CHAPTER 4

Much Ado About Nothing

Less is more.

—Ludwig Mies van der Rohe,
after Robert Browning

One step forward, two steps back, or so it seemed in our search for understanding our universe and accurately giving it a face. Even though observations had finally definitively determined the curvature of our universe—and in the process validated long-held theoretical suspicions—suddenly, even though it was known that ten times as much matter exists in the universe as could be accounted for by protons and neutrons, even that massive amount of dark matter, comprising 30 percent of what was required to produce a flat universe, was nowhere near sufficient to account for all the energy in the universe. The direct measurement of the geometry of the universe and the consequent discovery that the universe is indeed flat meant that 70 percent of the energy of the universe was still missing, neither in nor around galaxies or even clusters of galaxies!

Things were not quite as shocking as I have made them out to be. Even before these measurements of the curvature of the universe, and the determination of the total clustered mass within it (as described in chapter 2), there were signs that the by-then conventional theoretical picture of our universe—with sufficient dark matter (three times as much as we now know exists, in fact) to be spatially flat—was just not consistent with observations.

Indeed, as early as 1995, I wrote a heretical paper with a colleague of mine, Michael Turner, from the University of Chicago, suggesting that this conventional picture couldn't be correct, and in fact the only possibility that appeared consistent with both a flat universe (our theoretical preference at the time) and observations of the clustering of galaxies and their internal dynamics was a universe that was far more bizarre and that hearkened back to a crazy theoretical idea Albert Einstein had in 1917 to solve the apparent contradiction between the predictions of his theory and the static universe he thought we lived in and which he later abandoned.

As I recall, our motivation at the time was more to show that something was wrong with the prevailing wisdom than it was to suggest a definitive solution to the problem. The proposal seemed too crazy to really believe, so I don't think anyone was more surprised than we were when it turned out, three years later, that our heretical suggestion was precisely on the money after all!

Let's return to 1917. Recall that Einstein had developed general relativity and had heart palpitations of joy over discovering that he could explain the precession of the perihelion of Mercury, even as he had to confront that fact that his theory couldn't explain the static universe in which he thought he was living.

Had he had greater courage of his convictions, he might have *predicted* that the universe couldn't be static. But he didn't. Instead, he realized that he could make a small change in his theory, one that was completely consistent with the mathematical arguments that had led him to develop general relativity in the first place, and one that looked like it might allow a static universe.

While the details are complex, the general structure of Einstein's equations in general relativity is relatively straightforward. The left-hand side of the equations describes the curvature of the universe, and with it, the strength of the gravitational forces acting on matter and radiation. These are determined by the quantity on the right-hand side of the equation, which reflects the total density of all kinds of energy and matter within the universe.

Einstein realized that adding a small extra constant term to the left-hand side of the equation would represent a small extra constant *repulsive* force throughout all of space in addition to the standard gravitational attraction between distant objects that falls off as the distance between them increases. If it were small enough, this extra force could be undetectable on human scales or even on the scale of our solar system, where Newton's law of gravity is observed to hold so beautifully. But he reasoned that, because it was constant throughout all of space, it could build up over the scale of our galaxy and be large enough to counteract the attractive forces between very distant objects. He thus reasoned that this could result in a static universe on the largest scales.

Einstein called this extra term the *cosmological term.* Because it is simply a constant addition to the equations, it is now, however, conventional to call this term the *cosmological constant.*

Once he recognized that the universe is actually expanding, Einstein dispensed with this term and is said to have called the decision to add it to his equations his biggest blunder.

But getting rid of it is not so easy. It is like trying to put the toothpaste back in the tube after you have squeezed it out. This is because we now have a completely different picture of the cosmological constant today, so that, if Einstein had not added the term, someone else would have in the intervening years.

Moving Einstein's term from the left-hand side of his equations to the right-hand side is a small step for a mathematician but a giant leap for a physicist. While it is trivial mathematically to do so, once this term is on the right-hand side, where all the terms contributing to the energy of the universe reside, it represents something completely different from a physical perspective—namely a new contribution to the total energy. But what kind of stuff could contribute such a term?

The answer is, *nothing.*

By *nothing,* I do not mean nothing, but rather *nothing*—in this case, the nothingness we normally call empty space. That is to say, if I take a region of space and get rid of everything within it—dust, gas, people, and even the radiation passing through, namely absolutely *everything* within that region—if the remaining empty space *weighs something,* then that would correspond to the existence of a cosmological term such as Einstein invented.

