CHAPTER 6

The Free Lunch at the End of the Universe

Space is big. Really big. You just won't believe how vastly, hugely, mind-bogglingly big it is. I mean, you may think it's a long way down the road to the chemist's, but that's just peanuts to space.

—Douglas Adams, *The Hitchhiker's Guide to the Galaxy*

One out of two isn't bad, I suppose. We cosmologists had guessed, correctly it turned out, that the universe is flat, so we weren't that embarrassed by the shocking revelation that empty space indeed has energy—and enough energy in fact to dominate the expansion of the universe. The existence of this energy was implausible, but even more implausible is that the energy is not enough to make the universe uninhabitable. For if the energy of empty space were as large as the a priori estimates I described earlier suggested it should be, the expansion rate would have been so great that everything that we now see in the universe would quickly have been driven beyond the horizon. The universe would become cold, dark, and empty well before stars, our Sun, and our Earth could have formed.

Of all the reasons to suppose that the universe was flat, perhaps the simplest to understand arose from the fact that the universe had been well-known to be almost flat. Even in the early days, before dark matter was discovered, the known amount of visible material in and around galaxies accounted for perhaps 1 percent of the total amount of matter needed to result in a flat universe.

Now, 1 percent may not seem like much, but our universe is very old, billions of years old. Assuming that the gravitational effects of matter or radiation dominate the evolving expansion, which is what we physicists always thought was the case, then if the universe is not precisely flat, as it expands, it moves further and further away from being flat.

If it is open, the expansion rate continues at a faster rate than it would for a flat universe, driving matter farther and farther apart compared to what it would be otherwise, reducing its net density and very quickly yielding an infinitesimally small fraction of the density required to result in a flat universe.

If it is closed, then it slows the expansion down faster and eventually causes it to recollapse. All the while, the density first decreases at a slower rate than for a flat universe, and then as the universe recollapses, it starts to increase. Once again, the departure from the density expected for a flat universe increases with time.

The universe has increased in size by a factor of almost a trillion since it was 1 second old. If, at that earlier moment, the density of the universe was not almost exactly that expected of a flat universe but was, say, only 10 percent of that appropriate for a flat universe at the time, then today the density of our universe would differ from that of a flat universe by at least a factor of a trillion. This is far greater than the mere factor of 100 that was known to separate the density of the visible matter in the universe from the density of what would produce a flat universe today.

This problem was well-known, even in the 1970s, and it became known as the Flatness Problem. Considering the geometry of the universe is like imagining a pencil balancing vertically on its point on a table. The slightest imbalance one way or the other and it will quickly topple. So it is for a flat universe. The slightest departure from flatness quickly grows. Thus, how could the universe be so close to being flat today if it were not *exactly* flat?

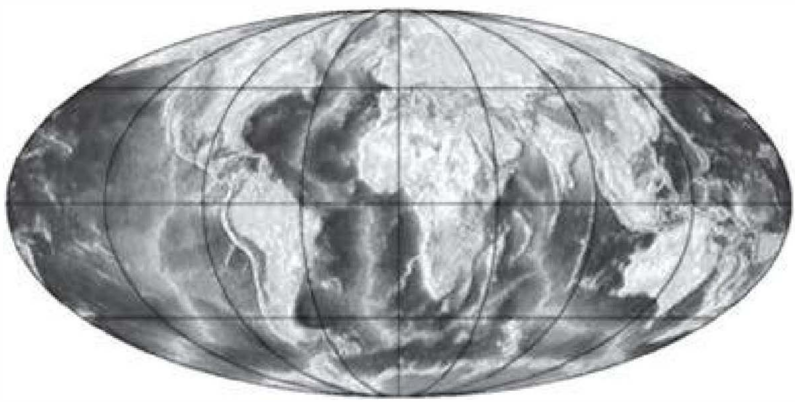
The answer is simple: it must be essentially flat today!

That answer's actually not so simple, because it begs the question, How did initial conditions conspire to produce a flat universe?

