Effective Theory

These notes[1] are part of a series concerning "Motifs in Physics" in which we highlight recurrent concepts, techniques, and ways of understanding in physics. In these notes we discuss the concept of effective theory and the related idea that no theory is truly fundamental.

Fundamental or Effective?

Once upon a time, when scientists would develop a particularly successful physical theory, one which displayed excellent agreement between theoretical predictions and experimental tests, they would subsequently conclude that said theory was a fundamental theory of the physical world. Here "fundamental" labels a theory which is the best possible theory one could ever have of the world. Said theory underlies all other theories and is therefore, in a sense, our most accurate description of reality. Thus, "fundamental" can be taken to be roughly synonymous with the philosophical notions of "absolute truth".

Although rarely articulated as such, this notion of a fundamental theory was the one ascribed to the classical physics developed during the 19th and 18th century. Newtonian mechanics had appeared to describe so many and such disparate phenomena that it was perfectly natural for physicists to believe the entire universe was Newtonian. And the theories of electrodynamics, although clearly associated with a novel set of phenomena, were largely consistent with the Newtonian picture of deterministic dynamics operating within a fixed background of space.

Certainly, it was fidelity to these edicts of classical physics which allowed scientists to develop the theories of statistical mechanics and electromagnetic waves to the extent that they did. However, it was an undue allegiance to the fundamental nature of these theories that led to so many incorrect predictions when physicists attempted to understand phenomena that extended far beyond the physical regimes they were used to. With the benefit of hindsight, we now know the questions which pointed to flaws in the implicit beliefs that these theories were fundamental, for if classical electrodynamics is correct on all length scales, why don't electrons orbiting nuclei radiate away all their energy? Or, if Newtonian gravitation is true to for all mass distributions, why does Mercury's orbit precess?

Nowadays, we know that the correct interpretation of classical physics—given its clear success in describing pulleys and small planets and its failure in describing electrons and some planetary orbits—is that it is an **effective theory**. An effective theory is a theory which is only true within certain parameter regimes or above certain length-scales, typically the regimes and length-scales used to experimentally verify the theory. Effective theory embodies the 20th century post-modern idea that absolute truth is not achievable, that our representations of truth have their limitations, and that whenever physicists or scientists come up with a theory of nature, they must have some level of humility and restraint in extrapolating the theory—and the premises about reality implicit in its construction—to other physical domains.

In this sense, all theories of nature are effective theories which is to say every theory is an approximation of some more fundamental theory about which we are largely ignorant. In the following sections we discuss one of the most important aspects of these effective theories: although one can develop an effective theory from a more fundamental theory, one cannot proceed in the opposite direction and develop a more fundamental theory from an effective theory (See Fig. 1). This fact is the reason for the historical precedent that developing more fundamental physics theories typically requires us to relinquish the principles which were foundational to the construction of their antecedents. We conclude by discussing how classical physics exists as an effective theory to various directions of modern physics.

[1] Inspired by a blog post by Philip Tanedo

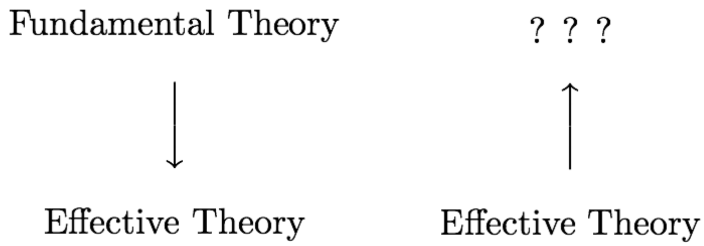

Figure 1: It is possible to unambiguously derive an effective theory from a more fundamental theory, but it is not possible to unambiguously derive a fundamental theory from an effective theory.

Effective Theory and Parameters

The basic features of effective theories can be illustrated with a simple example from classical mechanics. Say we have a spring that obeys Hooke's Law for small displacements from the origin. For such small displacements, the force exerted by the spring would be

$$F_0(x) = -kx, \tag{1}$$

where k is the spring constant. Now say that for larger displacements where $x \lesssim \ell_0$ for some length scale ℓ_0, we determine experimentally that the spring does not precisely obey Eq.(1). Instead, for this parameter regime, the spring has the force law

$$F(x) = -kx + k\lambda \frac{x^3}{3!\ell_0^2} + \cdots, \tag{2}$$

where $\cdots$ represents unknown higher-order terms which can contribute to the force, and λ is a dimensionless parameter. There are two features of note in this example: For this situation, any model developed from Eq.(1) would be an effective model of the more fundamental model developed using Eq.(2). In this sense Eq.(1) is the effective force of Eq.(2). Moreover, it would have been impossible to determine that the more fundamental force law is of the form Eq.(2), if we had never probed the experimental regime $x \sim \ell_0$. In this way, although we could derive Eq.(2) from Eq.(1) by taking $x \ll \ell_0$, it would not be possible to postulate the law Eq.(2) beginning from Eq.(1) presuming we had no additional experimental or theoretical information.

