

The Case for Strong Emergence

S. Hossenfelder

*Frankfurt Institute for Advanced Studies
Ruth-Moufang-Str. 1, D-60438 Frankfurt am Main, Germany*

Abstract

As everyone knows, physicists have proved that free will doesn't exist. That's because we are made of tiny particles which follow strict laws, and human behavior is really just a consequence of these particles' laws. At least that's what I used to think. But some years ago I stumbled over a gap in this argument. In this essay I want to tell you what made me rethink and why you should rethink, too.

1 Reductionism Works

Large things are made of smaller things, and if you know what the small things do, you can tell what the large things do. Physicists call this idea reductionism. You might not like it, but it arguably works well. Reductionism allowed us to understand molecular bonds and chemical elements, atomic fission and fusion, the behavior of the atom's constituents, and these constituent's constituents – and who knows what physicists will come up with next.

It took some centuries, but thanks to reductionism physicists now have a remarkably simple description for our universe that explains almost everything we observe. According to this description, matter is made of 25 particles, collected in what is known as the standard model of particle physics. The 25 particles interact through four forces: the electromagnetic force, the strong and weak nuclear force, and gravity. And everything else – from chemistry to biology to cosmology – follows from that, at least in principle.

This currently best explanation for the world around us is almost certainly incomplete. There might, for example, be a few more particles to account for dark matter. Something's fishy with the cosmological constant. And no one understands how gravity works when space-time is strongly curved. But for the following argument these unresolved puzzles do not play a role because we will be concerned only with the structure of the theories we know already.

2 What is Fundamental?

This essay contest posed the question “What is fundamental?” but I don't find it insightful to ask for the meaning of a word. One could just answer such a question by writing down a

definition, and where's the fun in that? A somewhat more interesting approach to answer the question would be to instead explain how the word is commonly used. But answering such a question requires a mix of history, linguistics, and sociology, none of which I know much about, and none of which I suspect this contests' audience wants to know much about.

Let me therefore move on by just defining what I mean by “fundamental” and then using this definition to instead answer a different question, one we argue about much better, namely whether it is rational to believe that you have free will. I promise I will get to this before the essay is over, but first I must clarify how I refer to physical theories:

*A **physical theory** is a set of mathematically consistent axioms combined with an identification of some of the theory's mathematical structures with observables.*

*If two physical theories give the same predictions for all possible observables they are **physically equivalent**.*

That having been said, the definition of “fundamental” that I will use here is:

*A physical theory A is **more fundamental than** B if B can be derived from A, but not the other way round. In this case, the theory B is **weakly emergent from** A. A physical theory is **fundamental** (without qualifier) if it is to best current knowledge not emergent from any other theory.*

This definition I think captures how the word is used in the foundations of physics today, though I will admit to not having polled my colleagues, so I may be mistaken. In Figure 1, I have depicted an example of a directed graph of theories with oriented links between them indicating possible derivations.

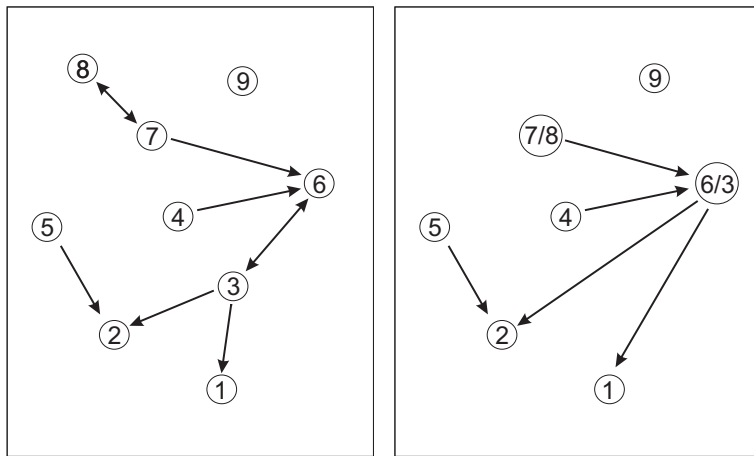


Figure 1: Left: Example of a graph of theories. Arrows indicate a known mathematical derivation. Right: Physically equivalent theories can be collected to one node.