Now, this makes Einstein's cosmological constant seem even crazier! For any fourth grader will tell you how much energy is contained in nothing, even if they don't know what energy is. The answer must be nothing.

Alas, most fourth graders have not taken quantum mechanics, nor have they studied relativity. For when one incorporates the results of Einstein's special theory of relativity into the quantum universe, empty space becomes much stranger than it was before. So strange in fact that even the physicists who first discovered and analyzed this new behavior were hard-pressed to believe that it actually existed in the real world.

The first person to successfully incorporate relativity into quantum mechanics was the brilliant, laconic British theoretical physicist Paul Dirac, who himself had already played a leading role in developing quantum mechanics as a theory.

Quantum mechanics was developed from 1912 to 1927, primarily through the work of the brilliant and iconic Danish physicist Niels Bohr and the brilliant young hot-shots Austrian physicist Erwin Schrödinger and German physicist Werner Heisenberg. The quantum world first proposed by Bohr, and refined mathematically by Schrödinger and Heisenberg, defies all commonsense notions based on our experience with objects on a human scale. Bohr first proposed that electrons in atoms orbit around the central nucleus, as planets do around the Sun, but demonstrated that the observed rules of atomic spectra (the frequencies of light emitted by different elements) could only be understood if somehow the electrons were restricted to have stable orbits in a fixed set of "quantum levels" and could not spiral freely toward the nucleus. They could move between levels by absorbing or emitting only discrete frequencies, or quanta, of light —the very quanta that Max Planck had first proposed in 1905 as a way of understanding the forms of radiation emitted by hot objects.

Bohr's "quantization rules" were rather ad hoc, however. In the 1920s, Schrödinger and Heisenberg independently demonstrated that it was possible to derive these rules from first principles if electrons obeyed rules of dynamics that were different from those applied to macroscopic objects like baseballs. Electrons could behave like waves as well as particles, appearing to spread out over space (hence, Schrödinger's "wave function" for electrons), and the results of measurements of the properties of electrons were shown to yield only probabilistic determinations, with various combinations of different properties not being exactly measurable at the same time (hence, Heisenberg's "Uncertainty Principle").

Dirac had shown that the mathematics proposed by Heisenberg to describe quantum systems (for which Heisenberg won the 1932 Nobel Prize) could be derived by careful analogy with the well-known laws governing the dynamics of classical macroscopic objects. In addition, he was also later able to show that the mathematical "wave mechanics" of Schrödinger could also be so derived and was formally equivalent to Heisenberg's formulation. But Dirac also knew that the quantum mechanics of Bohr, Heisenberg, and Schrödinger, as remarkable as it was, applied only to systems where Newton's laws, and not Einstein's relativity, would have been the appropriate laws governing the classical systems that the quantum systems were built with by analogy.

Dirac liked to think in terms of mathematics rather than pictures, and as he turned his attention to trying to make quantum mechanics consistent with Einstein's laws of relativity, he started playing with many different sorts of equations. These included complicated multicomponent mathematical systems that were necessary to incorporate the fact that electrons have "spin"—that is to say they spin around like small tops and have angular momentum, and they also can spin both clockwise and anticlockwise around any axis.

In 1929, he hit pay dirt. The Schrödinger equation had beautifully and accurately described the behavior of electrons moving at speeds much slower than light. Dirac found that if he modified the Schrödinger equation into a more complex equation using objects called matrices—which actually meant that his equation really described a set of four different coupled equations—he could consistently unify quantum mechanics with relativity, and thus in principle describe the behavior of systems where the electrons were moving at much faster speeds.

There was a problem, however. Dirac had written down an equation meant to describe the behavior of electrons as they interacted with electric and magnetic fields. But his equation appeared also to require the existence of new particles just like electrons but with opposite electric charge.

At the time, there was only one elementary particle in nature known with a charge opposite that of the electron—the proton. But protons are not at all like electrons. To begin with, they are 2,000 times heavier!

Dirac was flummoxed. In an act of desperation he argued that the new particles were in fact protons, but that somehow when moving through space the interactions of protons would cause them to act as if they were heavier. It didn't take long for others, including Heisenberg, to show that this suggestion made no sense.

