Supplement 7

The Universe in a Nutshell

Barrow, James. The Origin of the Universe. 1994

John D Barrow. The Origin of the Universe.
1994.

CHAPTER 1

THE UNIVERSE IN A NUTSHELL

> "I must thank you," said Sherlock Holmes, "for
> calling my attention to a case which certainly
> presents some features of interest."
> —*The Hound of the Baskervilles*

How, why, and when did the universe begin? How big is it? What shape is it? What's it made of? These are questions that any curious child might ask, but they are also questions that modern cosmologists have wrestled with for many decades. One of the attractions of cosmology for popular writers and journalists is that so many of the questions at the frontiers of the subject are easy to state. Look at the frontiers of quantum electronics, DNA sequencing, neurophysiology, or pure mathematics and you will not find that the problems of the expert translate so readily into the vernacular.

Until the early years of the twentieth century, neither philosophers nor astronomers had questioned the notion that space was absolutely fixed—an arena in which the stars, the planets, and all the other heavenly bodies played out their motions. But during the 1920s this simple picture was transformed: first by the suggestions of physicists exploring the consequences of Einstein's account of gravity, and then by the results of observations of light from stars in distant galaxies by the American astronomer Edwin Hubble.

Hubble made use of a simple property of waves. If their source moves away from the receiver, the frequency with which waves are received falls. To see this, wiggle your finger up and down in some still water and watch the wave crests moving off to some other point on the water's surface. Now move your finger away from that point as you make waves, and they will be received less frequently than they were emitted. Now move your finger toward the reception point, and the reception frequency goes up. This property is shared by all waves. In the case of sound waves, it is responsible for the change in pitch of a train whistle or a police siren as it passes you. Light is also a wave, and when its source is moving away from the observer the decrease in the frequency of the light waves means that visible light is observed to be slightly redder. Hence, this effect is called a "redshift." When the light source is approaching the observer, the reception frequency increases, visible light gets bluer, and it is called a "blueshift."

Hubble discovered that the light from the galaxies he was seeing displayed a systematic redshifting. By measuring the extent of the shift, he could determine how fast the sources of light were receding; and by comparing the apparent brightnesses of stars of the same sort (stars whose intrinsic brightnesses would be the same) he could deduce their relative distances away from us. What he discovered was that the farther away the source of light, the faster it was moving away from us. This trend is known as Hubble's Law, and its illustration with modern data is shown in figure 1.1. In figure 1.2 is shown an example of the light signal received from a distant galaxy, displaying the shift of the spectrum of various atoms toward the red, as compared with that emitted from the same atoms in the laboratory.

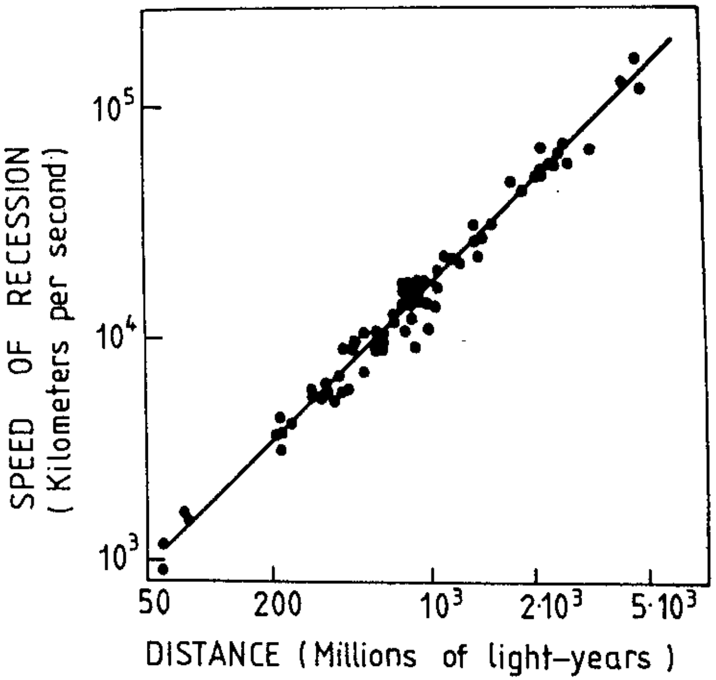

FIGURE 1.1
A modern illustration of Hubble's Law, displaying the increase of recession speed of galaxies growing in direct proportion to their distance.