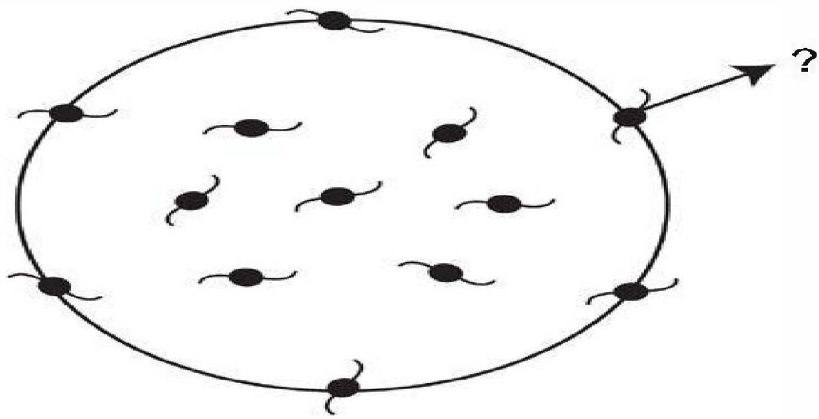
There are two answers to this second, more difficult question. The first goes back to 1981, when a young theoretical physicist and postdoctoral researcher at Stanford University, Alan Guth, was thinking about the Flatness Problem and two other related problems with the standard Big Bang picture of the universe: the so-called Horizon Problem and the Monopole Problem. Only the first need concern us here, since the Monopole Problem merely exacerbates both the Flatness and Horizon problems.

The Horizon Problem relates to the fact that the cosmic microwave background radiation is extremely uniform. The small temperature deviations, which I described earlier, represented density variations in matter and radiation back when the universe was a few hundred thousand years old, of less than 1 part in 10,000 compared to the otherwise uniform background density and temperature. So while I was focused on the small deviations, a deeper, more urgent question was, How did the universe get to be so uniform in the first place?

After all, if instead of the earlier image of the CMBR (where temperature variations of a few parts in 100,000 are reflected in different colors), I showed a temperature map of the microwave sky on a linear scale (with variations in shades representing variations in temperature, of say, ±0.03 degree [Kelvin] about the mean background temperature of about 2.72 degrees above absolute zero, or a variation of 1 part in 100 about the mean), the map would look like this:



Compare this image, which contains nothing discernible in the way of structure, to a similar projection of the surface of the Earth, with only slightly greater sensitivity, with color variations representing variations about the mean radius by about 1 part in 500 or so:



The universe is, therefore, on large scales, *incredibly uniform*!

How could this be? Well, one might simply assume that, at early times, the early universe was hot, dense, and in thermal equilibrium. This means that any hot spots would have cooled, and cold spots would have heated up until the primordial soup reached the same temperature throughout.

However, as I pointed out earlier, when the universe was a few hundred thousand years old, light could have traveled only a few hundred thousand light-years, representing a small percentage of what is now the total observable universe (this former distance would represent merely an angle of about 1 degree on a map of the complete cosmic microwave background last scattering surface as it is observed today). Since Einstein tells us that no information can propagate faster than light, in the standard Big Bang picture, there is simply no way that one part of what is now the observable universe at that time would have been affected by the existence and temperature of other parts on angular scales of greater than about 1 degree. Thus, there is no way that the gas on these scales could have thermalized in time to produce such a uniform temperature throughout!

Guth, a particle physicist, was thinking about processes that could have occurred in the early universe that might have been relevant for understanding this problem when he came up with an absolutely brilliant realization. If, as the universe cooled, it underwent some kind of phase transition—as occurs, for example, when water freezes to ice or a bar of iron becomes magnetized as it cools—then not only could the Horizon Problem be solved, but also the Flatness Problem (and, for that matter, the Monopole Problem).

If you like to drink really cold beer, you may have had the following experience: you take a cold beer bottle out of the refrigerator, and when you open it and release the pressure inside the container, suddenly the beer freezes completely, during which it might even crack part of the bottle. This happens because, at high pressure, the preferred lowest energy state of the beer is in liquid form, whereas once the pressure has been released, the preferred lowest energy state of the beer is the solid state. During the phase transition, energy can be released because the lowest energy state in one phase can have lower energy than the lowest energy state in the other phase. When such energy is released, it is referred to as "latent heat."

Guth realized that, as the universe itself cooled with the Big Bang expansion, the configuration of matter and radiation in the expanding universe might have gotten "stuck" in some meta-stable state for a while until ultimately, as the universe cooled further, this configuration then suddenly underwent a phase transition to the energetically preferred ground state of matter and radiation. The energy stored in the "false vacuum" configuration of the universe before the phase transition completed—the "latent heat" of the universe, if you will—could dramatically affect the expansion of the universe during the period before the transition.