Some comments on these definitions.

First, I am aware that other people have defined terms differently. For example what I call “weakly emergent” is sometimes referred to as “reducible,” and the word “emergent” doesn’t

seem to have any agreed upon definition (see eg [1, 2]). But please let us not quibble about the use of words. I have chosen these definitions because they will allow me to make my case sharply.

Second, note that according to the above what is fundamental depends on current knowledge. A theory considered fundamental today might be derived from another theory tomorrow, and would then cease to be fundamental. A theory that is emergent today, however, will remain emergent. (Leaving aside that a derivation might have been in error.) The standard model is, for all we currently know, fundamental. A good example for a weakly emergent theory is Fermi's theory of beta-decay, which can be derived from the standard model of particle physics but not the other way round.

Third, not in every pair of theories one must be derivable from the other. Some theories might not have any known connection to each other.

Fourth, several theories can be equally fundamental if they can mutually be derived from each other, in which case they are mathematically equivalent. A good example for this are duality relations, like those between the Thirring model and the sine-Gordon model [3]. But there is no particular reason why only two theories should be derivable from each other. In principle there could be infinitely many theories that start from different axioms and yet can be derived from one other.

Mathematically equivalent theories are also physically equivalent, though the opposite might not necessarily be the case: Two theories might give rise to all the same prediction without there being any (known) way to derive one from the other (which is the current situation for the AdS/CFT duality [6]).

Since we care only about the physics, we can collect physically equivalent theories to one node in the graph of theories (Figure 1, right). Note that this will remove loops only if they have an orientable component (plus possible further equivalences), so the graph doesn't have to be simple (though the depicted example is).

3 Weak vs Strong Emergence

Let us now add some empirical knowledge to the previous section's rather abstract discussion.

The first fact I want to draw upon is that our world can be described to good precision by a metric manifold in which matter occupies space. That the manifold is metric means we can measure distances and, with that, extensions.

Any experiment has an uncertainty on the measurement of distances, which I will refer to as the resolution of the experiment. For non-quantum ("classical") matter (say, a brick) this resolution can be identified with the actual extension of the matter. For matter with quantum properties (say, electrons) we can instead use the (center of mass) energy of the interaction that facilitates the measurement and define the resolution from the inverse of this energy.

As previously acknowledged, the description by ways of a manifold or quantum mechanics might break down on distances much shorter or much longer than we have tested. But this will not concern us in the following because for the present purposes we are interested only in what

happens in the range we have tested already.

We can then assign a resolution to every measurement and, since every physical theory allows the computation of measurement outcomes, we can assign a resolution to theories through the measurements which they (correctly) describe. This allows us to order the graph of theories as illustrated in Figure 2, left.

The second fact I want to draw upon is that nature does not allow mathematical inconsistencies. I consider this empirical knowledge because we have never witnessed a case in which we observed an inconsistency; indeed I am not even sure what this would mean. The consequence is that if we have two theories that are valid at the same resolution, they must be physically identical. This means that at any given resolution there can be only one (correct) physical theory, up to equivalence. This is illustrated in Figure 2, right

Of course this statement greatly oversimplifies the real situation because we often have theories at the same resolution but for different systems. Say, a theory for bricks and a theory for water both at a resolution of a micrometer. To picture this, you can imagine qualifiers for different systems as additional dimensions on the graph, which has the consequence that it is much rarer that two theories must be equivalent due to consistency. However, it is of little use trying to picture all these additional dimensions.

In the previous section I defined weakly emergent by the possibility of a mathematical derivation. As the dedicated reader will have anticipated, this is complemented by a notion of strong emergence which we can now define:

*A physical theory is **strongly emergent** if it is fundamental, but there exists at least one other fundamental theory at higher resolution.*

An example for this is theory nine in Figure 2. The rationale for this nomenclature is that loosely speaking going to lower resolution means going to larger extensions and hence larger objects. The existence of a strongly emergent physical theory then would mean that a large object could follow laws of nature which cannot be derived from any theory at higher resolution. Such laws, therefore, would be equally fundamental as the fundamental laws at high resolution that physicists are so proud of.

