CHAPTER 10

Nothing Is Unstable

Fiat justitia—ruat caelum. *(Do justice, and let the skies fall.)*

—ANCIENT ROMAN PROVERB

The existence of energy in empty space—the discovery that rocked our cosmological universe and the idea that forms the bedrock of inflation—only reinforces something about the quantum world that was already well established in the context of the kinds of laboratory experiments I have already described. Empty space is complicated. It is a boiling brew of virtual particles that pop in and out of existence in a time so short we cannot see them directly.

Virtual particles are manifestations of a basic property of quantum systems. At the heart of quantum mechanics is a rule that sometimes governs politicians or CEOs—as long as no one is watching, anything goes. Systems continue to move, if just momentarily, between all possible states, including states that would not be allowed if the system were actually being measured. These "quantum fluctuations" imply something essential about the quantum world: nothing always produces something, if only for an instant.

But here's the rub. The conservation of energy tells us that quantum systems can misbehave for only so long. Like embezzling stockbrokers, if the state that a system fluctuates into requires sneaking some energy from empty space, then the system has to return that energy in a time short enough so that no one measuring the system can detect it.

As a result, you might presume to safely argue that this "something" that is produced by quantum fluctuations is ephemeral—not measurable, unlike, say, you or I or the Earth on which we live. But this ephemeral creation, too, is subject to the circumstances associated with our measurements. For example, consider the electric field emanating from a charged object. It is definitely real. You can feel the static electric force on your hair or watch a balloon stick to a wall. However, the quantum theory of electromagnetism suggests that the static field is due to the emission, by the charged particles involved in producing the field, of virtual photons that have essentially zero total energy. These virtual particles, because they have zero energy, can propagate across the universe without disappearing, and the field due to the superposition of many of them is so real it can be felt.

Sometimes conditions are such that real, massive particles can actually pop out of empty space with impunity. In one example, two charged plates are brought close together and, once the electric field gets strong enough between them, it becomes energetically favorable for a real particle-antiparticle pair to "pop" out of the vacuum, with the negative charge heading toward the positive plate and the positive charge toward the negative one. In so doing, it is possible that the reduction in energy arising from reducing the net charge on each of the plates and hence the electric field between them can be greater than the energy associated with the rest mass energy required to produce two real particles. Of course, the strength of the field has to be huge for such a condition to be possible.

There is actually a place where strong fields of a different kind might allow a phenomenon similar to that described above to occur—but in this case due to gravity. This realization actually made Stephen Hawking famous among physicists in 1974, when he showed that it might be possible for black holes—out of which, in the absence of quantum mechanical considerations at least, nothing can ever escape—to radiate physical particles.

There are many different ways to try to understand this phenomenon, but one of these is strikingly familiar to the situation

I described above with electric fields. Outside of the core of black holes is a radius called the "event horizon." Inside an event horizon, no object can classically escape because the escape velocity exceeds the speed of light. Thus, even light emitted inside this region will not make it outside the event horizon.

Now imagine a particle-antiparticle pair nucleates out of empty space just outside of the event horizon due to quantum fluctuations in that region. It is possible, if one of the particles actually falls within the event horizon, for it to lose enough gravitational energy by falling into the black hole that this energy exceeds twice the rest mass of either particle. This means that the partner particle can fly off to infinity and be observable without any violation of energy conservation. The total positive energy associated with the radiated particle is more than compensated by the loss of energy experienced by its partner particle falling into the black hole. The black hole can therefore radiate particles.

The situation is even more interesting, however, precisely because the energy lost by the infalling particle is greater than the positive energy associated with its rest mass. As a result, when it falls into the black hole, the net system of the black hole plus the particle actually has less energy than it did before the particle fell in! The black hole therefore actually gets *lighter* after the particle falls in by an amount that is equivalent to the energy carried away by the radiated particle that escapes. Eventually the black hole may radiate away entirely. At this point we do not know because the final stages of black hole evaporation involve physics on such small distance scales that general relativity alone cannot tell us the final answer. On these scales, gravity must be treated as a fully quantum mechanical theory, and our current understanding of general relativity is not sufficient to allow us to determine precisely what will happen.

