

# Dependent Autonomy

Henry Blanchette

Metaphysics of Science  
Reed College

## Abstract

This is the abstract.

## 1 Introduction

This is the introduction.

A thought that a previous philosopher wrote about ([Wilson J. 2018](#)).

## References

- [1] Wilson J. 2018. *Metaphysical Emergence*
- [2] Fodor J. A. 1974. *Special Sciences*
- [3] Bedau M. 2018. *Downward Causation and the Autonomy of Weak Emergence*
- [4] Bedau M. 2008. *Is Weak Emergence Just in the Mind?*
- [5] Wolfram S. 1985. *Undecidability and Intractability in Theoretical Physics*
- [6] McLaughlin B. P. 1992. *The Rise and Fall of British Emergentism*
- [7] McLaughlin B. P. 1997. *Emergence and Supervenience*
- [8] Hempel C., Oppenheim P. 1965. *On the Idea of Emergence*
- [9] Searle J. 1992. *Reductionism and the Irreducibility of Consciousness*
- [10] Dennett D. C. 1991. *Real Patterns*
- [11] Humphreys P. 1997. *How Properties Emerge*

# Annotated Bibliography

**Wilson J. 2018. *Metaphysical Emergence***

**Chapters 1 & 2**

**Chapters 3 & 4**

**Chapter 5**

**Fodor J. A. 1974. *Special Sciences***

Consider and contrast two scientifically fundamental perspectives: the *Unity of Science* and the *Generality of Physics*. The Unity claim is actually stronger than the Generality claim; Unity of Science requires all sciences to have as their aim the construction of a physics-termed explanation of their studied phenomena, while Generality of Physics requires only that instances of studied phenomena be completely explainable in terms of physics. The Unity claim is, in fact, the central claim of Redictive Physicalism.

However, there is a seeming paradox in the Unity claim - it pronounces that the continued success of the special sciences is just more and more evidence that they ought to be discontinued; they stray further and further from an explanation in terms of physics. The *special sciences* are sciences that do not deal directly or appeal to physical explanations.

*Reductive Physicalism* is the view that all special sciences must reduce to physics. For some special-science relationship  $S_1 \rightarrow S_2$ , reductivism requires there be a reduction with to physical predicates  $P_1, P_2$  such that

$$S_1x \rightarrow S_2x \quad (1)$$

$$S_1x \rightleftharpoons P_1x \quad (2a)$$

$$S_2x \rightleftharpoons P_2x \quad (2b)$$

$$P_1x \rightarrow P_2x \quad (3)$$

This setup forms a sort of bridge between the special-science relationship and a physical relationship. Though, there are some problems with the existence of bridges like this and how it interacts with the meaning of  $\rightarrow$ .

A way to address the bridge problems is with *Token Physicalism*, where all events that the special sciences talk about are physical events.

- Token physicalism is weaker than materialism, which claims token physicalism and that every event falls under the laws of some science or other.
- Token physicalism is weaker than type physicalism, where every property mentioned in the laws of any science is a physical property. This in fact implies token physicalism.

- Token physicalism is weaker than reductive physicalism, which claims token physicalism and that there are no natural kind predicates in an ideally completed physics which correspond to each natural kind predicate in any ideally completed special science.

Every science implies a taxonomy of the events in its universe of discourse. It creates theoretically- and empirically- inspired vocabulary, which fall under the laws of the science by virtue of satisfying those predicates. Not every theoretical predicate is valid or good though.

To fix reductive physicalism, can allow bridge relations to be in the form

$$S_x \Leftrightarrow P_1x \vee \dots \vee P_nx$$

where  $P_1x \vee P_nx$  is not the kind of natural kind predicate in the reducing science. This allows for “bridge laws” to not really be laws, since they are not law-like. Reducing with this kind of bridge law looks like

$$P_1x \vee \dots \vee P_nx \Leftrightarrow P'_1 \vee \dots \vee P'_m$$

Thesis:

There are special sciences not because of the nature of our epistemic relation to the world, but because of the way the world is put together: not all natural kinds (not all the classes of things and events about which there are important, counterfactual supporting generalizations to make) are, or correspond to, physical natural kinds.