The false vacuum energy would behave just like that represented by a cosmological constant because it would act like an energy permeating empty space. This would cause the expansion of the universe at the time to speed up ever faster and faster. Eventually, what would become our observable universe would start to grow faster than the speed of light. This is allowed in general relativity, even though it seems to violate Einstein's special relativity, which says nothing can travel faster than the speed of light. But one has to be like a lawyer and parse this a little more carefully. Special relativity says nothing can travel *through space* faster than the speed of light. But *space itself* can do whatever the heck it wants, at least in general relativity. And as space expands, it can carry distant objects, which are at rest in the space where they are sitting, apart from one another at superluminal speeds.

It turns out that the universe could have expanded during this inflationary period by a factor of more than 10^{28}. While this is an incredible amount, it amazingly could have happened in a fraction of a second in the very early universe. In this case, everything within our entire observable universe was once, before inflation happened, contained in a region much smaller than we would have traced it back to if inflation had not happened, and most important, so small that there would have then been enough time for the entire region to thermalize and reach exactly the same temperature.

Inflation made another relatively generic prediction possible. When a balloon gets blown up larger and larger, the curvature at its surface gets smaller and smaller. Something similar happens for a universe whose size is expanding exponentially, as can occur during inflation—driven by a constant and large false vacuum energy. Indeed, by the time inflation ends (solving the Horizon Problem), the curvature of the universe (if it is non-zero to begin with) gets driven to an absurdly small value so that, even today, the universe appears essentially flat when measured accurately.

Inflation is the only currently viable explanation of both the homogeneity and flatness of the universe, based on what could be fundamental and calculable microscopic theories of particles and their interactions. But more than this, inflation makes another, perhaps even more remarkable prediction possible. As I have described already, the laws of quantum mechanics imply that, on very small scales, for very short times, empty space can appear to be a boiling, bubbling brew of virtual particles and fields wildly fluctuating in magnitude. These “quantum fluctuations” may be important for determining the character of protons and atoms, but generally they are invisible on larger scales, which is one of the reasons why they appear so unnatural to us.

However, during inflation, these quantum fluctuations can determine when what would otherwise be different small regions of space end their period of exponential expansion. As different regions stop inflating at slightly (microscopically) different times, the density of matter and radiation that results when the false vacuum energy gets released as heat energy in these different regions is slightly different in each one.

The pattern of density fluctuations that result after inflation—arising, I should stress, from the quantum fluctuations in otherwise empty space—turns out to be precisely in agreement with the observed pattern of cold spots and hot spots on large scales in the cosmic microwave background radiation. While consistency is not proof, of course, there is an increasing view among cosmologists that, once again, if it walks like a duck and looks like a duck and quacks like a duck, it is probably a duck. And if inflation indeed is responsible for all the small fluctuations in the density of matter and radiation that would later result in the gravitational collapse of matter into galaxies and stars and planets and people, then it can be truly said that we all are here today because of quantum fluctuations in what is essentially *nothing*.

This is so remarkable I want to stress it again. Quantum fluctuations, which otherwise would have been completely invisible, get frozen by inflation and emerge afterward as density fluctuations that produce everything we can see! If we are all stardust, as I have written, it is also true, if inflation happened, that we all, literally, emerged from quantum nothingness.

This is so strikingly nonintuitive that it can seem almost magical. But there is at least one aspect of all of this inflationary prestidigitation that might seem particularly worrisome. Where does all the energy come from in the first place? How can a microscopically small region end up as a universe-sized region today with enough matter and radiation within it to account for everything we can see?

More generally, we might ask the question, How is it that the density of energy can remain constant in an expanding universe with a cosmological constant, or false vacuum energy? After all, in such a universe, space expands exponentially, so that if the density of energy remains the same, the total energy within any region will grow as the volume of the region grows. What happened to the conservation of energy?

This is an example of something that Guth coined as the ultimate “free lunch.” Including the effects of gravity in thinking about the universe allows objects to have—amazingly—“negative” as well as “positive” energy. This facet of gravity allows for the possibility that positive energy stuff, like matter and radiation, can be complemented by negative energy configurations that just balance the energy of the created positive energy stuff. In so doing, gravity can start out with an empty universe—and end up with a filled one.