Nevertheless, all of these phenomena imply that, under the right conditions, not only can nothing become something, it is required to.

An early example in cosmology of the fact that "nothing" can be unstable and form something comes from efforts to understand why we live in a universe of matter.

You probably don't wake up each morning wondering about this, but the fact that our universe contains matter is remarkable. What is particularly remarkable about this is that, as far as we can tell, our universe does not contain substantial amounts of antimatter, which you will recall is required to exist by quantum mechanics and relativity, so that for every particle that we know of in nature, there can exist an equivalent antiparticle with opposite charge and the same mass. Any sensible universe at its inception, one might think, would contain equal amounts of both. After all, the antiparticles of normal particles have the same mass and similar other properties, so if particles were created at early times, it would have been equally easy to create antiparticles.

Alternatively, we could even imagine an antimatter universe in which all of the particles that make up the stars and galaxies were replaced with their antiparticles. Such a universe would appear to be almost identical to the one we live in. Observers in such a universe (themselves made of antimatter) would no doubt call what we call antimatter as matter. The name is arbitrary.

However, if our universe began sensibly, with equal amounts of matter and antimatter, and stayed that way, we wouldn't be around to ask "Why?" or "How?" This is because all particles of matter would have annihilated with all particles of antimatter in the early universe, leaving nothing but pure radiation. No matter or antimatter would be left over to make up stars, or galaxies, or to make up lovers or antilovers who might otherwise one day gaze out and be aroused by the spectacle of the night sky in each other's arms. No drama. History would consist of emptiness, a radiation bath that would slowly cool, leading ultimately to a cold, dark, bleak universe. Nothingness would reign supreme.

Scientists began to understand in the 1970s, however, that it is possible to begin with equal amounts of matter and antimatter in an early hot, dense Big Bang, and for plausible quantum processes to "create something from nothing" by establishing a small asymmetry, with a slight excess of matter over antimatter in the early universe. Then, instead of complete annihilation of matter and antimatter, leading to nothing but pure radiation today, all of the available antimatter in the early universe could have annihilated with matter, but the small excess of matter would have had no comparable amount of antimatter to annihilate with, and would then be left over. This would then lead to all the matter making up stars and galaxies we see in the universe today.

As a result, what might otherwise seem a small accomplishment (establishing a small asymmetry at early times) might instead be considered almost as the moment of creation. Because once an asymmetry between matter and antimatter was created, nothing could later put it asunder. The future history of a universe full of stars and galaxies was essentially written. Antimatter particles would annihilate with the matter particles in the early universe, and the remaining excess of matter particles would survive through the present day, establishing the character of the visible universe we know and love and inhabit.

Even if the asymmetry were 1 part in a billion there would be enough matter left over to account for everything we see in the universe today. In fact, an asymmetry of 1 part in a billion or so is precisely what was called for, because today there are roughly 1 billion photons in the cosmic microwave background for every proton in the universe. The CMBR photons are the remnants, in this picture, of the early matter-antimatter annihilations near the beginning of time.

A definitive description of how this process could have happened in the early universe is currently lacking because we have not yet fully and empirically established the detailed nature of the microphysical world at the scales where this asymmetry was likely to have been generated. Nevertheless, a host of different plausible scenarios has been explored based on the current best ideas we have about physics at these scales. While they differ in the details, they all have the same general characteristics. Quantum processes associated with elementary particles in the primordial heat bath can inexorably drive an empty universe (or equivalently an initially matter-antimatter symmetric universe) almost imperceptibly toward a universe that will be dominated by matter or antimatter.

If it could have gone either way, was it then just a circumstantial accident that our universe became dominated by matter? Imagine standing on top of a tall mountain and tripping. The direction you fall was not preordained, but rather is an accident, depending upon which direction you were looking in or at what point in your stride you tripped. Perhaps similarly our universe is like that, and even if the laws of physics are fixed, the ultimate direction of the asymmetry between matter and antimatter was driven by some random initial condition (just as in the case of tripping down the mountain, the law of gravity is fixed and determines that you will fall, but your direction may be an accident). Once again, our very existence in that case would be an environmental accident.