....

What Hubble had discovered was the expansion of the universe. Instead of a changeless arena in which we could follow the local perambulations of planets and stars, he found that the universe was in a dynamic state. This was the greatest discovery of twentieth-century science, and it confirmed what Einstein's general theory of relativity had predicted about the universe: that it cannot be static. The

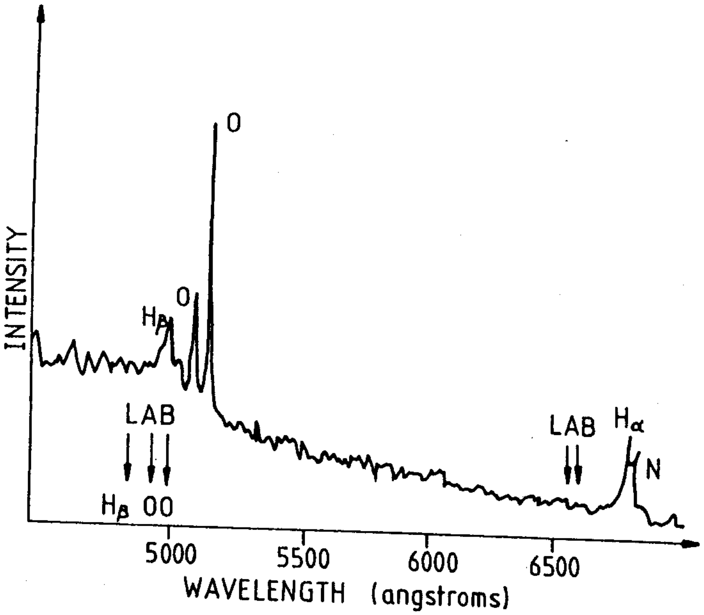

FIGURE 1.2
The spectrum of a distant galaxy (known as Markarian 609), showing how three spectral lines (marked H_β, O, and O) near 5000 angstroms and two (marked H_α and N) near 6500 angstroms are systematically shifted toward higher wavelengths than they have when measured in the laboratory. The positions of the lines in the laboratory are indicated by the arrows marked LAB; the measured positions are the labeled peaks on the graph of the light spectrum. The shift toward the red (optical red light lies at about 8000 angstroms) enables the recession speed to be calculated.

••••••••

gravitational attraction between the galaxies would bring them all together if they were not rushing away from each other. The universe can't stand still.

If the universe is expanding, then when we reverse the direction of history and look into the past we should find evidence that it emerged from a smaller, denser state—a state that appears to have once had zero size. It is this apparent beginning that has become known as the big bang.

But we are going a little too fast. There are important things to appreciate about the present expansion of the universe before we start delving into the past. First of all, what exactly is expanding? In the movie *Annie Hall*, Woody Allen is found on his analyst's couch telling of his anxiety about the expansion of the universe: "Surely this means that Brooklyn is expanding, I'm expanding, you're expanding, we're all expanding." Thankfully, he was wrong. *We* are not expanding. Nor is Brooklyn. Nor is the Earth. Nor is the solar system. Nor, in fact, is the Milky Way galaxy. Nor even those aggregates of thousands of galaxies that we call "galaxy clusters." These collections of matter are all bound together by chemical and gravitational forces between their constituents—forces that are stronger than the force of the expansion.

It is only when we get beyond the scale of great clusters of hundreds and thousands of galaxies that we see the expansion winning out over the local pull of gravity. For example, our near neighbor the Andromeda galaxy is moving toward us, because the gravitational attraction between Andromeda and the Milky Way is larger than the effect of the universal expansion. It is the galaxy clusters, not the galaxies themselves, that act as the markers of the cosmic expansion. A simple picture might be to think of specks of dust on the surface of an inflating balloon. The balloon will expand and the dust specks will move apart, but the individual dust specks will not themselves expand in the same way. They act like markers of the amount of stretching of the rubber that has occurred. Similarly, it is best to think of the expansion of the universe as the expansion of the space between clusters of galaxies, as illustrated in figure 1.3.

