CHAPTER 1

Why do Many Other Scientists Believe Time Began at a Big Bang?

Our everyday perception of the universe comes from looking up at the sky to see the Sun in the daytime and, more profoundly, to see thousands of stars in the night sky. Surely some of the oldest questions since the beginnings of human thought are: How large is the universe? Did it ever begin? What are the principal constituents of the present universe? Will time ever end?

Cosmology is the name for the scientific study of the universe. The present time is an unprecedented age for cosmology because it is fair to say that in the last twenty years we have learned more in cosmology than in all of previous human history. Despite this enormous and exciting growth of our knowledge as a result of many impressive observations, the universe has become more enigmatic in many ways. The more we learn, the more the extent of our ignorance becomes manifest.

DID TIME BEGIN? WILL TIME END? - Maybe the Big Bang Never Occurred

Cosmology has recently answered some of the old questions and in this chapter we shall give answers to the first two: How large is the universe? How long ago did it — at least the present expansion era — begin? We can all agree that the expansion stage we are presently in began a finite time ago but, as I shall explain later in this chapter, it is not obvious that time itself began then, if ever.

We do know how large the visible universe is, meaning how far away the most distant galaxies are whose light can reach us on the Earth. It is theoretically possible, and even favored in some theoretical scenarios, that our universe is actually much larger than the visible universe. In some very speculative scenarios the universe is spatially finite with non-trivial topology. This is at present not readily testable so we shall be content to try to convey just how gigantic the visible part is.

The observational means by which we know accurately the size of the universe will not concern us here but it is sufficient to say that present studies using the Hubble Space Telescope combined with the largest (up to 10 meters in diameter) ground-based optical telescopes tell us the size of the visible galaxy to an accuracy of a few per cent. This sort of accuracy has been achieved only since the turn of the 21st century.

* * *

Cosmological distances are so much bigger than any distance with which we may be familiar, it is not easy to grasp or comprehend them even in our imagination. So let us begin with the largest distance which is easily comprehensible from the viewpoint of our experience.

A very long airplane ride may take 15 hours and go 9000 miles, a significant part of halfway around the Earth. People who travel a lot may take such a flight a few times each year. One knows that the plane has a ground speed of about 600 miles per hour and the discomfort of sitting, especially in economy class, for such a long time gives a strong impression of just how far that distance is. Of course, people a hundred years ago would never travel that far in a day but now we do and it gives us a feel for the size of the planet so that makes it a length distance from which we can begin.

The next larger distance to think about is the distance between the Earth and the Moon. This is about thirty times the distance of the plane ride and so it would take equivalently some three weeks at the same airplane speed, or a few days in a NASA spacecraft. The distance to the Moon is thus imaginable: if you walked at four miles an hour non-stop without sleep it would take about eight years to arrive and another eight years to return. Nevertheless, the arrival on the Moon of astronauts Armstrong and Aldrin in July 1969 was one of the most memorable events of the last century. Only in part was it due to the distance to the Moon, it was equally due to the concept of humans walking for the first time on an astronomical object other than the Earth.

The Moon is visible in the night sky, and just as often present in the daytime thanks to its reflection of light from the Sun. The Sun is by far our nearest star and its radiated energy is crucial to the possibility of life on Earth. How far away is the Sun? It is about ten thousand times the length of the airplane ride and would take about twenty years to reach at the speed of an airplane. Not that any sane person would want to go there with a surface temperature well above that of molten iron. The Sun is about four hundred times further away than the Moon, and is already at such a large distance that it far exceeds anything with which we are familiar. This sets the scale of the Solar System with the Earth, rotating on its axis once a day, orbiting once a year around the Sun at a distance of some ninety-three million miles. Other planets like Mercury and Venus, circulate inside the Earth's orbit while six others including Mars, Jupiter and Saturn orbit outside the Earth.

It is almost inconceivable that any human being will travel outside of the Solar System in our lifetimes because of its enormous size. Yet on the scale of the visible universe, the Solar System is, in contrast, unimaginably tiny and insignificant. If there were no life other than on the Earth the universe would seem to be an absurdly large object if life were its primary goal.

In addition to the Moon and some planets, we can see thousands of stars with the naked eye. Most of these stars are similar to our Sun but appear much dimmer because of their distance from us. How distant are even the nearest stars? The answer is some two million times the distance to the Sun. So whereas we can reach the Sun in twenty years at the speed of an airplane, to reach the nearest star in twenty years would require a two million times faster speed. A quick calculation shows that this takes six hundred miles per hour into thirty-five thousand miles per second. To put such a speed into perspective, the speed of light is about one hundred and eighty thousand miles per second. This means our imaginary airplane, suitably coverted as a spacecraft, must travel at one fifth of the speed of light just to reach the nearest star in twenty years.