Independent of this uncertainty, however, is the remarkable fact that a feature of the underlying laws of physics can allow quantum processes to drive the universe away from a featureless state. Physicist Frank Wilczek, who was one of the first theorists to explore these possibilities, has reminded me that he utilized precisely the same language I have used previously in this chapter, in the 1980 *Scientific American* article he wrote on the matter-antimatter asymmetry of the universe. After describing how a matter-antimatter asymmetry might plausibly be generated in the early universe based on our new understanding of particle physics, he added a note that this provided one way of thinking about the answer to the question of why there is something rather than nothing: *nothing* is unstable.

The point Frank was emphasizing is that the measured excess of matter over antimatter in the universe appears on first glance to be an obstacle to imagining a universe that could arise from an instability in empty space, with nothingness producing a Big Bang. But if that asymmetry could arise dynamically after the Big Bang, that barrier is removed. As he put it:

> One can speculate that the universe began in the most symmetrical state possible and that in such a state no matter existed; the universe was a vacuum. A second state existed, and in it matter existed. The second state had slightly less symmetry, but was also lower in energy. Eventually a patch of less symmetrical phase appeared and grew rapidly. The energy released by the transition found form in the creation of particles. This event might be identified with the big

> bang . . . The answer to the ancient question "Why is there something rather than nothing?" would be that "nothing" is unstable.

Before I proceed, however, I am again reminded of the similarities between the discussion I have just given of a matter-antimatter asymmetry and the discussions we had at our recent Origins workshop to explore our current understanding of the nature of life in the universe and its origin. My words were different, but the fundamental issues are remarkably similar: What specific physical process in the early moments of the Earth's history could have led to the creation of the first replicating biomolecules and metabolism? As in the 1970s in physics, the recent decade has seen incredible progress in molecular biology. We learned of natural organic pathways, for example, that could produce, under plausible conditions, ribonucleic acids, long thought to be the precursors to our modern DNA-based world. Until recently it was felt that no such direct pathway was possible and that some other intermediate forms must have played a key role.

Now few biochemists and molecular biologists doubt that life can arise naturally from nonlife, even though the specifics are yet to be discovered. But, as we discussed all of this, a common subtext permeated our proceedings: Did the life that first formed on Earth *have* to have the chemistry that it did, or are there many different, equally viable possibilities?

Einstein once asked a question that, he said, was the one thing he really wanted to know about nature. I admit it is the most profound and fundamental question that many of us would like answered. He put it as follows: "What I want to know is whether *God* [*sic*] had any choice in the creation of the universe."

I have annotated this because Einstein's God was not the God of the Bible. For Einstein, the existence of order in the universe provided a sense of such profound wonder that he felt a spiritual attachment to it, which he labeled, motivated by Spinoza, with the moniker "God." In any case, what Einstein really meant in this question was the issue I have just described in the context of several different examples: Are the laws of nature unique? And is the universe we inhabit, which has resulted from these laws, unique? If you change one facet, one constant, one force, however slight, would the whole edifice crumble? In a biological sense, is the biology of life unique? Are we unique in the universe? We will return to discuss this most important question later in this book.

While such a discussion will cause us to further refine and generalize notions of "nothing" and "something," I want to return to taking an intermediate step in making the case for the inevitable creation of something.

As I have defined it thus far, the relevant "nothing" from which our observed "something" arises is "empty space." However, once we allow for the merging of quantum mechanics and general relativity, we can extend this argument to the case where space itself is forced into existence.

General relativity as a theory of gravity is, at its heart, a theory of space and time. As I described in the very beginning of this book, this means that it was the first theory that could address the dynamics not merely of objects moving through space, but also how space itself evolves.

Having a quantum theory of gravity would therefore mean that the rules of quantum mechanics would apply to the properties of space and not just to the properties of objects existing in space, as in conventional quantum mechanics.

Extending quantum mechanics to include such a possibility is tricky, but the formalism Richard Feynman developed, which led to a modern understanding of the origin of antiparticles, is well suited to the task. Feynman's methods focus on the key fact to which I alluded at the beginning of this chapter: quantum mechanical systems explore all possible trajectories, even those that are classically forbidden, as they evolve in time.