FIGURE 1.3
The expansion of the universe viewed as the expansion of space. Mark points on the surface of a balloon to represent galaxy clusters and inflate it. The space between the clusters increases, but the size of the clusters does not. This is analogous to a universe with two dimensions of space, represented by the surface of the balloon. Any cluster on the inflating surface sees all the other clusters receding from it. Notice that the center of the expansion does not lie on the surface of the balloon.

……

Next, we might worry about the implications of the fact that all the clusters are moving away from *us*. Why us? If we know anything about the history of science, it is that Copernicus demonstrated that the Earth is not at the center of the universe. Surely if we think that everything is moving away from *us* then we have reinstated ourselves back in the center of immensities. But this is not the case. The expanding universe is not like an explosion that has some origin at a point *in* space. There is no fixed background space into which the universe is expanding. The universe contains all the space there is!

Think of space as an elastic sheet. The presence and movement of material on this malleable space will produce indentations and curvature. The curved space of our universe is like the three-dimensional surface of a four-dimensional ball—something we cannot envisage. But imagine the universe as a flatland, with only two dimensions of space. It is then like the surface of a three-dimensional ball, which is easy to picture. Now imagine that this

three-dimensional ball can get bigger—like our inflating balloon in figure 1.3. The surface of the balloon is an expanding two-dimensional universe. If we mark two points on it, those points will recede from each other as the balloon is inflated. Now put many marks all over the surface of the balloon and inflate it again. What you find is that at whatever mark you locate yourself, all the other marks will appear to expand away from *you* as the balloon expands. You will see a Hubble Law of expansion, with the widely separated marks receding from one another faster than the ones closer together. The lesson from this example is that the surface of the balloon represents space, but the "center" of the expansion of the balloon does not lie on that surface at all. There *is* no center of expansion on the balloon's surface. Nor is there any edge. You cannot fall off the edge of the universe; the universe is not expanding into anything. It is everything there is.

One question we might raise at this stage is whether the state of expansion we witness in the universe will continue indefinitely. If we throw a stone in the air, it will return to Earth, pulled back by the force of the Earth's gravity. The harder we throw, the more energy we give the moving stone, and the higher it will go before it returns. Now, we know that if we launch a projectile faster than 11 kilometers per second it will escape the pull of the Earth's gravity. This is the critical launch speed for rockets. Space scientists call it the "escape velocity" of the Earth.

Similar considerations apply to any exploding or expanding system of material retarded by the pull of gravity. If the energy of outward motion exceeds that created by the inward pull of gravitation, the material will exceed the escape velocity and just keep on expanding. But if the attractive pull that gravity exerts between its parts is the greater, the objects in the expansion will eventually start to come back together again, just as the Earth and the stone do. So it is with expanding universes; there is a critical launch speed at the start of their expansion. If the speed exceeds this, the gravitational pull of all the material in such a universe will not be able to halt the expansion, and it will keep expanding forever. On the other hand, if the launch speed is less than the critical value, eventually the expansion will halt and reverse, culminating in a contraction back to zero size—the very same state in which it apparently began. In between, there exists what I call the "British compromise universe," which has exactly the critical launch speed—that is, the smallest value that will keep it expanding forever (see figure 1.4). One of the great mysteries about our universe is that it is currently expanding tantalizingly close to this critical case. So close, in fact, that we cannot yet say for sure on which side of the critical divide it lies. We do not know what the long-range forecast is.

Cosmologists regard the fact that we are so close to this critical divide as a peculiar property of the universe which requires an explanation. It is difficult to understand because, as the universe expands and ages, it will diverge farther and farther from the critical divide if it does not begin with precisely the critical launch speed. This creates a major puzzle. The universe has been expanding for about fifteen billion years, yet it is still so close to the critical divide that we cannot tell on which side it lies. To have remained so close after such a huge period of time turns out to require the universe's launch speed to have been "chosen" to differ from the critical one by no more than one part in ten followed by thirty-five zeros. Why?! Later, we shall see that our study of what might have happened during the first moments of the universal expansion offers a possible explanation for this highly unlikely state of affairs. But for now we shall content ourselves with understanding why any universe that contains human beings has to lie very close to that critical divide after billions of years of expansion.