Nature quickly came to the rescue. Within two years of the time Dirac proposed his equation, and a year after he had capitulated and accepted that, if his work was correct, then a new particle must exist, experimenters looking at cosmic rays bombarding the Earth discovered evidence for new particles identical to electrons but with an opposite electric charge, which were dubbed positrons.

Dirac was vindicated, but he also recognized his earlier lack of confidence in his own theory by later saying that his equation was smarter than he was!

We now call the positron the "antiparticle" of the electron, because it turns out that Dirac's discovery was ubiquitous. The same physics that required an antiparticle for the electron to exist requires one such particle to exist for almost every elementary particle in nature. Protons have antiprotons, for example. Even some neutral particles, like neutrons, have antiparticles. When particles and antiparticles meet, they annihilate into pure radiation.

While all this may sound like science fiction (and indeed antimatter plays an important role in *Star Trek*), we create antiparticles all the time at our large particle accelerators around the world. Because antiparticles otherwise have the same properties as particles, a world made of antimatter would behave the same way as a world of matter, with antilovers sitting in anticars making love under an anti-Moon. It merely is an accident of our circumstances, due, we think, to rather more profound factors we will get to later, that we live in a universe that is made up of matter and not antimatter or one with equal amounts of both. I like to say that while antimatter may seem strange, it is strange in the sense that Belgians are strange. They are not really strange; it is just that one rarely meets them.

The existence of antiparticles makes the observable world a much more interesting place, but it also turns out to make empty space much more complicated.

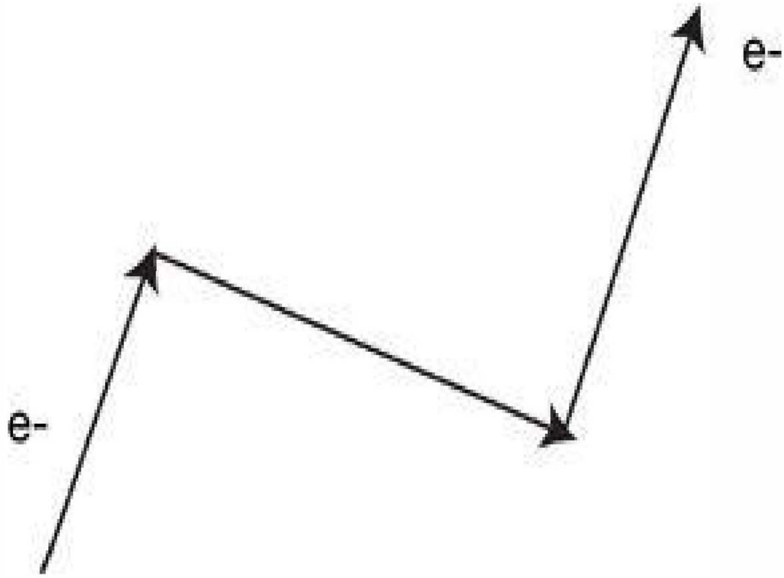
Legendary physicist Richard Feynman was the first person to provide an intuitive understanding of why relativity requires the existence of antiparticles, which also yielded a graphic demonstration that empty space is not quite so empty.

Feynman recognized that relativity tells us that observers moving at different speeds will make different measurements of quantities such as distance and time. For example, time will appear to slow down for objects moving very fast. If somehow objects could travel faster than light, they would appear to go backward in time, which is one of the reasons that the speed of light is normally considered a cosmic speed limit.

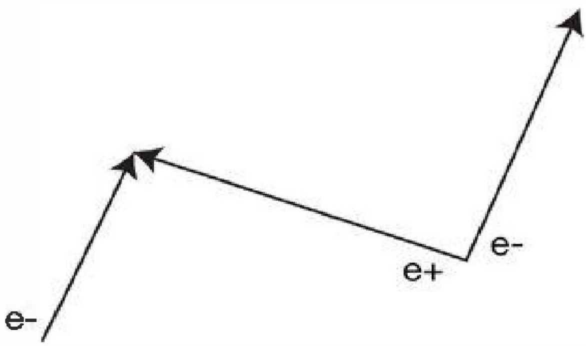
A key tenet of quantum mechanics, however, is the Heisenberg Uncertainty Principle, which, as I have mentioned, states that it is impossible to determine, for certain pairs of quantities, such as position and velocity, exact values for a given system at the same time. Alternatively, if you measure a given system for only a fixed, finite time interval, you cannot determine its total energy exactly.