A way of stating the classical reductionism view: things which belong to different physical kinds ipso facto can have no projectable descriptions in common; that if x and y differ in those descriptions by virtue of which they fall under the proper laws of physics, they must differ in those descriptions by virtue of which they fall under any laws at all.

If science is to be unified, then all such taxonomies must apply to the same things. If physics is to be basic science, then each of these things had better be a physical thing. But it is not further required that the taxonomies which the special sciences employ themselves reduce to the taxonomy of physics. It is not required, and it is probably not true.

## **Bedau M. 2018. *Downward Causation and the Autonomy of Weak Emergence***

In this paper, Bedau explores a foundation of Weak emergence as *derivability only by simulation* (or *underivable except by simulation*).

The core precepts for the concept of emergence are:

1. Emergent phenomena are dependent (ontologically) on underlying processes.
2. Emergent phenomena are autonomous (causally) from underlying processes.

All concepts of emergent base on the concept of the emergence property. The out-springing concepts include emergent entities (which are entities with emergent properties), emergent powers (which are powers bestowed by emergent properties), etc.

A property  $P$  is underivable except by simulation if and only if  $P$  can only be derived from a system, even a system causally closed at the lowest level, through a complete simulation of the low-level causes (e.g. *gliders* in the Game of Life). Bedau poses that  $P$  is Weakly emergent.

Basically, the idea is the if a property is interesting and complex in a formal way, in that it unpredictable arises from the low-level properties of a system, then it is indeed emergent (and, specifically, Weakly emergent).

It is important to note that this is a metaphysical point about emergent properties which is independent from epistemological concerns. For example, just because we haven't yet found a derivation sans simulation for a property doesn't mean that it is emergent - it just *looks* emergent. Of course, a proof that there is not derivation is good, but such proofs are extremely hard to come by (especially because of the open  $P \stackrel{?}{=} NP$  problem.)

Truely emergent properties are distinct from another kind of property that arises in similar circumstances - so-called *resultant properties*. A property  $P$  is a resultant property if and only if it is predictable from the properties of its parts. Often times,  $P$  is just of a whole in and of itself and not an interesting emergence.  $P$  of the whole that merely holds because of its being a whole as defined. For example, "being a circle" is a property that applies to a collection of points, but no single point in that collection shares the property of "being a circle". However, this is uninteresting and is directly predictable from the definition of a circle and the properties of points, without having to do any sort of scaling computation to figure it out.

In the *Game of Life*, there are many good examples of properties that are underivable except by simulation. For example, the property "finitely-expanding". A worldstate that, after a finite number of time steps, reaches a maximum  $m \times n$  area in the world such that no matter how many more time steps are run no cell will be born outside the area.

There are examples of non-finitely-expanding configurations, such as the famous *Gospel Glider Gun* that creates gliders in intervals and sends them out into infinity. And there are examples of finitely-expanding configurations, such as any that consists of only non-interacting *still-lives*.

However, to derive if a worldstate is finitely-expanding, one needs to run a simulation. There is no general mathematical theorem that decides without doing the equivalent of a brute simulation. Even if a mega-computer could do the simulation instantaneous, it would still count as needing a simulation and thus "finitely-expanding" is an emergent property.

If, however, someone came up with a shortcut, it would prove that "finitely-expanding" is not emergent.

The main problem for emergence is that of *downward causation* i.e. *overdetermination*. If  $S$  emerges from  $P$  but  $P$  is synchronically and ontologically dependent on  $P$ , then if  $S$  causes a  $P^*$

then surely we can say that  $P$  causes  $P^*$  as well. But if the low level of the  $P$ s is causally closed, then why do we need to talk about  $S$  causing anything at all? It seems irrelevant and problematic in an overdeterminative way. But if  $S$  was causally inefficacious, then it loses its status of autonomy from  $P$  and is not really emergent in the first place. This is Kim's overdetermination argument.

But Kim's concerns are ill-founded. The idea of Weak emergence avoids the problems by yielding overdetermination as unproblematic, since it is causally reductionistic and uses ontological dependence. But such weak emergence is important because it clearly defines autonomous and relevant structures.