In order to explore this, Feynman developed a "sum over paths formalism" to make predictions. In this method, we consider all possible trajectories between two points that a particle might take. We then assign a probability weighting for each trajectory, based on well-defined principles of quantum mechanics, and then perform a sum over all paths in order to determine final (probabilistic) predictions for the motion of particles.

Stephen Hawking was one of the first scientists to fully exploit this idea to the possible quantum mechanics of space-time (the union of our three-dimensional space along with one dimension of time to form a four-dimensional unified space-time system, as required by Einstein's special theory of relativity). The virtue of Feynman's methods was that focusing on all possible paths ends up meaning that the results can be shown to be independent of the specific space and time labels one applies to each point on each path. Because relativity tells us that different observers in relative motion will measure distance and time differently and therefore assign different values to each point in space and time, having a formalism that is independent of the different labels that different observers might assign to each point in space and time is particularly useful.

And it is most useful perhaps in considerations of general relativity, where the specific labeling of space and time points becomes completely arbitrary, so that different observers at different points in a gravitational field measure distances and times differently, and all that ultimately determines the behavior of systems are geometric quantities like curvature, which turn out to be independent of all such labeling schemes.

As I have alluded to several times, general relativity is not fully consistent with quantum mechanics, at least as far as we can tell, and therefore there is no completely unambiguous method to define Feynman's sum-over-paths technique in general relativity. So we have to make some guesses in advance based on plausibility and check to see if the results make sense.

If we are to consider the quantum dynamics of space and time then, one must imagine that in the Feynman "sums," one must consider every different possible configuration that can describe the different geometries that space can adopt during the intermediate stages of any process, when quantum indeterminacy reigns supreme. This means we must consider spaces that are arbitrarily highly curved over short distances and small times (so short and so small that we cannot measure them so that quantum weirdness can reign free). These weird configurations would then not be observed by large classical observers such as us when we attempt to measure the properties of space over large distances and times.

But let's consider even stranger possibilities. Remember that, in the quantum theory of electromagnetism, particles can pop out of empty space at will as long as they disappear again on a time frame determined by the Uncertainty Principle. By analogy, then, in the Feynman quantum sum over possible space-time configurations, should one consider the possibility of small, possibly compact spaces that themselves pop in and out of existence? More generally, what about spaces that may have "holes" in them, or "handles" like donuts dunking into space-time?

These are open questions. However, unless one can come up with a good reason for excluding such configurations from the quantum mechanical sum that determines the properties of the evolving universe, and to date no such good reason exists that I know of, then under the general principle that holds everywhere else I know of in nature—namely that anything that is not proscribed by the laws of physics must actually happen—it seems most reasonable to consider these possibilities.

As Stephen Hawking has emphasized, a quantum theory of gravity allows for the creation, albeit perhaps momentarily, of space itself where none existed before. While in his scientific work he was not attempting to address the "something from nothing" conundrum, effectively this is what quantum gravity may ultimately address.

"Virtual" universes—namely the possible small compact spaces that may pop into and out of existence on a timescale so short we cannot measure them directly—are fascinating theoretical constructs, but they don't seem to explain how something can arise from nothing over the long term any more than do the virtual particles that populate otherwise empty space.

However, recall that a nonzero real electric field, observable at large distances away from a charged particle, can result from the coherent emission of many virtual zero energy photons by the charge. This is because virtual photons that carry zero energy do not violate energy conservation when they are emitted. The Heisenberg Uncertainty Principle, therefore, does not constrain them to exist for only very brief times before they must be reabsorbed and disappear back into nothingness. (Again recall that the Heisenberg Uncertainty Principle states that the uncertainty with which we measure the energy of a particle, and hence the possibility that its energy may change slightly by the emission and absorption of virtual particles, is inversely proportional to the length of time over which we observe it. Hence, virtual particles that carry away zero energy can do so essentially with impunity—namely they can exist for arbitrarily long times and travel arbitrarily far away before being absorbed . . . leading to the possible existence of long-range interactions between charged particles. If the photon was not massless, so that photons always carried away non-zero energy due to a rest mass, the Heisenberg Uncertainty Principle would imply that the electric field would be short range because photons could propagate only for short times without being reabsorbed again.)