If a strongly emergent theory existed, it would imply that “more is different” as Anderson put it [4]. Your behavior, then, would not just be a consequence of the motion of the elementary particles that you are made of. It would mean that believing in free will would be compatible with particle physics. It would mean that reductionism is wrong.

(If you are bothered by the downward arrow in Figure 2, hang on, I’ll get to this in section 5.)

4 Strong Emergence Doesn’t Work

Most physicists are confident strong emergence doesn’t exist. The reason is not only that there isn’t any known example for it but that, more importantly, if there was an example it would

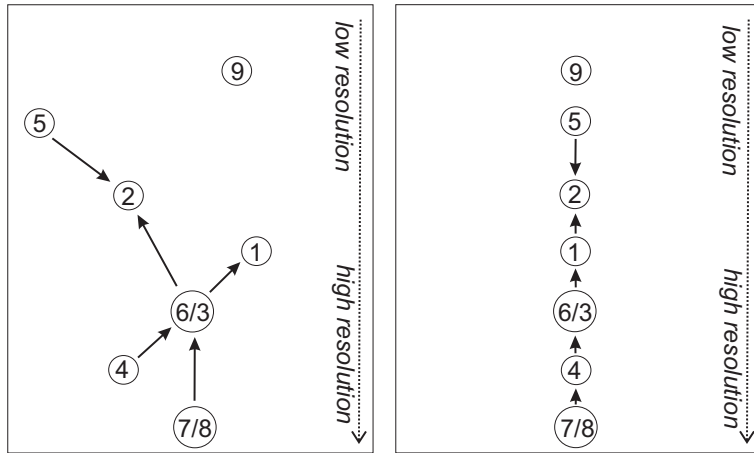


Figure 2: Left: Theories can be assigned a resolution at which they are valid. A theory’s range of validity is indicated by the size of the node and arrows extending from it. Right: Since two theories at the same resolution must agree on all predictions, the graph of theories becomes one-dimensional.

be incompatible with that they already know. And physicists know what they know with high confidence.

The argument – which I have made myself many times – goes like this. We know stuff is made of smaller stuff. We know this simply because it describes what we see. It’s extremely well-established empirical knowledge and rather idiotic to deny. No one has managed to cut open a frog and not find atoms.

Yes, It is interesting to ponder how it could have been any different, how it could possibly make sense that small stuff is made of larger stuff. The reason our universe doesn’t work this way is intricately linked with the ability of matter to occupy space and hence with space itself, something that – I admit – we don’t fully understand. Be that as it may, we have no working theory for building small things from large things. It doesn’t describe what we see. For all we know, stuff is made of smaller stuff.

This by itself, however, does not tell us what happens to the laws of the stuff. But for this, physicists have a mathematical framework called effective field theory; it tells us what happens with the laws if we join small things to large things.

It is worth emphasizing that effective field theories are a fairly recent development in the history of science. The idea has its roots in the 1950s, but key elements were only added in the 1990s (see eg [5] and references therein). It is still an active area of research, and I consider it origin of a paradigm shift that went largely unnoticed. I am emphasizing this because it means any discussion about emergence that predates or does not consider effective field theories is redundant.

Effective field theories are game changers because it used to be thought that theories which cease to work at high resolution (are “non-renormalizable”) are sick and cannot be correct as more fundamental theories. The modern way to think of them, in contrast, is that they may be approximations to the fundamental theory but that they must be completed. The paradigm change here is that a (correct) candidate for a fundamental theory might not reveal itself at first sight; indeed many theories which look wrong – because they break down at some resolution – are

compatible with an underlying theory that is perfectly healthy. They can thus be weakly emergent from a fundamental theory. The previously mentioned Fermi-theory of the weak interaction is such a case: It is non-renormalizable (“sick”) but can be completed by a renormalizable (“healthy”) theory.

Effective field theories work with quantum field theories, that is the type of theory that we presently use to describe nature at the highest resolution probed so far. The key equations of the framework (the “renormalization group equations”) connect a theory at high resolution with a theory at low resolution. That is, the theory at low resolution is always weakly emergent. It can be derived – at least in principle – from the theory at high resolution.

In practice the derivation of the low-resolution theory can only be done for simple systems, but from a philosophical standpoint this isn’t relevant. Relevant is merely that physicists *do* have equations that define the theory on low resolution from the theory at high resolution.