Here we see the limitations to any travel possibilities not only in our lifetime but what would seem to be forever. According to the theory of relativity, which there is no reason to doubt, nothing can travel faster than the speed of light. So even if the human lifetime is extended by medical advances to two hundred years or even a thousand years, it is impossible to travel during one lifetime to more than a few hundred times the distance to the nearest star. But the galaxy to which our Solar System belongs extends about ten thousand times the distance to the nearest star. So it would seem impossible ever to leave our particular galaxy — known as the Milky Way from its appearance spreading across the night sky.

There are however a couple of holes in this argument. First, according to relativity, time slows down as one travels when approaching the speed of light. Second, it is conceivable that some cryogenic method might be devised to slow down the speed of life and greatly enhance the effective human lifetime. Even so, to travel outside our galaxy does seem forever impossible and cosmology may remain just a spectator sport.

One hundred years ago it was generally believed that the the universe was comprised of only the Milky Way. The size of our galaxy is only ten thousand times the distance to the nearest star and already that is two hundred thousand times the distance to the Sun. Therefore, the galaxy size is two billion times (one billion is a thousand million) the Earth-Sun distance. The size of the galaxy seems to be relatively independent of time and so in ignorance of a universe very much bigger than a single galaxy, it was believed before the 1920s that the universe itself was static, neither expanding nor contracting.

When the General Theory of Relativity was proposed in 1915, this state of the observational knowledge stymied what could have been predicted, namely, the overall expansion of the universe. This expansion, which is a key feature of the universe and will lead us to the conclusion that it had a definite beginning, became an option only by observations somewhat later during the 1920s.

Now we arrive at the final leap in the distance scale. The visible universe turns out to be about four hundred thousand times the size of the Milky Way, very much larger than previously imagined. That is, not only is the Solar System of neglible size with respect to the universe but so is the entire Milky Way. In fact, in theoretical cosmology galaxies are treated as point particles. And the human race may be confined forever to be within one of these points!

We have seen that the size of the galaxy is tremendously larger, by a factor of billions, than the distance to the Sun. And then the visible universe is yet again so much larger than a galaxy that to study it each galaxy may be regarded as just a single dot within it. This should communicate in so many words an idea of just how big the visible universe is. Now we will show how we know that the present expansion (and possibly time itself) had a beginning some fourteen billion years ago.

* * * *

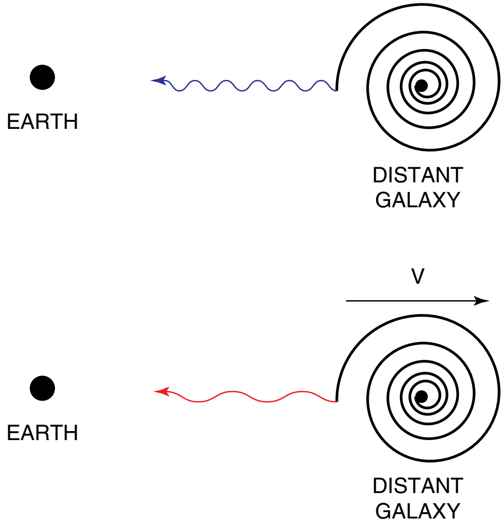
As has already been discussed, the size of the Milky Way has not expanded by even one order of magnitude since it was formed some ten billion years ago. Within the Milky Way the Sun and the Solar System appeared about five billion years ago. The Earth is a little younger, about four and a half billion years old. The point is the general arrangement of the Sun and planets in the Solar System has not markedly changed in the last few billion years. During that time we may regard the galaxy and its contents being of a constant size.

A truly astonishing revelation comes when we study the same question for the entire universe, including hundreds of thousands of the billions of galaxies outside of the Milky Way. The issue is: what is their typical motion relative to our galaxy?

Here it is important to understand a phenomenon well known in physics called the Doppler effect. It is a more familiar phenomenon for sound waves than for light. When a train blows its whistle and passes a listener the pitch of the whistle falls from a higher note to a lower note. In fact, not only the whistle but the entire train's noise exhibits the same Doppler effect. Why does this happen? It is because the motion of the train towards the listener compresses the sound waves to become a shorter wavelength and, because the velocity of sound is unaltered, to a higher frequency. Similarly, when the train is moving away the sound waves are stretched and the frequency lowers. The pitch for a stationary train would lie between the two pitches of the train approaching and receding. The shift in frequency is calculable simply in terms of the ratio of the speed of the train and the speed of sound.