Additionally, Weak emergence is not synchronic. It introduces a diachronic perspective on causation, where, for example, in the Game of Life for a macrostate to manifest an emergent property many time steps need to pass. For a microstate to have a property, it must be manifested in one or two time steps. In this case, emergent properties are causally relevant and distinct from lower-level properties.

Conclusion:

The advent of modern philosophy is conventionally presented as the Cartesian triumph over Aristotelian scholasticism. An Aristotelian thesis that attributed nature on the basis of a rich dependence on generating context was supplanted by the Cartesian antithesis that attributed reductionistic essences independent of context. Computer simulations allow weak emergence to extend reductionism into new territory, but they do so by embodying the idea that something's nature can depend on its genesis. Thus, the macro can depend on the context-sensitive process from which it arises and by which it is maintained. In this way, weak emergence can be viewed as a new synthesis.

### **Bedau M. 2008. *Is Weak Emergence Just in the Mind?***

The goal is to present a basis for Weak emergence in terms of *explanatory incompressibility*. Then, once this is established, Bedau will defend the metaphysical nature of Weak emergence by critically appealing to its *dynamical* nature.

In this paper, Bedau shifts from "underivability without simulation" to "unexplainability without simulation", and talk of macro-states that have incompressible explanations rather than incompressible derivations. This will be important for defending Weak emergence's metaphysical legitimacy.

The definition of Weak emergence is: if  $P$  is a macro-property of some system  $S$ , then  $P$  is Weakly emergent if and only if  $P$  is *generatively explainable* from all  $S$ 's prior micro-facts but only in an incompressible way. This definition defines Weakly emergent macro-phenomena by the distinctive way in which we explain how they are generated from underlying micro-states.

A *generative explanation* of a macro-state is one that exactly and correctly explains how macro-events unfold over time - how they are generated dynamically. This assumes complete information about both the micro-causal dynamics that drive the system and the system's earlier micro-states

and boundary conditions. *Incompressible explanations* cannot be replaced without explanatory loss by shorter explanations that avoid crawling the causal web.

If an explanation is *compressible*, then explaining the macro-property arbitrarily far into the future takes some fixed and finite amount of explanatory effort, no matter how far into the future your explanation reaches. For example, phenomena in *ALL Life* are compressible, where *ALL Life* is the cellular automata setup with one rule: A cell is alive at a given time whether or not it or any of its neighbors were alive or dead at the previous moment.

Only some cellular automata, including the Game of Life, support incompressible explanations for contained phenomena. This has been rigorously proven for some i.e. the Game of Life.

Weak emergence comes in degrees, scaling from “immediately deducible upon inspection of the specification or rules generating it” to behavior that is “deducible in hindsight from the specification after observing the behavior” to finally behavior that is “impossible to deduce from the specification”. Similarly, Explanatory incompressibility can be arrayed into similar stages. So, since weak emergence depends on explanatory incompressibility, weak emergence also comes in stages or degrees.

Bedau’s definition of Weak emergence applies to all systems that have macro-behavior with only incompressible explanations. If we could directly identify what it is about micro-causal interactions that make them incompressible, then we might be able to construct a direct definition of weak emergence.

Study of such properties have yielded the following observations about the intrinsic properties of micro-causal dynamics that require incompressible explanations:

- Massively parallel micro-level populations of independent and autonomous agents that interact with their neighbors and, restrictedly to some degree, in their local environment
- Interactions among the agents and their environments are typically non-linear and synergistic, so that the behavior of an agent is highly sensitive to its local context

The yield that the behavior of complex systems are impossible to predict, even given complete prior micro info, without full simulation.

Many scientists and philosophers alike assume that emergence and reduction are incompatible. A typical form of reductionism is mereological supervenience a la Kim. But, usefully, Weak emergence is compatible with most reduction (as opposed to Strong emergence which is compatible with no reduction).