What all this implies is that, for very short times, so short that you cannot measure their speed with high precision, quantum mechanics allows for the possibility that these particles act as if they are moving faster than light! But, if they are moving faster than light, Einstein tells us they must be behaving as if they are moving backward in time!

Feynman was brave enough to take this apparently crazy possibility seriously and explore its implications. He drew the following diagram for an electron moving about, periodically speeding up in the middle of its voyage to faster-than-light speed.

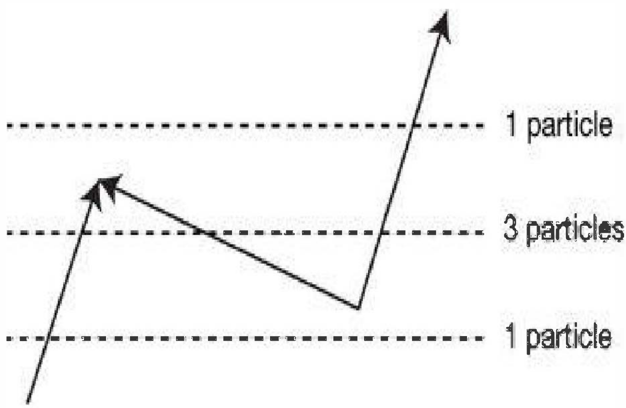


He recognized that relativity would tell us that another observer might alternatively measure something that would appear as shown below, with an electron moving forward in time, then backward in time, and then forward again.

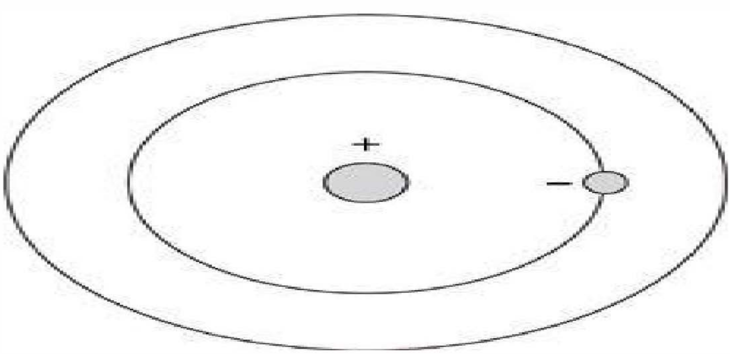


However, a negative charge moving backward in time is mathematically equivalent to a positive charge moving forward in time! Thus, relativity would require the existence of positively charged particles with the same mass and other properties as electrons.

In this case one can reinterpret Feynman's second drawing as follows: a single electron is moving along, and then at another point in space a positron-electron pair is created out of nothing, and then the positron meets the first electron and the two annihilate. Afterward, one is left with a single electron moving along.



If this doesn't bother you, then consider the following: for a little while, even if you start out with just a single particle, and end with a single particle, for a short time there are three particles moving about:



In the brief middle period, for at least a little while, something has spawned out of nothing! Feynman beautifully describes this apparent paradox in his 1949 paper, “A Theory of Positrons,” with a delightful wartime analogy:

> It is as though a bombardier watching a single road through the bomb-sight of a low-flying plane suddenly sees three roads and it is only when two of them come together and disappear again that he realizes that he has simply passed over a long switchback in a single road.

As long as this time period during this “switchback” is so short that we cannot measure all the particles directly, quantum mechanics and relativity imply that not only is this weird situation allowed, it is required. The particles that appear and disappear in timescales too short to measure are called *virtual* particles.

Now inventing a whole new set of particles in empty space that you cannot measure sounds a lot like proposing a large number of angels sitting on the head of a pin. And it would be about as impotent an idea if these particles had no other measurable effects. However, while they are not directly observable, it turns out their *indirect* effects produce most of the characteristics of the universe we experience today. Not only this, but one can calculate the impact of these particles more precisely than any other calculation in science.