This may also sound kind of fishy, but in fact it is a central part of the real fascination that many of us have with a flat universe. It is also something that you might be familiar with from high school physics.

Consider throwing a ball up in the air. Generally, it will come back down. Now throw it harder (assuming you are not indoors). It will travel higher and stay aloft longer before returning. Finally, if you throw it hard enough, it will not come down at all. It will escape the Earth’s gravitational field and keep heading out into the cosmos.

How do we know when the ball will escape? We use a simple matter of energy accounting. A moving object in the gravitational field of the Earth has two kinds of energy. One, the energy of motion, is called *kinetic energy,* from the Greek word for motion. This energy, which depends upon the speed of the object, is always positive. The other component of the energy, called *potential energy* (related to the potential to do work), is generally negative.

This is the case because we define the total gravitational energy of an object located at rest infinitely far away from any other object as being zero, which seems reasonable. The kinetic energy is clearly zero, and we define the potential energy as zero at this point, so the total gravitational energy is zero.

Now, if the object is not infinitely far away from all other objects but is close to an object, like the Earth, it will begin to fall toward it because of the gravitational attraction. As it falls, it speeds up, and if it smacks into something on the way (say, your head), it can do work by, say, splitting it open. The closer it is to the Earth's surface when it is let go, the less work it can do by the time it hits the Earth. Thus, potential energy *decreases* as you get closer to the Earth. But if the potential energy is zero when it is infinitely far away from the Earth, it must get more and more negative the closer it gets to the Earth because its potential to do work decreases the closer it gets.

In classical mechanics, as I defined it here, the definition of potential energy is arbitrary. I could have set the potential energy of an object as zero at the Earth's surface, and then it would be some large number when the object is infinitely far away. Setting the total energy to zero at infinity does make physical sense, but it is, at least at this point in our discussion, merely a convention.

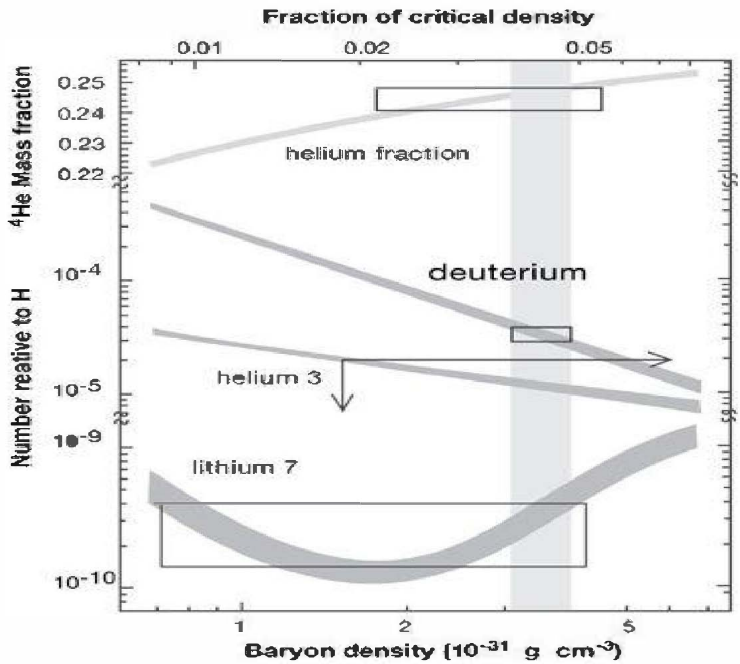
Regardless of where one sets the zero point of potential energy, the wonderful thing about objects that are subject to only the force of gravity is that the *sum* of their potential and kinetic energies remains a constant. As objects fall, potential energy is converted to the kinetic energy of motion, and as they bounce back up off the ground, kinetic energy is converted back to potential, and so on.