Effective field theories can fail [9] in the sense of methods becoming inapplicable, and there are certain theorems that can fail (such as the decoupling of scales), and there are some approximations that might become invalid (such as weak coupling), and so on. These are practical problems for sure. But in principle, none of this matters. Because even if we don’t know how to do a single calculation, the theory is still there. It doesn’t go away.

In principle, for example, we could use effective theories derived from the standard model plus general relativity to calculate, say, election outcomes. No one can do such a calculation, of course. And even if we could it’s questionable we could finish the calculation before we have the election results. But since there isn’t any reason why the known theories should stop working, we must conclude that indeed human behavior is weakly emergent from the underlying quantum field theory. In other words, you are nothing but a bag of particles, and science has proved it.

This is depicted in Figure 3, left, where “EFT” stands for the effective field theories derived from a (presently) fundamental theory. We can use the known mathematical tools it to obtain the theory at low resolution from the theory at higher resolution. As per the assumption that no logical contradictions are allowed and two theories that make the same predictions are physically equivalent, this means all other theories either agree with the predictions from effective field theory (and are hence weakly emergent) or they are wrong. And that’s why there is no strong emergence.

The previous argument is a sloppy version of the philosophically more elaborate “causal exclusion argument” [7, 8] which, roughly speaking, says that if a low-resolution effect can be derived from a theory at higher resolution, then the effect cannot have another cause.

The causal exclusion argument combined with effective field theory is the main reason why physicists believe that reductionism is correct. Another reason for their confidence is the absence of any known example of strong emergence, ie a case in which the properties of a system at large scales are known to be not calculable from the underlying theory. (Though there are certainly many examples in which they are not calculable by presently known methods.)

One example that is supposedly a case of strong emergence which I sometimes hear is superconductivity. But there is no reason to think superconductivity is strongly emergent. It’s

a novel feature that arises by the interaction of a system's constituents and by that it's entirely encoded in the system's microscopic properties already. No behavior has ever been observed that would imply superconductors are incompatible with the standard model. If that was so, you'd have seen the headlines.

It is true that we have to date no good theory for high temperature superconductivity, but the reason for this is that high temperature superconductors are believed to be strongly coupled, ie perturbative methods fail. This is one of the above mentioned cases in which calculations become intractable, but that doesn't mean the result of the calculation doesn't exist.¹

There are two examples in which the problem of calculating a property of a composite condensed matter system has been identified with the halting problem in computer science by using suitably configured (if somewhat contrived) systems [10, 11]. If the calculation of an emergent feature has an undecidable outcome, this would constitute a cases of strong emergence. However, both of these examples rely on infinitely large systems and/or the thermodynamic limit. The statement then comes down to saying that for an infinitely large system certain properties cannot be calculated on a classical computer in finite time, which is probably correct but doesn't teach us anything about reality.

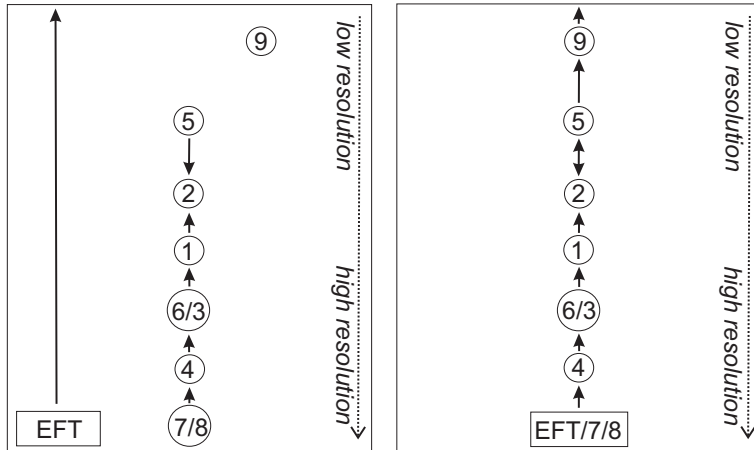


Figure 3: Left: Effective field theories derived from one fundamental theory (EFT) remain valid at all resolution. Right: As a consequence, all other known theories must be compatible with the already known theories derived by use of EFT.

