CHAPTER 3

WHY PHYSICS IS THE EASIEST SCIENCE ⸺ EFFECTIVE THEORIES


If we had to understand the whole physical universe at once in order to understand any part of it, we would have made little progress.  Suppose that the properties of atoms depended on their history, or whether they were in stars or people or labs ⸺ we might not understand them yet.  In many areas of biology and ecology and other fields systems are influenced by many factors, and of course the behavior of people is dominated by our interactions with others.  In these areas progress comes more slowly.  The physical world, on the other hand, can be studied in segments that hardly affect one another, as we will emphasize in this chapter.  If our goal is learning how things work, a segmented approach is very fruitful.  That is one of several reasons why physics began earlier in history than other sciences, and has made considerable progress ⸺ it really is the easiest science.  (Other reasons in addition to the focus of this chapter include the relative ease with which experiments can change one quantity at a time, holding others fixed; the relative ease with which experiments can be repeated and improved when the implications are unclear; and the high likelihood that results are described by simple mathematics, allowing testable predictions to be deduced.)

Once we understand each segment we can connect several of them and unify our understanding, a unification based on real understanding of how the parts of the world behave rather than on philosophical speculation.  The history of physics could be written as a process of tackling the separate areas once the technology and available understanding allow them to be studied, followed by the continual unification of segments into a larger whole.  Today one can argue that in physics we are finally working at the boundaries of this process, where research focuses on unifying all of the interactions and particles ⸺ this is an exciting time intellectually.  In recent years physicists have understood this approach better, and made it more explicit and formal.  The jargon for the modern way of thinking of theories and their relations is the method of "effective theories".  We'll use this bit of terminology in this chapter, and occasionally later in the book.  Sometimes this approach has been called a "reductionist" one.  But the word "reductionist" has different meanings and implications for different people, so we won't use it.  For physicists, "reductionist" implies simultaneously separating areas to study them, *and* integrating them as they become understood. This chapter aims to help understand where the Higgs boson discovery and supersymmetry fit into a broader picture, and more generally, to understand better how scientific research works.

ORGANIZING EFFECTIVE THEORIES BY DISTANCE SCALES

Probably the best way to organize effective theories is in terms of the typical size of structures studied by a particular effective theory, which we speak of as the distance scale of normal phenomena described by that theory. Imagine starting by thinking about the universe at very large distances, so large that our sun and all stars look like small objects from such distances. This is the effective theory of cosmology, where stars cluster in galaxies because of their gravitational attraction, and galaxies are attracted to each other and form clusters of galaxies. Because of gravitational attraction everything is moving, on a background of the expanding universe. The only force that matters is gravity. We can use simple Newtonian rules to describe motion — deviations due to effects described by quantum theory are tiny and can be ignored. We can study how stars and planets and galaxies form, their typical sizes, how they distribute themselves around the universe and so on. It doesn't matter if the particles that make up stars and planets are made of quarks or not, or how many forces there are at the small distances inside a nucleus. The large scale universe is insensitive to what its contents are except for their mass and energy. Because of this indifference, cosmology can make progress regardless of whether we understand how stars work, or whether protons are made of quarks, and so on. We can learn from astronomy data that there is dark matter. However, this also implies that if the dark matter is composed of particles we cannot learn from astronomy or cosmology what kind of particles they are, because cosmology is largely insensitive to the properties that distinguish one particle from another, such as their masses and what charges they carry.

Next consider smaller distances, about the size of stars. We can study how stars form, how they get their energy supply, how long they will shine — we can work out an effective theory of stars. While doing that we can ignore whether they are in galaxies, or whether there are top quarks or people. Let's go to a smaller distance, say people size, and consider the physics. Gravity keeps us on the planet, but otherwise it is the electromagnetic force that matters. All of our senses come from mechanical and chemical effects based on the electromagnetic force — sight consists of photons interacting with electrons in our eyes, followed by electrical signals traveling to our brains. Touch begins with pressure affecting cells in the skin, leading to electrical signals propagating to the brain. Hearing starts with air molecules hitting molecules in the eardrum, interacting via electromagnetic forces. Friction, essential for us to stay in place or to move, is due to electromagnetic forces between atoms. We don't need to know about the weak or strong forces, or the galaxy, to study people-sized physics. The sun inputs energy to the earth that provides all of our food and essentially all of our energy, mostly from stored solar energy, but otherwise how the sun works does not matter. Here is a case where phenomena from

one effective theory provide input to another in a very specific way — the earth can be viewed as a closed system except for the input of solar energy. From the point of view of the effective theory of the earth, how the sun generates its energy is irrelevant.