This allows us a marvelous bookkeeping tool to determine how fast one needs to throw something up in the air in order to escape the Earth, since if it eventually is to reach infinitely far away from the Earth, its total energy must be greater than or equal to zero. I then simply have to ensure that its total gravitational energy at the time it leaves my hand is greater than or equal to zero. Since I can control only one aspect of its total energy—namely the speed with which it leaves my hand—all I have to do is find the magic speed where the positive kinetic energy of the ball equals the negative potential energy it has due to the attraction at the Earth's surface. Both the kinetic energy and the potential energy of the ball depend precisely the same way on the mass of the ball, which therefore cancels out when these two quantities are equated, and one finds a single "escape velocity" for all objects from the

Earth's surface, namely about 7 miles per second, when the total gravitational energy of the object is precisely zero.

What has all this got to do with the universe in general, and inflation in particular, you may ask? Well, the exact same calculation I just described for a ball that I throw up from my hand at the Earth's surface applies to every object in our expanding universe.

Consider a spherical region of our universe centered on our location (in the Milky Way galaxy) and large enough to encompass a lot of galaxies but small enough so that it is well within the largest distances we can observe today:



If the region is large enough but not too large, then the galaxies at the edge of the region will be receding from us uniformly due to the Hubble expansion, but their speeds will be far less than the speed of light. In this case, the laws of Newton apply, and we can ignore the effects of special and general relativity. In other words, every object is governed by physics that is identical to that which describes the balls that I have just imagined trying to eject from the Earth.

Consider the galaxy shown above, moving away from the center of the distribution as shown. Now, just as for the ball from the Earth, we can ask whether the galaxy will be able to escape from the gravitational pull of all the other galaxies within the sphere. And the calculation we would perform to determine the answer is precisely the same as the calculation we performed for the ball. We simply calculate the total gravitational energy of the galaxy, based on its motion outward (giving it positive energy), and the gravitational pull of its neighbors (providing a negative energy piece). If its total energy is greater than zero, it will escape to infinity, and if less than zero, it will stop and fall inward.

Now, remarkably, it is possible to show that we can rewrite the simple Newtonian equation for the total gravitational energy of this galaxy in a way that reproduces *exactly* Einstein's equation from general relativity for an expanding universe. And the term that corresponds to the total gravitational energy of the galaxy becomes, in general relativity, the term that describes the curvature of the universe.

So what do we then find? In a flat universe, and *only* in a flat universe, the total average Newtonian gravitational energy of each object moving with the expansion is *precisely zero*!

This is what makes a flat universe so special. In such a universe the positive energy of motion is exactly canceled by the negative energy of gravitational attraction.

When we begin to complicate things by allowing for empty space to have energy, the simple Newtonian analogy to a ball being thrown up in the air becomes incorrect, but the conclusion remains essentially the same. In a flat universe, even one with a small cosmological constant, as long as the scale is small enough that velocities are much less than the speed of light, the Newtonian gravitational energy associated with every object in the universe is zero.

In fact, with a vacuum energy, Guth's "free lunch" becomes even more dramatic. As each region of the universe expands to ever larger size, it becomes closer and closer to being flat, so that the total Newtonian gravitational energy of everything that results after the vacuum energy during inflation gets converted to matter and radiation becomes precisely zero.

But you can still ask, Where does all the energy come from to keep the density of energy constant during inflation, when the universe is expanding exponentially? Here, another remarkable aspect of general relativity does the trick. Not only can the gravitational energy of objects be negative, but their relativistic "pressure" can be negative.

Negative pressure is even harder to picture than negative energy. Gas, say in a balloon, exerts pressure on the walls of the balloon. In so doing, if it expands the walls of the balloon, it does work on the balloon. The work it does causes the gas to lose energy and cool. However, it turns out that the energy of empty space is gravitationally repulsive precisely because it causes empty space to have a "negative" pressure. As a result of this negative pressure, the universe actually does work *on* empty space as it expands. This work goes into maintaining the constant energy density of space even as the universe expands.

Thus, if the quantum properties of matter and radiation end up endowing even an infinitesimally small region of empty space with energy at very early times, this region can grow to be arbitrarily large and arbitrarily flat. When the inflation is over, one can end up with a universe full of stuff (matter and radiation), and the total Newtonian gravitational energy of that stuff will be as close as one can ever imagine to zero.

So when all the dust is settled, and after a century of trying, we have measured the curvature of the universe and found it to be zero. You can understand why so many theorists like me have found this not only very satisfying, but also highly suggestive.

A universe from Nothing . . . indeed.