5 Top Down Causation Doesn't Help

Top down causation is the idea that the laws of a system at low resolution can dictate the laws at high resolution. I have depicted this with theory five in Figures 2 and 3. Again we don't know any case in which this happens. But even if there was it wouldn't make strong emergence possible; it would merely mean that in at least some range of resolution the existing theories must be equally fundamental. The reason is, as previously, that (a) we already have a bottom-up

¹It may be possible to address this problem by using the the AdS/CFT correspondence which maps the strongly coupled condensed matter system to a weakly coupled gravitational system. So maybe we are not all that far from actually deriving a theory for high temperature superconductivity.

causation by way of effective field theory and (b) any other theory is either compatible with that or wrong.

I said above that we have no examples of top-down causation, but we certainly have wrong examples. Since these seem to be widespread, allow me some comments.

A typical argument goes like follows. The chief of CERN speaks the word “Go,” and in response someone pushes a button which will set into motion two proton beams that collide and produce a Higgs-boson. Human speech, as a perturbation of density fluctuations in the air, takes place at much lower energies (ie lower resolution) than elementary particle collisions. Hence, a top level process has caused a lower level process. Another example is that I swallow a pill, so that a big, low-resolution object like my arm causes a chain of molecular reactions. Reductionism must be wrong!

But such examples merely show that large systems often have interactions at a variety of energies at different places and at different times. Therefore, some parts of the system might lend themselves to a description at low energies (sound waves) while others do not (proton collisions). To demonstrate top-down causation, you would have to show that it is not possible to derive the sound wave’s propagation from the high-resolution theory for the air and its atoms and these atom’s constituents and so on. And there is no reason to think this isn’t possible, never mind that you won’t be able to actually do the calculation.

Another type of argument uses (possibly global) boundary conditions. Since the boundaries are usually large-scale (say, conducting plates) and yet constrain the behavior of the system at shorter scales (possibly large integer fraction of the plate’s distance), this is taken to mean a top-down causation took place. Again, however, to demonstrate top-down causation it would be necessary to show that the boundary conditions (the plates) could not themselves have been described at high resolution.

A related but somewhat different case are topological constraints. The equations of general relativity, for example do not determine the topology of space-time. But just because the equations do not determine some property of a system doesn’t mean that property cannot be determined from the system’s (entire!) small-scale configuration. A good way to see this is to think of a chain. Each link of the chain has two neighbors. If you look at any element of the chain and its neighbors (the “local” information) you cannot tell whether the chain is closed (ie, you cannot tell its topology). But of course if you have the complete information about the neighbor-couplings you will be able to tell that the chain is closed.

Yet another argument that seems different at first sight but is wrong for the same reason as the example with the chain is that entanglement realizes top-down causation [12]. The argument here is that entanglement is a non-local property of a system. Hence, if you have information only about a small part of a system, you have no way of knowing whether the system will begin to show novel effects due to entanglement if you look at the full system. Again, though, it is clearly possible to derive the behavior of the whole system if you have information about its entire microscopic constituents which, of course, includes entanglement between them.

In summary, we have no viable example of either strong emergence or top-down causation. Free will isn’t free. Effective field theory seems a fool-proof argument. So far.

6 The Loophole

Now that it's clear what's at stake, it doesn't take many words to state what's wrong with the previous argument. It's simply that we don't know for sure the equations of effective field theories (RGEs) have solutions which can be analytically continued from high resolution to all lower resolutions.

Landau poles are typical examples. A Landau pole is a divergence in a coupling constant that determines the strength of an interaction. Such a divergence happens, for example in QCD at around 100 MeV or in QED at energies far beyond the Planck energy. These poles are clearly non-physical and must mean that the extrapolation for the running of the coupling breaks down because the theories become strongly coupled. And QED of course is believed to be absorbed in a grand unified symmetry long before the Landau pole, which may or may not actually happen.²

So, a theory can't be extended beyond its Landau pole which would mean strong emergence is viable, but also Landau poles shouldn't be there to begin with because they are not physical. Landau poles, thus, don't help. But note that just because a function can't be continued doesn't necessarily mean it diverges and therefore can be discarded as non-physical. A function can be perfectly regular, indeed be differentiable up to all orders, and still can't be continued.