The distinction between explanations or reductions that hold only *in principle* versus those that hold also *in practice* is important. For example, a reductive generative explanation might exist to explain from macro to micro *in principle*, but be unhelpful for explaining weak emergent phenomena *in practice*. Reasons why could be

- Some relevant micro-details required for the explanation might be unknown and/or inaccessible
- the explanation might be too complex and tedious for anyone to work through without the aid of a computer simulation (but still workable in principle)

Simulation implies that all Weakly emergent properties do in fact have micro-to-macro explanations. This distinction between *in principle* and *in practice* helps defend that Weak emergence fits the two hallmarks of emergence: dependence of macro on micro, autonomy of macro from micro. The incompressible structure of the macro makes it independent from the micro.

“The subtle way in which weak emergence balances principles and practices is summarized with the awkward but apt notion of in principle irreducibility in practice.”

“We can put these points together by saying that weak emergent phenomena are in principle irreducible in practice.”

There is still a distinction of *robustness* of Weak emergence. For example: “Physicists in some instances have mathematically proved that the critical behavior of some large class of physical systems is insensitive to almost all details about the system, but in most cases one has merely empirical evidence that a physical system exhibits universal behavior. Nevertheless, this empirical evidence can be very strong.”

But why can’t we just say that Weak emergence is all in the mind? A worry: “The existence of in principle irreducible downward causation is an ontological matter, because it involves the real existence of a certain kind of causal process.” Most people studying emergence are concerned with finding possible *new powers* like in a Strongly emergent context. But Weak emergence differs precisely on this point. “This difference between reducibility in principle and in practice is the difference between strong and weak emergence.”

Other attempts with Weak emergence, like with Silverstein and McGeeve, brand Weak emergence as merely epistemic and not genuinely ontological. They define epistemological emergence to apply to any property that is “reducible to or determined by the intrinsic properties of the ultimate constituents of the objects or system” but is “very difficult for us to explain, predict, or derive ... on the basis of the ultimate constituents”. And a reductive consequence for macro phenomena immediately follows: “In principle, in such cases the higher-level feature, rule or law is a logical consequence of some lower-level feature, rule or law”

But the key observation to have is that this is just a result of definitions. It is “presumed that all apparently emergent phenomena are merely apparent, and have a true, reductive and non-emergent explanation. This implies that attributions of emergence are merely admissions of our ignorance of the true, reductive and non-emergence explanation”.

“If our best scientific theories construe certain phenomena as emergent (because in principle they are irreducible), that does not show us anything about nature.”



However, the Game of Life and similar cellular automata show clear counter-examples to this line of reasoning. We are clearly not missing the necessary scientific theories to explain behavior in the Game of Life. We know all the information. The epistemic essence of emergence that has been pointed to does not generalize to Weak emergence as Bedau defines. “If something has an indirect epistemological definition, it does not follow that it is just in the mind”.

A succinct argument for Weak emergence not just being in the mind:

the indirect epistemological definition is produced by and reflects a distinctive underlying ontological status or structure in nature. Incompressibility of explanations is a consequence of the objective complexity of the local micro-causal interactions that are ultimately generating the emergent behavior being explained. The micro-causal web is real and objective, and the incompressible causal pathways of weak emergent phenomena have a distinctive epistemological consequence. Note that the explanatory incompressibility that defines weak emergence applies to the explanations of any naturalistic epistemic agent, in principle. Just like us, any non-human epistemic agent will have to work through the objective complexity of the local micro-causal interactions. Thus, weak emergence is not merely in the mind, but refers to objective complexity in the objective natural world that is in principle irreducible in practice. (pp. 11)

Computer simulations are useful for talking about Weak emergence because they “produce some of the most striking examples of weak emergent phenomena”. The computer itself is exhibiting Weakly emergent behavior. So, computers are important as evidence for Weak emergence in complex natural systems (we have no practical alternative, considering the definition of Weak emergence. it would be hard to test such hypotheses in nature).