Now consider atomic size. Here we'll be able to see even better how powerful the effective theory idea is. To be able to use the basic equations that govern atoms we have to input some information, essentially a few properties of the electron (its mass and electric charge and spin), and the same properties for each of the naturally occurring nuclei if we want to describe the whole Periodic Table of the Elements. A description of atoms doesn't need to take account of whether stars or galaxies or people exist, nor if the nucleus is made of protons and neutrons, nor if protons are made of quarks.

Before we go to even smaller distances, this is a good place to stop and consider some of the implications of this way of thinking. When we have a tentative theory of atoms, we want to test its predictions and whether it explains phenomena we already know. The predictions for the lifetimes of excited states of atoms, for the energies of photons emitted by atoms, for the sizes of atoms, and so on, depend on pieces of input information — on the masses and charges and spins of the electron and the nuclei. Every effective theory has some input parameters such as these. Without input information about the electron and nuclei, we could solve the equations but not evaluate the results numerically, so we could not test the theory or make any useful predictions, or know if we had the right theory. For example, the radius of the hydrogen atom is $h^2/e^2 m_e$ (h is Planck's constant, $m_e$ is the electron's mass, and e is the magnitude of the electric charge of the electron). If we measure the size of the hydrogen atom we still can't check whether the theory is working unless we know (by measurement or calculation) the mass and charge of the electron and Planck's constant (all three were measured nearly a century ago). Note that the size of the atom gets larger if the electron mass is smaller (since it is in the denominator in the expression for the radius above), and the radius would become infinite if the electron mass were zero, as it would be without its interaction with the Higgs field.

Now we can emphasize an important aspect of effective theories. When we study the effective theory of the nucleus we want to be able to calculate the masses and spins and charges of the nuclei — those parameters are results derived in the effective theory at that level. But for the next higher level, the atom, those parameters are the input. They can be input if they are measured, whether or not they are understood in the effective theory of nuclei. Similarly, the electron is a fundamental particle whose mass and electric charge will, it is hoped, be calculable someday in a fundamental theory such as M or string theory, but for the effective theory of atomic physics it does not matter whether they are calculable or understood if they have been measured. We have known the numerical value of the mass

of the electron since the beginning of the 20<sup>th</sup> century, but we do not yet understand why it has that value.  The value of the electron's mass can be input into every effective theory that depends on the electron.  Every effective theory so far has some input that is for that effective theory a "given", not something to be questioned, such as the mass of the electron for the effective theory of atoms.

*If a theory has inputs, it is an effective theory*.  From this point of view the goal of particle physics is to learn the ultimate theory at the smallest distances, recognizing it as the theory for which no parameters have to be input to calculate its predictions.  For the ultimate theory it is not satisfactory to input the electron mass; rather, it is necessary to be able to calculate it from basic principles, and to explain why it has the value it does.