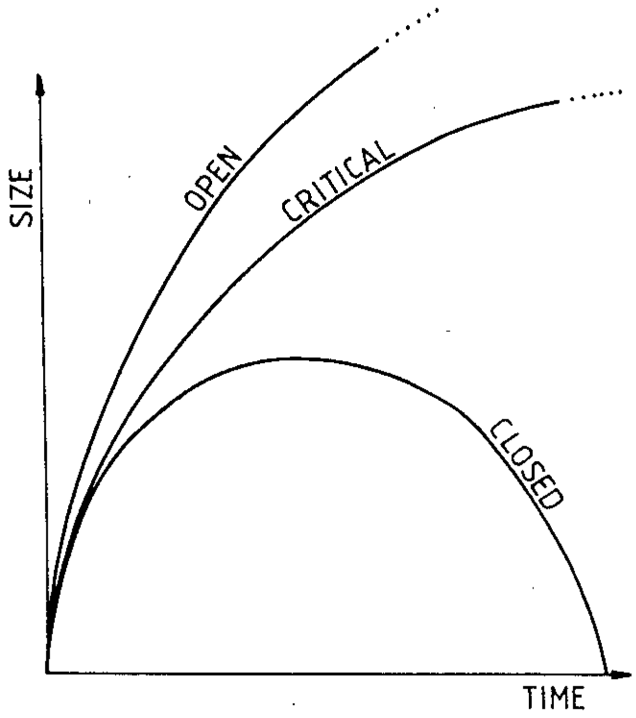

FIGURE 1.4
The three varieties of expanding universe. "Open" universes are infinite in extent and expand forever. "Closed" universes are finite and contract back to a "big crunch." The divide between the two is marked by the "critical" universe, which is infinitely large and expands forever.

....

If the universe starts expanding far faster than the critical speed, then gravity can never draw together local islands of material to form galaxies and stars. The formation of stars is a crucial step in the evolution of a universe that is to be observed. Stars are condensations of matter large enough to create at their centers pressures great enough to initiate spontaneous nuclear reactions. Those reactions burn hydrogen into helium throughout a long and sedate period of their history—a period that our sun is in the midst of—but in the final stages of their lives stars encounter a nuclear energy crisis. They undergo an explosive period of rapid change, in which helium is transformed into carbon, nitrogen, oxygen, silicon, phosphorus, and all the other elements that play a vital role in biochemistry. When stars explode in supernovas, these elements are dispersed into space and ultimately find their way into planets and people. The stars are the source of all the elements upon which complexity, and therefore life, is based. Every nucleus of carbon in our bodies originated in the stars.

So we see that universes which expand faster than the critical divide will never give birth to stars, and hence will never produce the building blocks required to make "living" entities as complex as human beings or silicon-based computers. Similarly, if a universe expands at far less than the critical speed, its expansion will be reversed into contraction before the stars have had time to form, explode, and create the constituents of living things. Again, we are left with a universe unable to give rise to life.

Thus we learn a surprising lesson: only those universes that still expand very close to the critical divide after billions of years can produce the material out of which any structure complex enough to qualify as an observer must be made (see figure 1.5). We should not be surprised to find our universe expanding so close to the critical divide. We could not exist in any other sort of universe.

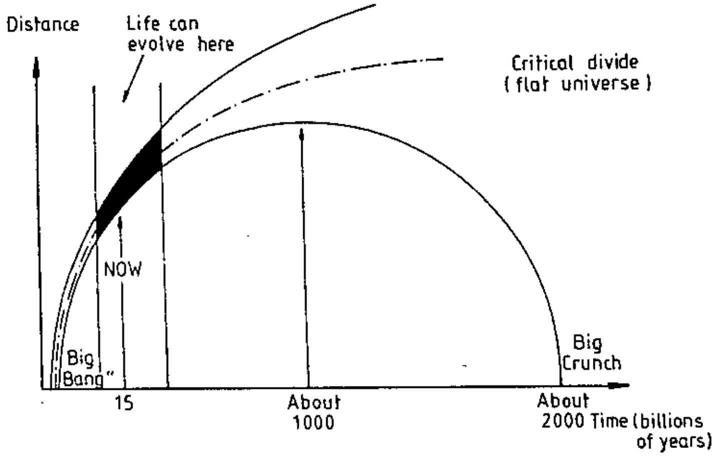

FIGURE 1.5
Universes that are too far above the critical divide expand too fast for matter to condense into stars and galaxies; such universes therefore remain devoid of life. Those that fall too far below the critical divide collapse before stars form. The shaded region indicates the range of cosmological expansions and epochs in which observers could evolve.