The indirect epistemological role of computer simulations in explaining weak emergence might fuel a revival of the belief that weak emergence is in some sense merely epistemological. But this would be a mistake. The weak emergence exhibited in jamming traffic and dividing vesicles is not merely epistemological. Traffic jams and vesicles require incompressible explanations because of their objective, intrinsic micro-causal complexity. **Traffic jams and vesicles are not just in the mind.** (pp. 13)

Finally, Weak emergence is special because of its *dynamical* nature. The canonical example of emergence is the mapping of a mental state to a brain state, *synchronically*. This kind of emergence is *static* (as opposed to *dynamic*) because of the synchronic conditions. But, this allows the trivial derivability from the macro-state from the micro-state. Weak emergence’s dynamic nature disallows this, there is no such trivial derivation since there are many time steps that are related differently and complicatedly at the high- and low-level. This is an ontological property, and “underscores why weak emergence is not merely in the mind”.

Conclusions:

Any naturalistic epistemic agent who tries to explain it [macro-state] will have to use incompressible explanations. Weak emergence is the macro-level mark of incompressible complexity in a network of micro-causal interactions. When the objective micro-causal web is sufficiently complex, all explanations of its macro-behavior are incompressible. The resulting weak emer-

gence is not just in the mind

**Wolfram S. 1985. *Undecidability and Intractability in Theoretical Physics***

**McLaughlin B. P. 1992. *The Rise and Fall of British Emergentism***

This paper is about the ideation and history of a certain emergentism that prospered at the beginning of the 20th century. Ultimate, it has all but dropped out of accepted philosophical discourse by now.

The center of British Emergentism is that reality is strictly hierarchical. there are hierarchies of laws and of entities. Of course, this still employs the “dependent but autonomous” view of emergence (as extracted by Bedau).

A classic example of this view is the claim that chemical bonding is a higher-level causal feature of molecules that is not explainable in terms of atoms (and was so for a long while). In this way, there is a separate set of laws that apply to molecules and atoms interacting (bonding) than just atoms and molecules not interacting. So the bonds are an emergent property that is dependent upon the participants, but pertains to a whole that is greater and more complicated than the low-level interactions of its parts.

This view can be extrapolated to say that the sciences form a great hierarchy of natural investigations:

$$\begin{aligned} \text{Physics} &\rightarrow \text{Chemistry} \rightarrow \text{Biology} \rightarrow \text{Psychology} \rightarrow \dots \\ \dots &\rightarrow \text{Sociology} \rightarrow \text{Political Science} \rightarrow \dots \end{aligned}$$

Each level is distinct not by specificity, but by following the actual metaphysical and lawful differences in reality. The laws of physics are not enough to describe the laws of chemistry, and so on.

Another center is the idea of causal non-composability. If  $A$  is an emergence from some  $a_i$ , then saying that  $A$  causes  $B$  says nothing on the same level of any  $a_i$  causing  $B$ . And if an  $a_i$  does cause  $B$ , then the same argument applies visa versa. So, it seems that in emergent cases, there is something special about an  $A$  being emergent rather than just a group. For a group can easily satisfy this causal non-composability.

The chemistry example is key because during the rise of British Emergentism, there was strong intuition and evidence behind the idea that chemical bonds involved causal powers that were not merely physical (at the lowest level). But later in the 20th century, quantum mechanics offered just such a physical explanation. If this could happen for the primary motivating case for British Emergentism, why not for the other cases? And so this emergentism it is not held in high regard even if we don’t have an explanation of all of biology in terms of physics.

**McLaughlin B. P. 1997. *Emergence and Supervenience***

**Hempel C., Oppenheim P. 1965. *On the Idea of Emergence***

The key observation is that people attribute to emergence much more than they should. For example, emergent phenomenon are commonly thought of as

- unpredictable
- mysterious
- a manifestation of newness

But, upon close analysis, this paper concludes that these intuitions about emergence would leave emergence as an empty concept - nothing could possibly satisfy it. Even the classic examples of emergent phenomenon, like the wetness of water, are still predictable given the properties of their parts.

Rather than thinking of emergence as an ontological even, think of it as merely epistemic. Consider psychology. Our brain is made of chemicals and we do study the chemistry involved in the processes of psychology, but we do not have an explanation of high-level psychological concepts in terms of the low-level chemistry (or physics, for that matter). So, the psychological phenomena are emergent from the chemistry, but they are not metaphysically new or unpredictable. We just, with our present theories and understanding, have not explained the high level in terms of the low level. So, as an epistemic condition, psychological properties are emergent. But when we do figure such an explanation, psychology will no longer be emergent.

