Predict the Behavior of Complex Systems**."** *Foundations of Science*

van Fraasen, B. (1996) "Science, Materialism, and False Consciousness", pp. 149-181 in Jonathan Kvanvig (ed.) *Warrant in Contemporary Epistemology: Essays in Honor of Alvin Plantinga's Theory of Knowledge*. Rowman Littlefield, 1996.

Wilson, J. (2006) "On Characterizing the Physical" *Philosophical Studies* (131) 61-99

Wilson, J. (2010). "What is Hume's Dictum, and Why Believe It?" *Philosophy and Phenomenological Research*, *80*(3), 595-637.

Wolfram, S. (2002) *A New Kind of Science.* Champaign IL: Wolfram Media