Every effective theory is based on others — it is effective theories all the way down, until the ultimate one, which we will name the final theory.  Each effective theory has certain structures, that bind together at its level — stars, atoms, nuclei, protons.  Stars are made of nuclei and electrons bound by the gravitational force as viewed from the effective theory of stars, but for the effective theory of cosmology they are just inputs characterized by a mass and brightness.  Nuclei are bound states of neutrons and protons for the effective theory of nuclei, but for the effective theory of atoms they are merely pointlike inputs.  All systems and structure are inputs at one level of effective theories, but something to be derived and explained by the effective theory at a smaller distance.  In a sense a given effective theory can be explained in terms of shorter distance theories, and the input from them.  Dirac said that his equation that unified special relativity and quantum theory for the interactions of electrons and nuclei explained all of chemistry, and in a sense he was right.  His equation, plus the input parameters describing electrons and nuclei, in principle explained all chemical processes.  In another sense he was not right, because in practice one could never start from the Dirac equation and calculate the properties of molecules, or figure out their structure, or how to construct new molecules with certain desired properties — the questions are just too complicated to solve.   For example, Dirac could not have deduced that water is wet from his equation.  For each effective theory new regularities or laws are found, and properties arise that are not predictable *in practice*.  They are often called emergent properties.  Life is an emergent property.  Physics tells us everything that molecules can do and cannot do.  In particular, physics can tell us that life will not emerge on some planets if circumstances are too adverse, but it cannot guarantee that life will emerge on a planet where conditions are favorable, even though that may be very probable.

Another way to view effective theories is in terms of types of understanding.  At its own level, an effective theory provides a "how understanding", a description of how things work.  But for the effective

theory above it, at larger distances, the smaller distance effective theory explains all or some of the input parameters, thus providing a why understanding. For example, nuclear physics describes the properties of nuclei, using the proton and its electric charge, spin, mass, and magnetic properties as given, unexplained input. But the Standard Model provides the explanation, allowing the calculation of all those properties of the proton in terms of quarks bound by gluons.

Yet another perspective appears if we observe that in general all areas of science are intrinsically open-ended — chemistry, the physics of materials, geology, biology, and so on. There is no end to the number of possible systems and variations that can be studied. But particle physics and cosmology are different. If the fundamental laws that govern the universe are found and understood, and the inputs are calculated in that theory, that's it — these two fields (that are merging into one) will end.

Having examined some implications of effective theories, let's return to the progression to smaller distances. We can go from atoms through the effective theory of nuclei, protons and neutrons, to quarks and leptons. Another important point is that each effective theory works well at its level, but it breaks down as we go to smaller distances and find new kinds of structure. When we went inside protons we found quarks, so we could not make a theory of protons unless we understood quarks and their interactions as well.

To go deeper into matter it will help to keep track of the distance scales numerically. Since we will cover a huge range of distances we need to use powers of ten — remember that each step in the power is a factor of ten in the result — $10^{-1}$\$ is a dime, and $10^{-2}$\$ is a penny. $10^3$ \$ is one thousand dollars. There are two scales that are useful for us to keep track of, meters that are a typical human size, and another length called the Planck length, after Max Planck who first introduced it soon after he took the initial step toward the quantum theory in 1899. The Planck length is extremely small — we'll understand it better later in this chapter. People are typically a meter or two in size. All of our analysis will be very approximate, so we won't worry about whether we talk of the height or width or radius of a system. We'll keep track of powers of ten, but not worry about distinguishing between things that might differ by a factor of two or so in size. People are about $10^{35}$ Planck lengths (1 followed by 35 zeros) tall, 1 or 2 meters. Atoms are about $10^{-10}$ meters (one ten-billionth of a meter) – about $10^{25}$ Planck lengths – in radius. Protons are about 100,000 times smaller than atoms, $10^{-15}$ meters or $10^{20}$ Planck lengths. Many particle physicists currently expect that quarks, leptons, photons, W and Z, and gluons will ultimately be understood as having a string-like extension if we could view them at a distance

scale of about 1 Planck length or $10^{-35}$ meters (a decimal point followed by 34 zeros and a 1) ⸺ they should seem point-like until we can study them at that scale.

In the language of this chapter we can think of the Standard Model (the previous chapter) as the effective theory of quarks and leptons interacting on a scale of about $10^{-17}$ meters, or $10^{18}$ Planck lengths, about 100 times smaller than protons and neutrons. Sometimes we call this the "collider scale" since it is associated with the typical energies at which the experimental collider facilities operate.