* * *

Exactly the same Doppler effect occurs for light. If a galaxy is approaching our galaxy, its light appears with a higher frequency. For the visible spectrum the highest frequency is for blue light so we may say that the light is blue-shifted. On the other hand, if the galaxy is receding from ours its light appears shifted to a lower frequency and



DOPPLER EFFECT: Wavelength of light from distant galaxy lengthened (red shifted) by speed of recessan from Earth is red-shifted toward the red or lowest frequency end of the visible spectrum.

In fact, what are observed are the spectral lines of light emitted from known atoms and whose frequencies are accurately known here on Earth. If all the lines are systematically shifted towards the blue then the galaxy hosting these atoms is approaching the Milky Way: if toward the red then it is receding from us. This can be made precise by a mathematical expression for the frequency shift which gives the approach or recession speed as a fraction of the speed of light.

When this is studied for a large number of galaxies it might be expected that roughly half would be blue-shifted and half red-shifted if the overall universe were static and the galaxies were moving randomly.

What was observed, however, to the astonishment of Hubble and Einstein in 1929 is that almost all galaxies are red-shifted. Apart from a few galaxies in our immediate neighborhood like the nearby Andromeda galaxy all the hundreds of thousands of more distant galaxies measured are receding from us. This means that the entire universe is expanding and the galaxies are moving away from us and, as we shall see later, from each other. This phenomenon is called the Hubble expansion.

The next important question is: how does the recession velocity depend on the distance of the galaxy from the Milky Way? The galaxies can be classified into types such as spiral, elliptical or irregular and the total light emitted may be assumed to show regularity within each type. But their apparent brightness on Earth depends on the distance and falls off as an inverse square law. So the apparent brightness can be converted into a distance. When the recession velocity is compared to the distance a very important regularity appears. This is a most significant discovery in cosmology. It is called Hubble's law, which states that the recession velocity is proportional to the distance.

The ratio of the velocity and the distance is thus a constant, i.e., the Hubble parameter. Its value is notoriously difficult to measure but now we do know it to be close to seventy, within ten percent, in certain units. These units, which are not crucial to the general discussion, are kilometers per second per megaparsec, where a megaparsec is the distance light travels in about three million years.

How does this tell us when the present expansion phase began? This requires the use of the equations of the General Theory of Relativity together with two assumptions. The first assumption is that the universe at the large scale is the same, on average, in each of the three directions of space. This is called isotropy and is supported by observations which indicate that no preferred direction exists in the universe. The second assumption is that, on average, all positions in the universe are equivalent. This means that in all galaxies (hypothetical!) observers would see the recession of other galaxies according to Hubble's law. This assumption is called homogeneity. The combination of these two assumptions, isotropy and homogeneity, is technically known as the cosmological principle.

Combining the cosmological principle with the general theory of relativity gives rise to a mathematical equation known after its inventor as the Friedmann equation which characterizes the expansion of the universe in terms of a scale factor which is a function of time and specifies the typical distance between galaxies. Inserting the known Hubble parameter and the present known composition of the universe then enables us to calculate the scale for all past times. Note, in passing, that we cannot do this for future times with confidence because we do not know with certainty how the composition of the universe will evolve in the future. For the past we have good confidence and we find a striking conclusion: run in reverse, the contracting universe is seen mathematically to shrink to a point at a well-defined past time.

At that time the universe may have begun in some unimaginably powerful explosion called the Big Bang. Our present expansion seems to have a beginning and we know when. It was some 13.7 billion years ago, give or take two hundred million years. The age of the universe is now established to an accuracy of about two percent.

* * *

There is one serious problem with extrapolating the Friedmann equation back in time, namely that about 13.7 billion years ago the equation becomes singular. The density and temperature become infinite and the classical theory of general relativity breaks down. Thus the Big Bang, only much more recently named, was the *initial singularity* to the early workers of the 1920s and 1930s.

Within the theory, we know therefore that the Big Bang must be avoided. A common response is to invoke quantum mechanics. From the fundamental constants, the speed of light, Newton's gravitational constant and Planck's constant one can construct a time known as the Planck time which is a tiny fraction (10^{-44}) of a second. One may say that at that short time after the would-be Big Bang the classical theory of general relativity must break down. Quantum gravity must play a role but till today, no completely satisfactory theory exists. So one may argue that quantum mechanics rescues the day. Indeed an entire field known as quantum cosmology has been built up around such an idea.