We can see this more clearly by asking a higher level question:

> If we took Eq.(2) to be an effective form for a more fundamental force law, what is that more fundamental force law?

We will argue that this question has no unambiguous answer. Given that Eq.(2) reduces to Eq.(1) when we take $x \ll \ell_0$, we can similarly expect that this hypothetical "more fundamental" force law reduces to Eq.(2) when we take its Taylor expansion in x/ℓ_0. So, rephrasing the question of what is the more fundamental force law of Eq.(2), we can ask what function (when Taylor expanded in x/ℓ_0) gives us Eq.(2)? With some thought it is clear that there are many function which yield Eq.(2) upon expansion, and it would be impossible to know which function was correct unless we had more information. Take the four functions listed below:

$$\begin{aligned} &-kx + k\lambda\frac{x^3}{3!\ell_0^2}, \qquad && -k\ell_0\sqrt{\frac{2}{\lambda}}\tanh\left[\frac{x}{\ell_0}\sqrt{\frac{\lambda}{2}}\right], \\ &-\frac{k\ell_0}{\sqrt{\lambda}}\sin\left[\frac{x}{\ell_0}\sqrt{\lambda}\right], \qquad && -\frac{3k\ell_0^2}{\lambda x}\ln\left[1 + \frac{\lambda}{3}\frac{x^2}{\ell_0^2}\right]. \end{aligned} \tag{3}$$

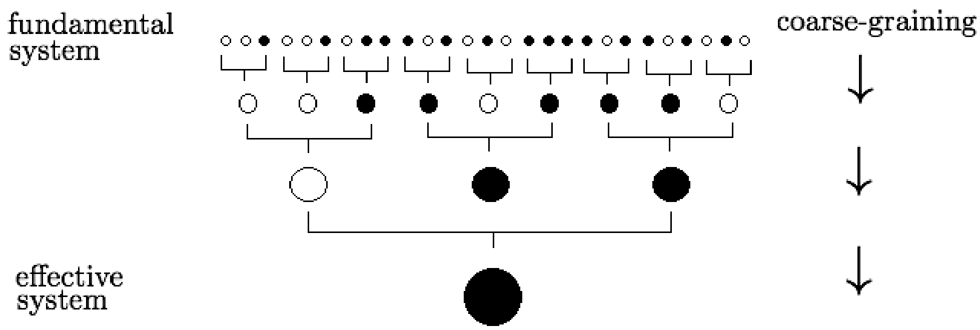

Figure 2: As we successively coarse grain our system, numerous degrees of freedom are consolidated into fewer degrees of freedom on longer length scales. This process of coarse graining takes us from a fundamental to an effective system, and although it is possible to uniquely determine the effective system from the fundamental system, it is not possible to uniquely determine the fundamental system from the effective system.

Each yields Eq.(2) when expanded to $\mathcal{O}(x^3/\ell_0^3)$, and thus each is a candidate for the more fundamental version of Eq.(2). To determine which is the correct fundamental force law, we would need some criteria to rule out a few of the candidates. One way would be to predict and test a theoretical result which depends on the $\mathcal{O}(x^5/\ell_0^5)$ contribution to $F(x)$. But even then, the force law would not be totally unambiguous because it could be corrected by higher order terms whose relevance we have yet to test. At the end the of the day, the most we would be able to say is that the force law we find is true to a certain order of x/ℓ_0 or to a certain decimal place precision.

In general, it is not possible to go from an effective theory to a more fundamental theory unless there is additional information to rule out the possible fundamental theories we can choose from. This additional information can come in various forms. It can come as experiments which eliminate unphysical possibilities, or as as theoretical principles (like symmetry principles) which constrain the space of theories we explore. However, such theoretical principles are often unreliable because it is not always evident that nature will remain faithful to them well beyond the physical regimes we have already explored.

Indeed, as is discussed in the next section, it is often not possible to know the precise physics which governs systems at length scales smaller than the ones to which we have access.

Effective Theory and Length Scale

In the previous section, we showed how the concept of effective theory is manifest when we consider the parameter-regime validity of equations in physics, but, historically, the concept of effective theory was first developed in reference to length scales. The basic idea of how effective theory is relevant to length scale is similar to the idea of how effective theory is relevant to parameters: More fundamental theories on smaller length scales uniquely determine their effective theory derivatives, but effective theories do not uniquely determine their more fundamental primaries.