Consider, for example, a hydrogen atom—the system Bohr tried to explain by developing his quantum theory and Schrödinger later tried to describe by deriving his famous equation. The beauty of quantum mechanics was that it could explain the specific colors of light emitted by hydrogen when it was heated up by arguing that electrons orbiting around the proton could exist only in discrete energy levels, and when they jumped between levels they absorbed or emitted only a fixed set of frequencies of light. The Schrödinger equation allows one to calculate the predicted frequencies, and it gets the answer almost exactly right.

But not exactly.

When the spectrum of hydrogen was observed more carefully, it was seen to be more complicated than had previously been estimated, with some additional small splittings between levels observed, called the “fine structure” of the spectrum. While these splittings had been known since Bohr’s time, and it was suspected that perhaps relativistic effects had something do to with them, until a fully relativistic theory was available, no one could confirm this suspicion. Happily, Dirac’s equation managed to improve the predictions compared to Schrödinger’s equation and reproduced the general structure of the observations, including fine structure.

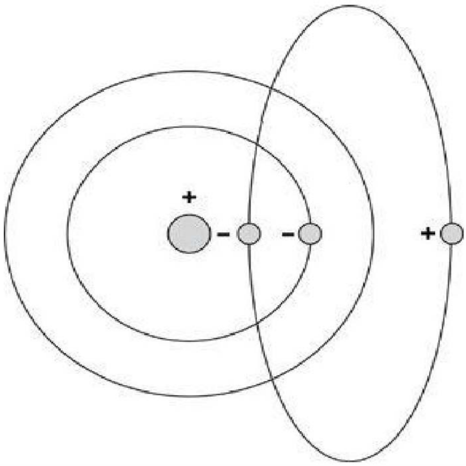
So far so good, but in April of 1947, United States experimentalist Willis Lamb and his student Robert C. Retherford performed an experiment that might otherwise seem incredibly ill motivated. They realized that they had the technological ability to measure the energy-level structure of the level of hydrogen atoms with an accuracy of 1 part in 100 million.

Why would they bother? Well, whenever experimentalists find a new method to measure something with vastly greater precision than was possible before, that is often sufficient motivation for them to go ahead. Whole new worlds are often revealed in the process, as when the Dutch scientist Antonie Philips van Leeuwenhoek first stared at a drop of seemingly empty water with a microscope in 1676 and discovered it was teeming with life. In this case, however, the experimenters had more immediate motivation. Up until the time of Lamb’s experiment, the available experimental precision could not test Dirac’s prediction in detail.

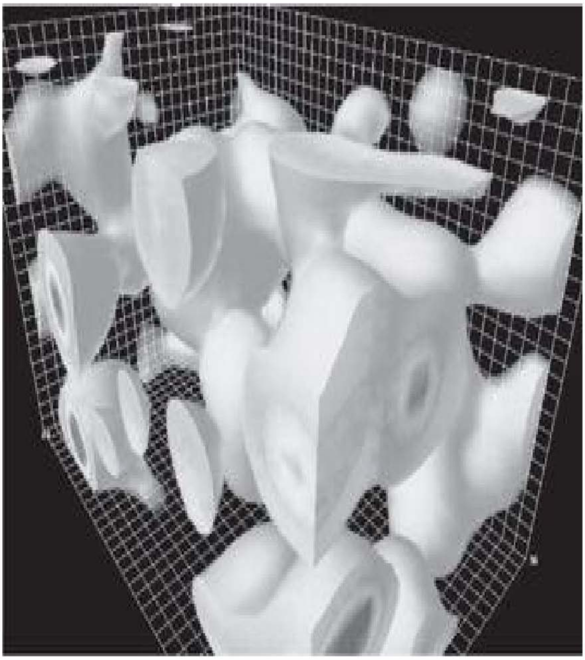
The Dirac equation did predict the general structure of the new observations, but the key question that Lamb wanted to answer was whether it predicted it in detail. This was the only way to actually test the theory. And when Lamb tested the theory, it seemed to give the wrong answer, at a level of about 100 parts per billion, well above the sensitivity of his apparatus.

Such a small disagreement with experiment may not seem like a lot, but the predictions of the simplest interpretation of the Dirac theory were unambiguous, as was the experiment, and they differed.