In this book we shall not appeal to quantum mechanics in this way but examine whether the Big Bang can be avoided in a purely classical context.

* * *

Before proceeding there is one amusing anecdote about the origin of the graphic term Big Bang which seems apt to describe the beginning of the universe. Before the scenario we have just described was firmly established, a competing theory was the steady-state theory which postulated that despite the Hubble expansion there was a steady state and no beginning because of the continuous spontaneous creation of new galaxies. As a derogatory term for the competing theory, Big Bang was coined by a leading exponent of the steady-state theory. Unfortunately for Hoyle, it was the Big Bang theory and not his viewpoint which was confirmed by subsequent measurements.

* * * *

There is one alternative view (we shall discuss another view in Chapter 7) of the Big Bang where the lifetime of the universe is also infinite. The process of the Big Bang is in that view something which has occurred repeatedly, indeed infinitely, resulting in an infinite number of different universes of which we are aware of just one. This is technically called eternal inflation and the resultant universe becomes a multiverse for obvious reasons.

It is difficult, if not impossible, to test eternal inflation because the other universes would seem to be forever hidden from our view. Our best chance may be to make a probabilistic treatment of the multiverse to estimate the probability of the universe we observe having the properties it has in terms of its fundamental consituents or building blocks. Some research is indeed being pursued along this line.

* * * *

In this book we shall assume the beginning of the present expansion era to have taken place approximately fourteen billion years ago and followed, as will be discussed, by an inflationary era of rapid expansion. The latter explains two different kinds of extraordinary smoothness observed. We know that there was temperature uniformity to one part in one hundred thousand in the universe when it was only four thousand years old. Then there is the proximity of the observed density of the universe to a special value known as the critical density which would, without inflation, require preternatural fine-tuning in the early universe.

Inflation appears now to be ubiquitous in almost all theoretical cosmology, in one form or another. As we shall discuss, it can account for the otherwise puzzling smoothness properties of the universe. On the other hand, it is exceedingly difficult to make direct measurements which are sensitive to such an early era, the inflationary era, which occurred even earlier than a billionth of a second after the would-be Big Bang.

Normal observations involving electromagnetic radiation go back only to a few hundred thousand years after the would-be Big Bang, far too recent for us to study inflation. Studies of abundances of light elements like helium and hydrogen probe indirectly back to a cosmic time of one second after the would-be Big Bang. Potential neutrino astronomy measurements could directly probe a similar era.

The only chance of direct observation of the inflationary era would appear to be by gravitational radiation — waves created by strong gravitational fields in the early universe. The observability of such radiation depends on how early inflation took place, the earlier being the easier to detect. For later inflation it looks presently impossible to detect such gravity waves. The word "presently" is essential since how technology will evolve, and what consequent scientific apparatus will be enabled by the end of the 21st century is impossible to predict. It is a lesson from the history of physics that to decree anything to be impossible is a dangerous prediction.

* * *

One may ask what happened before the Big Bang, if it did occur? This is beyond scientific investigation and it is easier to assume that time began then. Very ambitious and speculative theories discuss prior times using ideas such as T-duality in string theory or eternal inflation with its resultant multiverse. If such theories become testable and shed light on the physics of our universe, then they must be taken very seriously in a more general domain of applicability. At present, such ideas remain speculation.

Another question which we shall address at length in this book is what will happen to the universe in the future? This is less well-understood than the past, and depends critically on the properties of the newly-discovered Dark Energy which comprises almost three quarters of the total energy density of the universe.

Concerning space one will ask whether it too, like past time, will be finite in extent. Is it possible that by proceeding in a straight line one will return, after a finite time and distance, to the starting point due to a non-trivial topology of space? There is no compelling evidence for this possibility although certain data on the cosmic microwave background radiation can be interpreted as supporting such an assumption. Alternative explanations for the data come from arguments about cosmic variance or from small distortions in the hypothetical inflation potential, so the case for non-trivial spatial topology is not strong. If there were non-trivial topology, it could be of one of three types. Positive curvature corresponds to a closed universe, negative curvature to an open universe and a flat geometry, without curvature, as predicted by inflation. The local properties in such a universe satisfy the same general relativity equations as for the case of infinite space with trivial topology; only the global toplogical properties differ so it is not obvious from, say, the study of our galaxy alone which option Nature chooses. The notion of non-trivial topology of space necessarily introduces at least one fundamental length which, to be consistent with observational data, must be comparable to the radius of the visible horizon of a few gigaparsecs.