We can understand this with a toy example. Say we have 27 circles in a row each of which can be filled or unfilled. If our visual resolution is such that we can see all 27 circles, then we could precisely specify the state of the system. If, however, our visual resolution was such that we could only see every set of three circles as one circle and that the color of this set was dominated by the majority color (that is two unfilled circles and one filled circle looked like an unfilled circle) then system we observe would be an approximation of the more precise underlying one. In this case, we would have a **coarse-grained** view of the system where instead of 27 individual circles we would see 9. Defining the resolution as the inverse of the length scale at which we can perceive differences in the system, the 9 units we perceive on this larger length scale and lower resolution would be an effective description of the 27 units which exist on a smaller length scale and higher resolution.

We can imagine performing such coarse-graining twice more so that instead of seeing 9 circles we see only one circle which represents the effective state of the entire 27 circle system. In such a scenario, this effective state would be to the full 27 circle state what the magnetic field per mole of a permanent magnet is to the $\sim 10^{26}$ magnetic dipoles which give rise to it.

What should be clear from this construction is that the effective state of the system at the lowest resolution (i.e., largest length scale) is uniquely determined by the state of the system at the highest resolution, but the converse is not true (See Fig. 2). This is because in coarse-graining the system we are essentially performing an all-or-nothing average over the filled and unfilled states of the circles, and as is true of all averages, the distribution uniquely determines the value of the average but the average does not uniquely determine the distribution.

This is the mathematical reason why effective theories defined on particular length scales cannot be used to determine the theories which exist at smaller length scales. In going from a larger length scale to a smaller one, there is an irretrievable loss of information from averaging over degrees of freedom. Moreover, it is often not possible to determine whether physics at smaller length scales is governed by the same set of laws that govern physics at larger length scales. This was the problem posed by quantum physics in the early 20th century and the problem posed by quantum gravity today.

> **Whole $>\sum$ (Parts) :** Although it is possible to go from the physical laws of a fundamental theory to the physical laws of an effective theory it is not always (or, rather, hardly ever) clear what properties result from having many degrees of freedom obeying those effective laws within the same system. This was the challenge posed by statistical mechanics in the 19th century (solved!), by superconductivity in the 1950s (solved!) and is currently the problem posed by the molecular organization of life (unsolved!).

The Dense, the Fast, and the Small

Now that we have discussed the basic ideas of effective theory, we can consider how effective theory informs our understanding of modern theories of physics. The classical physics of the 19th century and earlier serves as a collection of effective theories of the many more fundamental theories which came later. For our purposes we will discuss the theories which extend in three directions: gravitation (the really dense), relativity (the really fast), and quantum physics (the really small).

It is useful to consider how these various theories relate to one another and how they are themselves effective theories of more fundamental variants, both real and hypothetical. For example, we can associate each theory with a physical parameter and then represent the parameter on an axis such that the origin represents a classical mechanics without the theory present and moving away from the origin "turns on" said theory.

To define these axes in a non-ambiguous way, we would need to parameterize them by dimensionless quantities. Quantum mechanics is relevant when the action [2] S is of the same order of magnitude as $\hbar$, so we define the quantum mechanics axis by $\hbar/S$. Turning off quantum mechanics would amount to taking $\hbar \to 0$ or $\hbar \ll S$. Gravity becomes relevant when Newton's Gravitational constant leads to a force strong enough to change the energy of a particle, so we could define the gravity axis by GMm/ER where M is the mass of the gravitational object, m is the mass of the gravitating object, E is the energy of the gravitating object, and R is the radial extent of the system. We turn off gravity by taking $G \to 0$ or $E \gg GMm/R$. Relativity becomes relevant when the speed v of objects approaches the speed of light c.[3], so we could define the relativity axis by v/c, and we would turn off relativity by taking $c \to \infty$ or $v \ll c$.

[2]For spin systems we would replace the action with the average spin which, conveniently, is represented by the same letter.