**Searle J. 1992. *Reductionism and the Intractability of Consciousness***

**Dennett D. C. 1991. *Real Patterns***

Consider the question “are beliefs real?” Is it correct to *believe* in centers of gravity? Why or why not? There are examples of people that argue these kinds of beliefs are obviously unreal and others that argue that they are obviously real (e.g. Dretske).

What about the case of a completely arbitrary *x*-center, such as the center of Dennett’s lost socks. Are such concepts and beliefs about just as real as with other, more “legitimate” *x*-centers? Or does deciding the reality of concepts and beliefs about just to do with usefulness or interestingness?

A question that naturally arises is “should we treat mental states/patterns (e.g. belief) as *real* in the same way and to the same degree as other patterns, such as electrons?”

Dennett gives the canonical *Bar Code* example. It is unclear that the underlying Bar Code pattern is there for all of the cases, except for *E* and *F* where it starts to get iffy. In fact, in *F*, it is

technically indistinguishable from random noise, from strictly a retrospective perspective (without looking at the generating code).

Dennett refers to Chaitin's idea of *incompressibility*, in that patterns are recognizable because they are compressible while noise is incompressible. But compressibility comes with two degrees of freedom: accuracy and simplicity. Usually, it is a trade-off from one to the other. It is merely a design choice for which is more important in which circumstance, and that it isn't an inherent metaphysical fact about the reality of the concerned patterns themselves. There are also good examples of this trade-off in the *Game of Life*.

A final thought:

Fine tuning could of course reduce these probabilities (of modeling which method was used to create a bar code pattern), but that is not my point. My point is that even if the evidence is substantial that the discernible pattern is produced by one process rather than another, it can be rational to ignore these differences and use the simplest pattern description (e.g. bar code) as one's way of organizing the data ...

## **Humphreys P. 1997. *How Properties Emerge***

The exclusion argument:

1. If every event  $x$  is causally sufficient for every other event  $y$ , then there is not  $x'$  distinct from  $x$  that is causally relevant to  $y$ . (exclusion)
2. For every physical event  $y$ , some physical event  $x$  is casually sufficient for  $y$ . (physical determinism)
3. For every physical event  $x$  and mental event  $x'$ ,  $x$  is distinct from  $x'$ . (dualism)
4. For every physical event  $y$ , no mental event  $x'$  is causally relevant to  $y$ . (conclusion)

This is bad for dualist theories, including ones incorporating supervenience. However, this argument is invalid.

Say that  $x$  is *causally connected* to  $z$  if and only if  $x$  causes  $z$  or visa versa. Then a fixed formulation of the same idea is:

- 1'. If an event  $x$  is causally sufficient for an event  $y$ , then no event  $x'$  distinct from and causally disconnected from  $x$  is causally relevant to  $y$ . (exclusion)
- 2'. For every  $I$ -level event  $y$ , some  $I$ -level event  $x$  is causally sufficient for  $y$ . ( $I$ -determinism)

But what about physical bio-determinism? It doesn't seem right because the first biological events must have been caused from some events from a non-biological level. So, this more general formulation 2' doesn't work, and we should restrict to the 0-level.

- 2'. For every 0-level event  $y$ , some 0-level event  $x$  is causally sufficient for  $y$ . (0-determinism)
- 3'. For every 0-level event  $x$  and every  $i$ -level event  $x'_i$ , where  $i > 0$ ,  $x$  is distinct from  $x'_i$ .  
(pluralism)

The it immediately follows that we can write

- 4'. For every 0-level event  $y$ , no  $i$ -level event  $x'_i$  that is causally disconnected from every 0-level event antecedent to  $y$ , where  $i > 0$ , is causally relevant to  $y$ . (new conclusion)

From this line of revised argumentation, we can pose a new definition of “emergent property”:  
A property is emergent if and only if it has novel causal powers from its object’s parts.