Over the next few years, the best theoretical minds in physics jumped into the fray and tried to resolve the discrepancy. The answer came after a great deal of work, and when the dust had settled, it was realized that the Dirac equation actually gives precisely the correct answer, but only if you include the effect of virtual particles. Pictorially, this can be understood as follows. Hydrogen atoms are usually pictured in chemistry books something like this, with a proton at the center and an electron orbiting around it, jumping between different levels:



However, once we allow for the possibility that electron-positron pairs can spontaneously appear from nothing for a bit before annihilating each other again, over any short time the hydrogen atom really looks like this:



At the right of the figure I have drawn such a pair, which then annihilate at the top. The virtual electron, being negatively charged, likes to hang around closer to the proton, while the positron likes to stay farther away. In any case, what is clear from this picture is that the actual charge distribution in a hydrogen atom is *not,* at any instant, described by simply a single electron and proton.

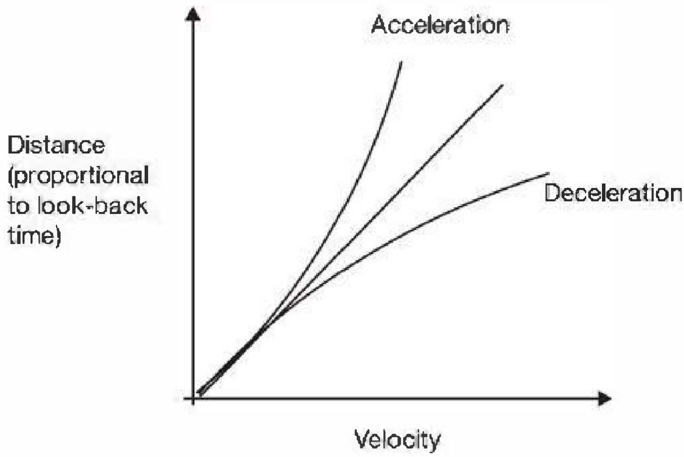
Remarkably, we physicists have learned (after all the hard work by Feynman and others) that we can use Dirac's equation to calculate to an arbitrarily high precision, the impact on the spectrum of hydrogen of all the possible virtual particles that may exist intermittently in its vicinity. And when we do, we come up with *the best, most accurate prediction in all of science*. All other scientific predictions pale in comparison. In astronomy, the most recent observations of the cosmic microwave background radiation allow us to compare with theoretical predictions at the level of perhaps 1 part in 100,000, which is remarkable. However, using Dirac's equation, and the predicted existence of virtual particles, we can calculate the value of atomic parameters and compare them with observations and have remarkable agreement at the level of about 1 part in a billion or better!

Virtual particles therefore exist.

While the spectacular precision available in atomic physics is hard to match, there is nevertheless another place where virtual particles play a key role that may actually be more relevant to the central issue of this book. It turns out that they are responsible for most of your mass, and that of everything that is visible in the universe.

One of the great successes in the 1970s in our fundamental understanding of matter came with the discovery of a theory that accurately describes the interactions of quarks, the particles that make up the protons and neutrons that form the bulk of material from which you and everything you can see are made. The mathematics associated with the theory is complex, and it took several decades before techniques were developed that could handle it, particularly in the regime where the strong interaction between the quarks became appreciable. A herculean effort was launched, including building some of the most complicated parallel processing computers, which simultaneously utilize tens of thousands of individual processors, in order to try to calculate the fundamental properties of protons and neutrons, the particles we actually measure.

After all of this work, we now have a good picture of what the inside of a proton actually looks like. There may be three quarks contained therein, but there is also a lot of other stuff. In particular, virtual particles reflecting the particles and fields that convey the strong force between quarks are popping in and out of existence all the time. Here is a snapshot of how things actually look. It is not a real photograph of course, but rather an artistic rendering of the mathematics governing the dynamics of quarks and the fields that bind them. The odd shapes and different shadings reflect the strength of the fields interacting with one another and with the quarks inside the proton as virtual particles spontaneously pop in and out of existence.



The proton is intermittently full of these virtual particles and, in fact, when we try to estimate how much they might contribute to the mass of the proton, we find that the quarks themselves provide very little of the total mass and that the fields created by these particles contribute most of the energy that goes into the proton's rest energy and, hence, its rest mass. The same is true for the neutron, and since you are made of protons and neutrons, the same is true for you!

Now, if we can calculate the effects of virtual particles on the otherwise empty space in and around atoms, and we can calculate the effects of virtual particles on the otherwise empty space inside of protons, then shouldn't we be able to calculate the effects of virtual particles on truly empty space?