The goal of particle physics is an ultimate theory of the natural world. What should we call it? People have called it a Theory of Everything, but that name is somewhat misleading since it is really not a theory of weather, stars, psychology, and everything at once. Steven Weinberg called it the final theory. That's a good name, but can be misinterpreted as the last in a succession of theories that replaced each other, as if all the theories on the way to the final one should be discarded. In fact all the effective theories coexist simultaneously, all are part of our description of nature. A good name would be the "primary theory", following Lucretius, suggesting the theory one arrives at after going through a sequence of effective theories at smaller and smaller distances. On balance, sticking with "final" might be best for communicating with people easily, so I will do that. As we will see more clearly in a few paragraphs, the final theory should be the description of nature at a distance scale of about 1 Planck length, or about $10^{-35}$ meters. How can we journey the many orders of magnitude from the Standard Model to the Planck length?

SUPERSYMMETRY IS AN EFFECTIVE THEORY TOO

If nature is indeed supersymmetric one of the wonderful bonuses we may get is a way to carry out the journey through those orders of magnitude. Supersymmetry is an effective theory too, but it may be the penultimate one that will take us from the Standard Model to the theory near or at the Planck scale. Supersymmetry is an effective theory because it still needs some input parameters to describe the masses and interactions of the particles ⸺ those inputs should be predictable by the theory near the Planck scale, e.g M/string theory. Qualitatively, the supersymmetric Standard Model should become the effective theory at distances of $10^{-17}$ to $10^{-18}$ meters, and remain the effective theory down to nearly the Planck scale. It has special properties that allow it to cover that large range, rather than breaking down at shorter distances as most effective theories do.

In the past we have been able to do the experiments that were essential to make progress as the technology developed and allowed us to probe more deeply. The Planck length is too small ⸺ there will never be direct experiments possible at that scale. This statement is not just an extrapolation based

on current technologies or costs.  It is not just a matter of getting higher energy probes — the probes have to have the energy concentrated into a region smaller than the scale of interest, and before we can do that we run not only into limits like cost, we run into natural limits.  Nevertheless, there are a variety of ways to test ideas about Planck scale physics (this is explained further in Chapter 9). Saying we cannot test Planck scale physics is like saying we cannot test the big bang theory.  Even though there was no one present at the big bang, there are relics that provide convincing tests that it occurred.  There are already some indirect ways to test ideas about physics at the Planck length, but supersymmetry will allow us to add many systematic tests.   It will give us techniques to take a prediction at the Planck length and calculate what is predicted at the distances colliders can probe in the coming years (about $10^{-18}$ meters), or to take data from colliders and calculate the form of the theory at the Planck length implied by the data.  With supersymmetry we will be able to test ideas about M/string theories (Chapter 9) or whatever form the final theory may take — without it we do not know how to do that.  Of course, that bonus does not guarantee that nature is indeed supersymmetric, but it is a powerful motivation to study the theory and do the experiments needed to find out.


THE PHYSICS OF THE PLANCK SCALE

Whenever we describe a segment of nature we have to talk about the actual quantities that are calculated or predicted or explained in units — meters, or seconds, or kilograms or other appropriate units.  In Carl Sagan's novel *Contact* a signal from an extraterrestrial intelligence has been detected, with instructions on how to build a machine to facilitate communication.  There is a conversation between a scientist and an administrator:

" 'Don't ask why we need two tons of erbium.  Nobody has the faintest idea.'

'We wasn't going to ask that.  We want to know how they told you how much a ton is.'

'They counted it out for us in Planck masses.  A Planck mass is — '

'Never mind, never mind.  It's something that physicists all over the universe know about, right? And we've never heard of it.'"


For every effective theory there is a natural system of units, one where the description of phenomena is simple and not clumsy.  It would be silly to measure room sizes in Planck lengths just because the final theory is best talked about in those units.  Consider the units for atoms more closely. The radius of an atom can be expressed in terms of the properties of the electron plus Planck's constant h that sets the scale of all quanta.  h is the fundamental, universal constant of quantum theory.