[3]Alternatively, when E approaches pc or mc^2

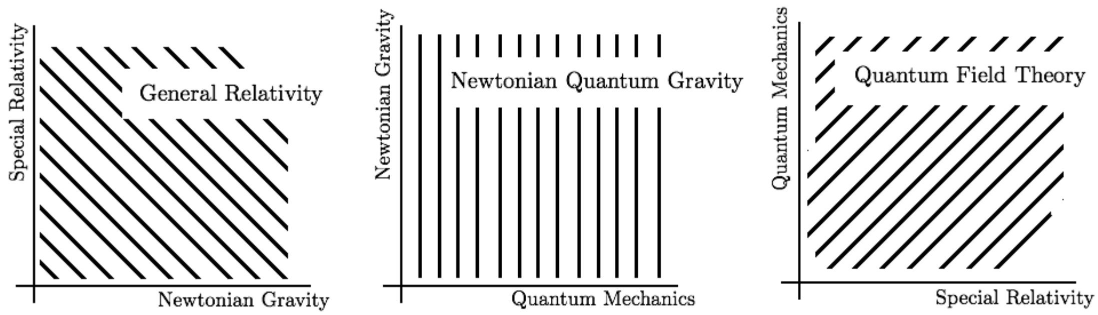

Figure 3: Various effective theories of modern physics and their even more effective limiting cases.

The plots in Fig. 3 depict how newtonian gravity, special relativity, and quantum mechanics are each effective theories of some more fundamental theory even though they are all, in a way, more fundamental than non-gravitational classical mechanics[4]. We can even go further and depict all of these axes on the same plot (called the ”Bronstein cube”) in which case the most fundamental theory is the as-of-yet-undiscovered theory of relativistic quantum gravity[5]. Such a theory would proverbially contain general relativity and the Standard model of particle interactions as special limiting cases and is thus often termed a unified theory of fundamental interactions [6]. But as the previous discussions should illustrate, it is not at all easy to reason our way to such a theory given our knowledge of its limiting cases. Instead, developing such a theory would require stepping outside of the frameworks of the limiting cases themselves and postulating new frameworks (which can hopefully be experimentally probed) to encompass the length scales of interest.

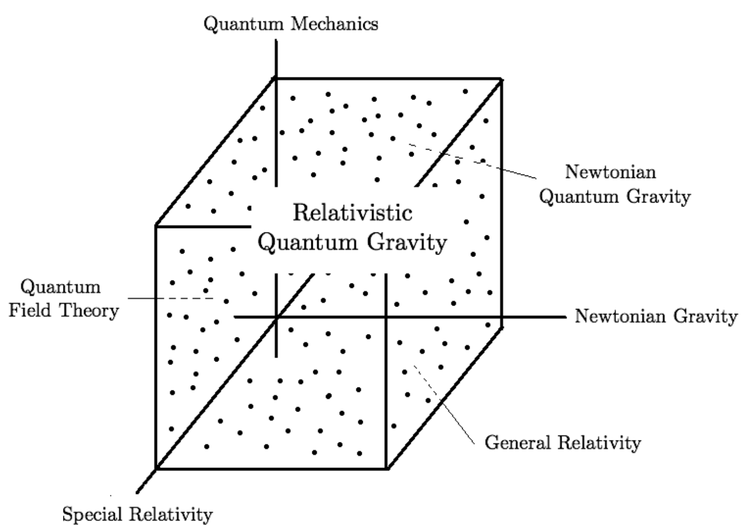

Figure 4: Bronstein cube

[4]We should note that non-gravitational mechanics is typically not seen as an effective theory of Newtonian mechanics. We include it as such here because such an interpretation leads to the ”Bronstein cube”.

[5]You will never hear someone say ”relativistic” quantum gravity because today ”gravity” is synonymous with Einsteinian gravity which already includes relativity.

[6]This is not a perfect picture of either modern physics or attempts to understand quantum gravity. See ”The cube of physical theories” [2] and the associated comments for a criticism of this cube.

Renormalization and Effective Theory

The concept of effective theory was first given a precise formulation in the context of **effective field theory**, a quantum field theory specification of effective theory developed by Kenneth Wilson in his work on renormalization [1].

Prior to Wilson's work, renormalization was a questionable procedure which involved isolating and then ignoring the infinite quantities in quantum field theory. Wilson made the procedure precise and justifiable by relating it to the fact that all physical theories are developed within some frame of ignorance. Wilson's claim was that the divergent results incurred from summing up small length-scale contributions to physical quantities were beside the point because physicists should not be including these contributions anyway. Given that quantum field theories are developed and tested within a specific length-scale regime, it is unrigorous to extrapolate the theories to much smaller scales. More generally, Wilson concluded that *any* physical theory we develop comes with an implicit length-scale below which we cannot claim to know much of anything.

Although Wilson gave a clear mathematical exposition of effective theory and why the idea was relevant to our most fundamental theories of nature, the idea applies more generally to all physical theories. Key to the idea is that the theories we think are fundamental are actually just effective theories of even more fundamental theories to which we do not yet have experimental access.

References

[1] K. G. Wilson, "The renormalization group and critical phenomena," *Nobelprize.org*, Decenber 1982.

[2] S. hossenfelder, "The cube of physical theories," May 2011.