A good example for a non-divergent function that can't be continued is the function $f(x) := \exp(-1/x^2)$ for $x \geq 0$, which cannot be Taylor-expanded around zero and hence can't be continued to $x < 0$. If you haven't come across this function before, I encourage you to do the Taylor-expansion at zero. You will find it's just identical to zero at all orders.

Because of this you can complete the function $f(x)$ beyond zero with any other function that has a similar behavior, say, $g(x) := \alpha \exp(-1/x^2)$ for any value of α . The combination of both functions ($f(x)$ for $x \geq 0$ and $g(x)$ for $x < 0$) will then be well-defined and differentiable at all orders. And yet, you cannot continue the function from $x > 0$ to $x < 0$.

To translate the mathematical example to the physical case, $f(x)$ corresponds to some coupling constant of the effective theory, x correspond to the scale of resolution, and of course the transition would not be at zero, but should be shifted to some finite value, say a distance of a nanometer. But the central conclusion remains: There isn't a priori any reason why it must be possible to continue the constants of the theory at high resolution to any lower resolution. If you run into a point where the coupling can't be continued, you will need new initial values that have to be determined by measurement. Hence, strong emergence is viable.

I will admit that this example would be more convincing if I could come up with a system that has a beta-function which actually displays such a feature. I don't have any such example, and if I had I'd have written a proper paper and not an essay with many pictures and few equations. But I also do not know of any reason why it should not happen.

With this, the ball is back in the court of physicists. The argument that effective field theory proves reductionism even though no one is able to at least derive the properties of an atomic nucleus from QCD undeniably has an air of physicists' hubris to it. It is thus only fair on those philosophers who like to believe that strong emergence exists that physicists first show that

²It is actually the hypercharge coupling of the electroweak theory that diverges.

the coupling constants of a quantum field theory can always be continued to low energies for physically realistic systems.

7 Conclusion

In this essay, I have presented a new example for strong emergence. While this example is purely hypothetical, it illustrates how truly new fundamental laws could emerge for composite objects, at least theoretically.

I herewith grant you permission to believe in free will again.

References

- [1] Butterfield, J. “*Less is Different: Emergence and Reduction Reconciled*,” *Foundations of Physics*: 41, 6 (2011): 1065-1135.
- [2] Butterfield, J. “*Reduction, Emergence and Renormalization*,” *The Journal of Philosophy*, 111 (2014): 5-49.
- [3] Coleman, S. “*Quantum sine-Gordon equation as the massive Thirring model*,” *Phys. Rev. D* 11 (1975): 2088.
- [4] Anderson, P. W. “*More Is Different*,” *Science* 177, 4047 (1972): 393-396.
- [5] Burgess, C. P. “*Introduction to Effective Field Theory*,” *Ann. Rev. Nucl. Part. Sci.* **57**, 329 (2007) [hep-th/0701053].
- [6] Maldacena, J. M. “*The Large N limit of superconformal field theories and supergravity*,” *Int. J. Theor. Phys.* **38**, 1113 (1999) [*Adv. Theor. Math. Phys.* **2**, 231 (1998)] [hep-th/9711200].
- [7] Kim, J. “*Making sense of emergence*,” *Philosophical studies* 95.1 (1999): 3-36.
- [8] Kim, J. “*Emergence: Core ideas and issues*,” *Synthese* 151.3 (2006): 547-559.
- [9] Donoghue, J. F. “*When Effective Field Theories Fail*,” *PoS EFT* **09**, 001 (2009) [arXiv:0909.0021 [hep-ph]].
- [10] Gu, M., Weedbrook, C., Perales, A., Nielsen, M. A. “*More Really is Different*,” *Physica D* 238 (2009): 835-839 arXiv:0809.0151 [cond-mat.other].
- [11] Cubitt, T. S., Perez-Garcia, D., Wolf, M. M. “*Undecidability of the spectral gap*,” *Nature* 528 (2015): 207211.
- [12] Gambini, R., Pullin J., “*Event ontology in quantum mechanics and downward causation*,” *International Journal of Quantum Foundations* 2, 89 (2016) arXiv:1605.07531 [quant-ph].