Denoting the electric charge of the electron by e, and the mass of the electron by $m_e$, the radius (R) of the hydrogen atom, the simplest atom, is $R=h^2/e^2m_e$. The size of the atom is fully determined by these inputs. Nothing else matters — the nucleus, for example, is just a tiny object at the center. Once we know R, we can express the sizes of all atoms in terms of R; we don't need to use the input of h or the electron properties any more. R is the natural size unit for atoms. Atoms with different numbers of electrons will have somewhat different sizes, with radii such as 1.2R, or 2.4R, but all will be some number that is not too big or small times R. R is expressed in terms of parameters that are givens for atomic physics — we hope e and $m_e$ can be calculated someday in M/string theory, but they cannot be understood by atomic physics.

We can learn a great deal from this kind of analysis. For example, this expression for the size of an atom has major implications. It tells us that the size of atoms is essentially a universal quantity. Given the basic quantities (Planck's constant and the mass and charge of the electron), the size of all atoms of all kinds, anywhere in the universe, is determined. Since mountains and plants and animals are all made of atoms, their sizes are approximately determined by the size of atoms and the electromagnetic and gravitational forces. Combining atoms into genes and cells to evolve an organism that can manipulate and deal with the world requires a large number of cells, and sets a minimum size for the organism. Having a brain with enough neurons to make enough connections to make decisions about the world requires a minimum size brain, since the atoms cannot be made smaller than the size determined by the radius R. Nothing the size of a butterfly will be able to think, anywhere in the universe. Thinking organisms could be much larger than people, but they do not need to be, so it follows from general principles that all intelligent life is expected to be about our size, not much larger or much smaller.

Suppose now that we have just discovered the final theory. To present the results we have to express the predictions and explanations in appropriate units. What units should we use? We expect the natural units for the final theory to be very universal ones, not dependent on whether the universe has people or stars. There is only one known way to make universal units. There are only three universal constants in nature common to all aspects of nature, all interactions and all particles. They are Planck's constant h, the speed of light (denoted by c) that is constant under all conditions, and Newton's constant G that measures the strength of the gravitational force. Since Einstein proved that energy and mass are convertible into one another, and gravitation is a force proportional to the amount of energy a system has, everything in the universe feels the gravitational force. In fact, using these three quantities h, c, G it is possible to construct combinations that have the units of length, time, and energy. We

expect all the quantities that enter into the final theory or are solutions of the equations of the final theory to be expressible in terms of the units constructed from h, c, G. (For the interested reader, the result for the Planck length is $(Gh/c^3)^{1/2}$, which is about equal to $10^{-35}$ meters as we said before. For completeness, the Planck time is $(hG/c^5)^{1/2}$ which is about $10^{-44}$ seconds, and the Planck mass is $(hc/G)^{1/2}$ which is about $10^{-8}$ kilograms.) The Planck distance and time are extremely small, while the Planck mass (or equivalently energy) is very large for a particle.

Max Planck understood fully a century ago the universality of those units we call the Planck units. He wrote in his book *The Theory of Heat Radiation* (reprinted by Dover Publications, 1991) "All the systems of units which have hitherto been employed….owe their origin to the coincidence of accidental circumstances, inasmuch as the choice of the units lying at the base of every system has been made, not according to general points of view which would necessarily retain their importance for all places and all times, but essentially with reference to the special needs of our terrestrial civilization. Thus the units of length and time were derived from the present dimensions and motion of our planet….In contrast with this it might be of interest to note that….we have the means of establishing units of length, mass, time…which are independent of special bodies or substances, which necessarily retain their significance for all times and for all environments, terrestrial and human or otherwise, and which may, therefore, be described as "natural units". The means of determining the units of length, mass, and time …are given by the constant h, together with the magnitude of the velocity of propagation of light in a vacuum, c, and that of the constant of gravitation, G. These quantities retain their natural significance as long as the law of gravitation and that of the propagation of light in a vacuum [and quantum theory] remain valid. They therefore must be found always the same, when measured by the most widely differing intelligences according to the most widely differing methods." (We have left out some words to make this read smoothly, replacing them with ellipses, and since Planck didn't then know about the completion of the development of the quantum theory we have added that term in brackets as he would presumably have included it.)