A similar argument suggests that one can imagine one specific type of universe that might spontaneously appear and need not disappear almost immediately thereafter because of the constraints of the Uncertainty Principle and energy conservation. Namely, a compact universe with zero total energy.

Now, I would like nothing better than to suggest that this is precisely the universe we live in. This would be the easy way out, but I am more interested here in being true to our current understanding of the universe than in making an apparently easy and convincing case for creating it from nothing.

I have argued, I hope compellingly, that the average Newtonian gravitational energy of every object in our flat universe is zero. And it is. But that is not the whole story. Gravitational energy is not the total energy of any object. To this energy we must add its rest energy, associated with its rest mass. Put another way, as I have described earlier, the gravitational energy of an object at rest isolated from all other objects by an infinite distance is zero, because if it is at rest, it has no kinetic energy of motion, and if it is infinitely far away from all other particles, the gravitational force on it due to other particles, which could provide potential energy to do work, is also essentially zero. However, as Einstein told us, its total energy is not merely due to gravity, but also includes the energy associated with its mass, so that, as is famously known, $E = mc^2$.

In order to take this rest energy into account, we have to move from Newtonian gravity to general relativity, which, by definition, incorporates the effects of special relativity (and $E = mc^2$) into a theory of gravity. And here things get both subtler and more confusing. On small scales compared to the possible curvature of a universe, and as long as all objects within these scales are moving slowly compared to the speed of light, the general relativistic version of energy reverts to the definition we are familiar with from Newton. However, once these conditions no longer hold, all bets are off, almost.

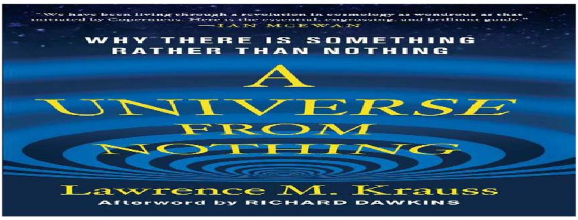
Part of the problem is that it turns out that energy as we normally think of it elsewhere in physics is not a particularly well-defined concept on large scales in a curved universe. Different ways of defining coordinate systems to describe the different labels that different observers may assign to points in space and time (called different "frames of reference") can lead, on large scales, to different determinations of the total energy of the system. In order to accommodate this effect, we have to generalize the concept of energy, and, moreover, if we are to define the total energy contained in any universe, we must consider how to add up the energy in universes that may be infinite in spatial extent.

There is a lot of debate over precisely how to do this. The scientific literature is replete with claims and counterclaims in this regard.

One thing is certain, however: There is one universe in which the total energy is definitely and precisely zero. It is not, however, a flat universe, which is in principle infinite in spatial extent, and therefore the calculation of total energy becomes problematic. It is a closed universe, one in which the density of matter and energy is sufficient to cause space to close back upon itself. As I have described, in a closed universe, if you look far enough in one direction, you will eventually see the back of your head!

The reason the energy of a closed universe is zero is really relatively simple. It is easiest to consider the result by analogy with the fact that in a closed universe the total electric charge must also be zero.

Since the time of Michael Faraday we think of electric charge as being the source of an electric field (due in modern quantum parlance to the emission of the virtual photons I described above). Pictorially, we imagine “field lines” emanating out radially from the charge, with the number of field lines being proportional to the charge, and the direction of field lines being outward for positive charges and inward for negative charges, as shown below.



We imagine these field lines going out to infinity, and as they spread out, getting farther apart. This implies that the strength of the electric field gets weaker and weaker. However, in a closed universe, the field lines associated with a positive charge, for example, may start out spreading apart but eventually, just as the lines of longitude on a map of the Earth come together at the North and South Poles, the field lines from the positive charge will come together again on the far side of the universe. When they converge, the field will get stronger and stronger again until there is enough energy to create a negative charge that can “eat” the field lines at this antipodal point of the universe.