....

The development of our picture of the expanding universe and the reconstruction of its past history moved very slowly. During the 1930s, the Belgian priest and physicist Georges Lemaître played a leading role in its inception. His theory of the "primeval atom" was a precursor of what is now known as the big-bang model. The most important steps were taken during the late 1940s by George Gamow, a Russian émigré to the United States, together with two of his young research students, Ralph Alpher and Robert Herman. They began to take seriously the possibility of applying known physics to figure out what the early stages of an expanding universe might have been like. They recognized one key point: if the universe began in a hot, dense state in the distant past, there should remain some radiation from this explosive beginning. More specifically, they realized that when the universe was just a few minutes old it should have been hot enough for nuclear reactions to occur everywhere. Later, these important insights would be confirmed by much more detailed predictions and observations.

In 1948, Alpher and Herman predicted that the remnant radiation from the big bang, having been cooled by the expansion of the universe, should now have a temperature of around five degrees above absolute zero (absolute zero is equal to -273°C)—that is, five degrees Kelvin. Their prediction lay buried in the physics literature, however. A decade and a half later, several other scientists were considering the problem of the origin of a hot, expanding universe, but none of them knew about Alpher and Herman's paper. Communications were not then what they are now. Reconstructing the details of the universe's early history was not a very serious activity in the minds of most physicists in the 1950s and early 1960s. But in 1965, everything changed. Alpher and Herman's cosmic radiation field—manifested as microwave noise coming with the same intensity from all directions in the sky—was discovered serendipitously by Arno Penzias and Robert Wilson, two radio engineers at Bell Labs, in New Jersey, who were calibrating a sensitive radio antenna for tracking the first Echo satellite. Meanwhile, only a few miles away at Princeton University, a group led by the physicist Robert Dicke had independently recalculated what Alpher and Herman had long ago published, and had set about designing a detector to mount a search for remnant radiation from the big bang. They learned of the unexplained noise in the Bell Labs receiver and soon interpreted it as the relic radiation they were looking for. If the source was indeed heat radiation, the temperature was 2.7°K—very close to Alpher and Herman's inspired estimate. The phenomenon was dubbed the "cosmic microwave background radiation."

The discovery of the cosmic microwave background marked the beginning of the serious study of the big-bang model. Gradually, other observations revealed further properties of the background radiation. It had the same intensity in every direction to at least one part in a thousand. And as its intensity was measured at different frequencies, it began to reveal the characteristic variation of intensity with frequency which is the signature of pure heat. Such radiation is called "blackbody" radiation. Unfortunately, the absorption and emission of radiation by molecules in the Earth's atmosphere prevented astronomers from confirming that the whole spectrum of the radiation was indeed that of heat radiation. Suspicions remained that it might have been produced by violent events that occurred nearby in the universe long after the expansion began. These doubts could be overcome only by observing the radiation from above the Earth's atmosphere, and measuring the whole spectrum from space was the first great success of NASA's Cosmic Background Explorer (COBE) satellite in 1989. It was the most perfect blackbody spectrum ever seen in nature, and a striking confirmation that the universe was once hundreds of thousands of degrees hotter than it is today (see figure 1.6). For only under such extreme conditions could the radiation in the universe assume a blackbody form to such high precision.

Another key experiment to confirm that the background radiation did not have a recent origin nearby in the universe was carried out by high-flying U2 aircraft. These former spy-planes are extremely small, with large wingspans, which makes them very stable platforms for making observations. On this occasion, they were looking up rather than down, and they detected a small but systematic variation in the intensity of the radiation around the sky—a variation which had been predicted to appear if the radiation had originated in the distant past. If the radiation formed a uniformly expanding sea, emerging from the early stages of the universe, then we would