The Planck length and time can also be interpreted as the smallest length and time that we can make sense of in a world described by quantum theory and having a universal gravitational force. The arguments that teach us that are interesting and not too complicated, but to explain them we have to recall the definition of a black hole. Basically the idea of a black hole is simple. Imagine being on a planet and launching a rocket. If you give the rocket enough speed it can escape the gravitational attraction of the planet and travel into outer space. If you increase the mass of the planet you have to increase the speed needed to escape. If you increase the mass so much that the required speed

exceeds the speed of light then the rocket can't escape since nothing can go faster than light.  The rocket and everything is trapped.  Light also feels gravitational forces, so beams of light are trapped too.  Since gravitational forces increase with decreasing distance, if you pack some mass into a sphere of smaller radius it is harder to escape from it, so the condition for having a black hole depends on both the amount of mass and the size of the sphere you pack the mass into.

A fascinating thing is that if we put an object having the Planck energy in a region with a radius of the Planck length we satisfy the conditions to have a black hole!  We cannot separate such a region into parts, or get information out from a measurement, so we cannot define space to a greater precision than the Planck length!  Since distance is speed x time, and speed can be at most the speed of light, and there is a minimum distance we can define, there is also a minimum time we can define — that comes out to be the Planck time.  We saw above that the Planck scale provides the natural units for expressing the final theory when the units are constructed from the fundamental constants h, c, and G.  Now we see a second reason for expecting the Planck scale to be the distance scale for the final theory — there does not appear to be a way even in principle to make sense of smaller distances or times.  The times when events occur cannot be specified, or even put in order, more precisely than the Planck time.

There is a third interesting argument that gives the same answer.  The gravitational force between two objects is proportional to their energies, and grows as the distance between them decreases.  Consider for example two protons.  Normally the repulsive electrical force between them is much larger than the attractive gravitational force.  But if the energies of the protons are increased to the Planck energy, then the gravitational force between them becomes about equal to the electrical force between them.  All the forces become about the same strength at the Planck scale, rather than being widely different in strength as they are in our everyday world.  Thus we might expect the gravitational force to unify with the others at the Planck scale, just as one might hope for in the final theory.

The arguments of this chapter have led in several ways to the idea that it makes sense to analyze the physical world with effective theories organized by the distance scale to which they apply, and move toward a final theory that unifies the forces and particles and is valid at the smallest scale that makes sense, the Planck scale.  Of course these arguments do not prove that is how nature works; we will not know that until we achieve such a description.  The Planck scale is very small, but not beyond our imagination.  Some readers may recall the delightful "Powers of Ten" book and movie of the designers Charles and Ray Eames, in which the universe was looked at in snapshots each ten times smaller than the previous one, starting with the largest cosmological distances.  When the Eameses did

this work, shortly before the Standard Model was discovered, they could not meaningfully go to smaller distances than the proton.  Today the Standard Model takes us nearly three powers of ten smaller than the proton.  From the universe down to the Standard Model domain is about 46 powers of ten, and from the Standard Model to the Planck scale only about 16 more powers of ten — looked at that way perhaps it does not seem so far.

THE HUMAN SCALES

Since the time of Copernicus, who taught us that the earth was not at the center of the universe, we have learned that if we want to understand the world we have to go beyond how the world seems to be and ask for evidence of how it is. It looks like the sun rises, but actually the earth orbits the sun.   We have learned that matter in the heavens and matter on earth obey the same natural laws, that we are made of the same atoms as the earth and the stars, that we and all organisms on earth evolved from cells, that we have unconscious minds which affect our behavior, that our star is only one of a hundred billion stars in our galaxy.  We have learned that the rules that govern nature (quantum theory and special relativity) are not apparent in our everyday classical world, and that the laws of nature have symmetries that are hidden from us but important (such as the particle-interchange symmetry of the Standard Model).  Perhaps even the number of space dimensions of the world will be larger than the three of which we are aware.  That may seem surprising, but so is the fact that the earth orbits the sun.  To understand the universe we must recognize that additional hidden aspects of nature may arise at scales far different from the human scale, and learn how to uncover them.  Supersymmetry is such a hidden aspect of nature.