It turns out a very similar argument, in this case associated not with the “flux” of field lines but with the “flux” of energy in a closed universe, tells us that the total positive energy, including that associated with the rest masses of particles, must be exactly compensated for by a negative gravitational energy, so that the total energy is precisely zero.

So if the total energy of a closed universe is zero, and if the sum-over-paths formalism of quantum gravity is appropriate, then quantum mechanically such universes could appear spontaneously with impunity, carrying no net energy. I want to emphasize that these universes would be completely self-contained space-times, disconnected from our own.

There is a hitch, however. A closed expanding universe filled with matter will in general expand to a maximum size and then recollapse just as quickly, ending up in a space-time singularity where the no-man's land of quantum gravity at present cannot tell us what its ultimate fate will be. The characteristic lifetime of tiny closed universes will therefore be microscopic, perhaps on the order of the "Planck time," the characteristic scale over which quantum gravitational processes should operate, about 10^{-44} seconds or so.

There is a way out of this dilemma, however. If, before such a universe can collapse, the configuration of fields within it produces a period of inflation, then even an initially tiny closed universe can rapidly, exponentially expand, becoming closer and closer to an infinitely large flat universe during this period. After one hundred or so doubling times of such inflation, the universe will be so close to flat that it could easily last much longer than our universe has been around without collapsing.

Another possibility actually exists, one that always gives me a slight twinge of nostalgia (and envy), because it represented an important learning experience for me. When I was first a postdoc at Harvard, I was playing with the possible quantum mechanics of gravitational fields, and I learned of a result by a good friend from graduate school, Ian Affleck. A Canadian who had been a graduate student at Harvard when I was at MIT, Affleck joined the Society of Fellows a few years before I did and had used the mathematical theory of Feynman that we now use for dealing with elementary particles and fields, called quantum field theory, to calculate how particles and antiparticles could be produced in a strong magnetic field.

I realized that the form of the solution that Ian had described, something called an "instanton," resembled very much an inflating universe, if one took over his formalism to the case of gravity. But it looked like an inflating universe that began from nothing! Before writing up this result, I wanted to address my own confusion about how to interpret what physics such a mathematical solution might correspond to. I soon learned, however, that while I was cogitating, just down the road the very creative cosmologist I mentioned earlier, Alex Vilenkin, who has since become a friend, had actually just written a paper that described in exactly this fashion how quantum gravity indeed might create an inflating universe directly from nothing. I was scooped, but I couldn't be that upset because (a) I frankly didn't understand in detail at that point what I was doing, and (b) Alex had the boldness to propose something that at the time I didn't. I have since learned that one doesn't have to understand all the implications of one's work in order to publish. Indeed, there are several of my own most important papers that I only fully understood well after the fact.

In any case, while Stephen Hawking and his collaborator Jim Hartle have proposed a very different scheme for trying to determine the "boundary conditions" on universes that may begin from nothing at all, the important facts are these:

1. In quantum gravity, universes can, and indeed always will, spontaneously appear from nothing. Such universes need not be empty, but can have matter and radiation in them, as long as the total energy, including the negative energy associated with gravity, is zero.
2. In order for the closed universes that might be created through such mechanisms to last for longer than infinitesimal times, something like inflation is necessary. As a result, the only long-lived universe one might expect to live in as a result of such a scenario is one that today appears flat, just as the universe in which we live appears.

The lesson is clear: quantum gravity not only appears to allow universes to be created from nothing—meaning, in this case, I emphasize, the absence of space and time—it may require them. "Nothing"—in this case no space, no time, no anything!—*is* unstable.

Moreover, the general characteristics of such a universe, if it lasts a long time, would be expected to be those we observe in our universe today.

Does this prove that our universe arose from nothing? Of course not. But it does take us one rather large step closer to the plausibility of such a scenario. And it removes one more of the objections that might have been leveled against the argument of creation from nothing as described in the previous chapter.

There, "nothing" meant empty but preexisting space combined with fixed and well-known laws of physics. Now the requirement of space has been removed.

But, remarkably, as we shall next discuss, even the laws of physics may not be necessary or required.