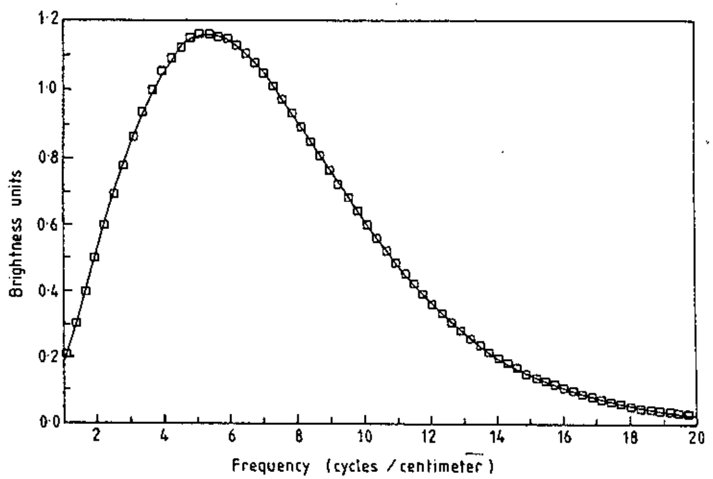

FIGURE 1.6
The variation of the intensity of the microwave background radiation with its frequency, as observed by the COBE satellite from above the Earth's atmosphere. The observations (boxes) display a perfect fit with the (solid) curve expected from pure heat radiation with a temperature of 2.73°K.

………

be moving through it. The aggregate of the Earth's motion around the sun, the sun's motion around the Milky Way, the Milky Way's motion among its neighbors, and so on, means that we are moving through the radiation in some direction (see figure 1.7). The radiation intensity will appear greatest when we look in that direction and least intense 180 degrees away, and should display a characteristic cosine variation with angle in between (see figure 1.8). It is rather like running in a rainstorm. You get wettest on your chest and stay driest on your back. Here it is microwaves that are swept up in our net direction of motion. The observations revealed a perfect cosine variation, as predicted.

Subsequently, several different experiments confirmed the discovery of "The Great Cosine in the Sky," as it became known. We, and the local cluster of galaxies in which we

reside, are moving relative to the sea of cosmic microwaves. The radiation cannot therefore have arisen locally, because it would then have shared our motion, and the cosine variation in its intensity would not have been seen.

Our motion through the background radiation from the big bang is not the only thing that can cause its intensity to vary slightly from one direction to another. If the universe is expanding at slightly different rates in different directions, the radiation will be less intense (cooler) in the directions of faster expansion. Moreover, there are large concentrations of matter, as well as regions devoid of matter, in some directions; these, too, should alter the intensity of radiation coming from those directions. It was the search for these variations that motivated the COBE satellite mission, and their discovery that made headline news all over the world in 1992.

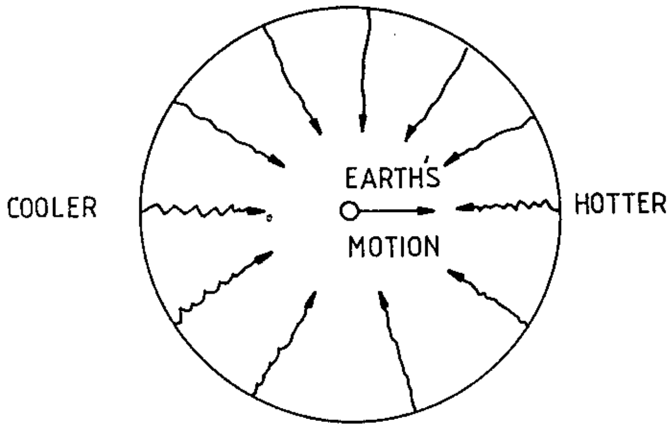

FIGURE 1.7
Our motion through the isotropic cosmic sea of microwaves arriving from the big bang. We measure the maximum intensity in the direction we are moving, and a minimum in the opposite direction, with a steady cosine variation in between.