Well, this calculation is actually harder to do. This is because, when we calculate the effect of virtual particles on atoms or on the proton mass, we are actually calculating the total energy of the atom or proton *including* virtual particles; then, we calculate the total energy that the virtual particles would contribute without the atom or proton present (i.e., in empty space); and *then* we subtract the two numbers in order to find the net impact upon the atom or proton. We do this because it turns out that each of these two energies is formally infinite when we attempt to solve the appropriate equations, but when we subtract the two quantities, we end up with a finite difference, and moreover one that agrees precisely with the measured value!

However, if we want to calculate the effect of virtual particles on empty space alone, we have nothing to subtract, and the answer we get is therefore infinite.

Infinity is not a pleasant quantity, however, at least as far as physicists are concerned, and we try to avoid it whenever possible. Clearly, the energy of empty space (or anything else, for that matter) cannot be physically infinite, so we have to figure out a way to do the calculation and get a finite answer.

The source of the infinity is easy to describe. When we consider all possible virtual particles that can appear, the Heisenberg Uncertainty Principle (which I remind you says that the uncertainty in the measured energy of a system is inversely proportional to the length of time over which you observe it) implies that particles carrying ever more energy can appear spontaneously out of nothing as long as they then disappear in ever-shorter times. In principle, particles can therefore carry almost infinite energy as long as they disappear in almost infinitesimally short times.

However, the laws of physics as we understand them apply only for distances and times larger than a certain value, corresponding to the scale where the effects of quantum mechanics must be considered when trying to understand gravity (and its associated effects on space-time). Until we have a theory of "quantum gravity," as it is called, we can't trust extrapolations that go beyond these limits.

Thus, we might hope that the new physics associated with quantum gravity will somehow cut off the effects of virtual particles that live for less time than the "Planck-time," as it is called. If we then consider the cumulative effects of only virtual particles of energies equal to or lower than that allowed by this temporal cutoff, we arrive at a finite estimate for the energy that virtual particles contribute to nothing.

But there is a problem. This estimate turns out to be about

1,000,000,000,000,000,000,000,000,000,000,
000,000,000,000,000,000,000,000,000,000,
000,000,000,000,000,000,000,000,000,000,

000,000,000,000,000,000,000,000,000,000
times larger than the energy associated with all the known matter in the universe, including dark matter!

If the calculation of the atomic energy level spacings including virtual particles is the best computation in all of physics, this estimate of the energy space—120 orders of magnitude larger than the energy of everything else in the universe—is undoubtedly the worst! If the energy of empty space were anywhere near this large, the repulsive force induced (remember the energy of empty space corresponds to a cosmological constant) would be large enough to blow up the Earth today, but more important, it would have been so great at early times that everything we now see in our universe would have pushed apart so quickly in the first fraction of a second of the Big Bang that no structure, no stars, no planets, and no people would ever have formed.

This problem, appropriately called the Cosmological Constant Problem, has been around since well before I was a graduate student, first made explicit by the Russian cosmologist Yakov Zel'dovich around 1967. It remains unsolved and is perhaps the most profound unsolved fundamental problem in physics today.

In spite of the fact that we have had no idea how to solve the problem for more than forty years, we theoretical physicists knew what the answer had to be. Like the fourth grader who I suggested would have guessed that the energy of empty space had to be zero, we too felt that when an ultimate theory was derived, it would explain how the effects of virtual particles would cancel, leaving empty space with precisely zero energy. Or nothing. Or rather, Nothing.

Our reasoning was better than the fourth grader's, or so we thought. We needed to reduce the magnitude of the energy of empty space from the truly gargantuan value that the naïve estimate suggested to a value consistent with the upper limits allowed by observation. This would require some way to subtract from a very large positive number another very large positive number so the two would cancel to 120 decimal places, leaving something non-zero in the 121st decimal place! But there is no precedent in science for canceling two large numbers to such accuracy, with only something minuscule left over.

However, zero is a number that is easy to produce. Symmetries of nature often allow us to demonstrate that there are precisely equal and opposite contributions coming from different parts of a calculation, canceling out exactly, with precisely nothing left over. Or, again, Nothing.

Thus, we theorists were able to rest easy and sleep at night. We didn't know how to get there, but we were sure what the final answer had to be.

Nature, however, had something different in mind.