A common question by an educated non-physicist is: into what does the universe expand or, equivalently, what is "outside" the universe? So let us try to give a clear answer. The answer is not obvious only because of the limitations to the human imagination. All of us can easily imagine three spatial dimensions but four is exceedingly difficult. Unfortunately, the spacetime manifold of the universe is itself four dimensional and this is both why the question naturally arises and why the answer is slightly elusive. If we scale down by one dimension there is an analogous situation which is, by contrast, very easy to grasp. Take a balloon with spots on the surface to represent galaxies. As time passes we inflate the balloon and the spots get further apart as for the expansion of the universe. Now a two-dimensional being on the balloon surface may ask: into what is this two dimensional space expanding? The answer is that there is nothing "outside" the two-dimensional surface as obviously it is a closed surface without a boundary. Similarly the three-dimensional space of our universe has no boundary and no "outside".

* * * *

Among so many interesting yet unanswered questions, the one about the beginning of the present expansion 13.7 billion years ago seems settled. The expansion itself was universally accepted only in 1965 as a result of the discovery of the remnant background microwave radiation. The uncertainty of the future scenario for the universe is under much study as a result of the discovery of dark energy dating from 1998. Thus we are at a very exciting time in the subject.

The establishment of a finite time of the present expansion of about fourteen billion years is clearly of fundamental importance which could be equated by establishment of a finite spatial extent to the universe. There is absolutely no compelling evidence for such an idea although some data from the WMAP analysis of the cosmic background radiation, particularly the unexpectedly small values of the low multipoles, have been interpreted as suggestive of finite size and non-trivial topology.

Certainly these cosmological discoveries change our picture of our own history.

Finally, in our discussion of the universe's longevity, it is important that we use a linear time, rather than logarithmic time, in the above discussion. The two are dramatically different. Firstly, in logarithmic time the age becomes infinite. But the difference can be better seen in a concrete analogy.

Suppose we condense the entire cosmic history of fourteen billion years into one day of twenty-four hours starting and ending at midnight. First we use linear time. The nucleosynthesis takes place just a trillionth of a second after midnight; recombination and the surface of last scatter are three seconds later; galaxy formation starts around 1.40 am; the Sun is created about 4 pm in the afternoon and Julius Caesar invades Gaul about a hundredth of a second before midnight.

But if we map the same history using logarithmic time starting at the Planck time (since we must now start at a finite time in the past) then the occurrence of major events looks completely different.

Nucleosynthesis waits until 5 pm in the afternoon; recombination and the surface of last scatter are at 10 pm in the evening; galaxy formation begins at 11.25 pm; the Sun is created at 11.48 pm and Julius Caesar appears now only a trillionth of a second before midnight.

This illustrates how the use of linear time in cosmology effects the relative spacing of subsequent events. From the viewpoint of fundamental physics more happened in the first second of the Big Bang than has happened in the subsequent fourteen billion years, more in parallel with a logarithmic picture of time. But it is in linear time, with which we are familiar in measuring all everyday events, that the time since the would-be Big Bang does have the finite value of 13.7 billion years.

* * *

The answer to the question in this chapter's title has already been alluded to, that many other scientists (if not this author) explain away the initial singularity of the Friedmann equation by an appeal to quantum mechanics and quantum gravity. The absence of a fully satisfactory theory of quantum gravity can act as a further security for such scientists as no one can definitively refute the argument.

However, the tentative attempts at quantum gravity have problems. The concept of the wave function of the universe, as employed in quantum cosmology, is problematic for several reasons, not least of which is that the observer is inside the system. The Planck time is much shorter than the time expected to pass between the would-be Big Bang and the onset of inflation. Explaining away the initial singularity by quantum mechanical arguments was useful only when it was *faute de mieux*. This is no longer the case.

I find it more satisfactory to make a cosmological model where the density and temperature are never infinite. This precludes a Big Bang and replaces it with a different picture of time where time never begins and never ends. This is in contrast with the standard cosmological model where time begins at the Big Bang and never ends during an infinite future expansion.

While I cannot prove rigorously that the conventional wisdom is wrong, it does entail the singularities and concomitant breakdown of general relativity that we have mentioned. The existence of plausible alternatives now, however, makes the Big Bang idea less plausible.

As we shall show later in the book, there is an alternative version where time begins at a Big Bang and ends in a "Big Rip" at a finite time in the future. I regard this as preferable aesthetically to the conventional picture. But best of three possibilities about time is the "infinite in both directions, past and future" as exemplified by a cyclic model that my student and I constructed only in the twenty-first century based on the dark energy component. Observations of dark energy and its properties, especially its equation of state, will confirm or refute such more satisfactory ideas about time which insist there was never a Big Bang.