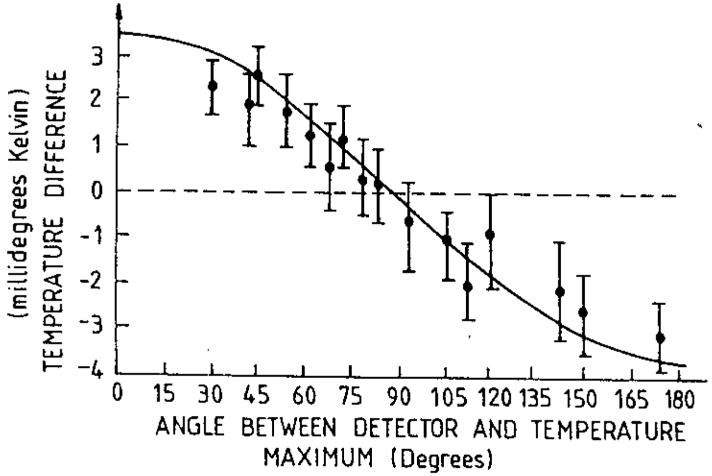

FIGURE I.8
"The Great Cosine in the Sky" shows the actual differences in temperature of the microwave background radiation in millidegrees Kelvin, as one varies the angle of observation from the direction in which it is maximum to the direction in which it is minimum. The error bars show the accuracy of each temperature measurement.

....

When we examine all these measurements of the intensity of radiation coming to us from different directions in the sky, we learn a number of striking things about the structure of the universe. We find that it is expanding at the same rate in every direction to an accuracy better than one part in a thousand. We say that the expansion is isotropic—that is, the same in every direction. If one had been picking possible universes at random from some cosmic menagerie, there would be countless varieties that expand far faster in some directions than others, or that rotate at high speed, or even that contract in some directions while expanding in others. Yet our universe is peculiar: it seems to be in an improbably well-ordered state, in which the expansion proceeds at the same rate in every direction to high precision. It is as if you were to find all your children's bedrooms perfectly tidy—a highly unlikely state of affairs. You decide that some outside influence must have been exerted. Likewise, there must be some explanation for the striking isotropy of the expansion.

Cosmologists have long regarded the isotropy of the universe's expansion as a great mystery that must be explained. The approaches taken to it illustrate something of the styles of thinking within the subject. The first line that could be taken is to say that the universe began expanding isotropically from the outset and that the present state is just a reflection of its special starting conditions. Things are as they are because they were as they were. As it stands, this is not very helpful. It doesn't explain anything. It's like invoking the tooth fairy. But it could, of course, be true. If so, we might hope to find some deep "principle" calling for an initial state of isotropic expansion. Such a principle might have other, more local applications by which it could reveal itself. The unsavory feature of this approach is that it places the onus for explaining the present state of the universe entirely upon its unknown (and perhaps unknowable) starting state.

The second approach is to regard the present state of affairs as a consequence of physical processes still going on in the universe. Thus perhaps no matter how irregular was its initial state, after billions of years the irregularities all get washed out, leaving a state of isotropic expansion. This approach has the merit of suggesting possible research programs: Are there cosmic processes that can smooth out nonuniformities in the expansion? How long does the smoothing take? Can these processes get rid of any amount of irregularity by the present day, or can they eradicate only a small amount? This approach allows us to say that no matter how the universe began, there are processes inevitably arising within it during its early history which insure that after fifteen billion years of expansion it looks pretty much the way it does today.

Although the second philosophy sounds wonderfully appealing, it does have a downside. If we succeed in showing that the present state of the universe emerges regardless of the starting conditions, then our observations of its structure will not be able to tell us anything about those starting conditions—for the present state would be compatible with any starting state. But if, on the contrary, the present structure of the universe—its expansion isotropy, and the patterns displayed by the clustering of galaxies—are partial reflections of the way the universe began, then it might be possible to determine something about the initial state of the universe by observing it today.

CHAPTER 2

THE GREAT UNIVERSAL CATALOG

> All other men are specialists, but his specialism is omniscience.
> —*The Bruce-Partington Plans*

When Einstein published his general theory of relativity in 1915, there was no widespread belief that the universe was populated by those huge collections of stars we know as galaxies. It was commonly held that these extraterrestrial sources of light—or "nebulae," as they were then called—lay within our own Milky Way galaxy. Nor had there ever been any proposal by astronomers or philosophers that the starry universe was anything but static. It was into this intellectual ambience that Einstein launched his new theory of gravitation. Unlike Newton's classical description of gravitational forces, which Einstein's theory included and superseded, the general theory of relativity had the extraordinary ability to describe entire universes, even if they were infinite in extent. Only the simplest of solutions to Einstein's equations have ever been found. Fortunately, the very simple ones describe rather well the universe we see.

When Einstein began to explore what his new equations revealed about the universe, he set about doing what scien-