8 The Evolution of the Universe

MARTIN REES

Cosmologists study evolution on the grandest scale of all. They aim to set our earth and our solar system in an evolutionary scheme stretching right back to the formation of the Milky Way galaxy – right back even to a so-called 'Big Bang' that set our entire observable universe expanding and imprinted the physical laws that govern it.

Evolution within our galaxy

Let us start with something fairly well understood – the life cycle of our sun, a typical star. About 4.5 billion years ago it condensed from an interstellar cloud, and contracted until the centre became hot enough to ignite fusion of hydrogen into helium. This process will keep it shining until, after another five billion years, the hydrogen runs out. The sun will then flare up, becoming large enough to engulf the inner planets, and to vaporize all life on earth. After this 'red giant' phase the inner regions contract into a white dwarf – a dense star no larger than the earth, though nearly a million times more massive.

We are quite confident about these calculations because the relevant physics has been well studied in the laboratory – atomic and nuclear physics, Newtonian gravity and so forth. Astrophysicists can just as easily compute the life cycles of stars with half the sun's mass, or twice, four times, etc. Heavier stars burn brighter, and trace out their life cycle more quickly.

Stars live so long compared to astronomers that we are granted just a single 'snapshot' of each star's life. But we can test our theories, by looking at the whole *population* of stars. Trees can live for hundreds of years. But for a newly landed Martian who had never seen a tree before, it would take no more than an afternoon wandering around in a forest to deduce the life cycle of trees: from looking at saplings, fully grown specimens and some which had died.

In the Orion Nebula, for instance, new stars are even now condensing within glowing gas clouds. The best 'test beds' for checking such calculations are globular clusters – swarms of a million different stars, held together by their mutual gravity, which all formed at the same time.

But not everything in the cosmos happens slowly; sometimes stars explode catastrophically as supernovae. The closest supernova of the twentieth century occurred in 1987. Its sudden brightening and gradual fading have been followed not only by optical astronomers (Figure 1) but by those using the other modern techniques – radio, X-ray and gamma-ray telescopes – which have opened new 'windows' on the universe.

In about 1000 years, it will look like the Crab Nebula (Figure 2), the relic of a supernova witnessed and recorded by Chinese astronomers in A.D. 1054. Now, nearly a thousand years later, we see the expanding debris from the explosion. The Crab Nebula will remain visible, gradually expanding and fading, for a few thousand years; it will then become so diffuse that it merges with the very dilute gas and dust that pervades interstellar space.

Cosmic alchemy

Supernovae fascinate astronomers, but why should anyone else care about explosions thousands of light years away? Because, were it not for supernovae, there would be no planets, still less any complex evolution on them.

Of the ninety-two chemical elements that occur naturally, some are vastly more common than others. For every ten atoms of carbon, you would find, on average, twenty of oxygen, and about five each of nitrogen and iron. But gold is 100 million times rarer than oxygen, and others – uranium, for instance – are rarer still. Why are carbon and oxygen common, but gold and uranium so rare? This everyday question is not unanswerable – but the answer involves ancient stars that exploded in our Milky Way more than five billion years ago, before our solar sytem formed.

Stars much heavier than the sun evolve in a more complicated and dramatic way. After they have used up their central hydrogen (and turned into helium) gravity squeezes them further. Their centres get still hotter, until helium atoms can themselves stick together to make the nuclei of heavier atoms: carbon (six protons), oxygen (eight protons) and iron (26 protons). A kind of 'onion skin' structure develops: where the hotter inner layers have been transmuted further up the periodic table.

When their fuel has all been consumed (when their hot centres are

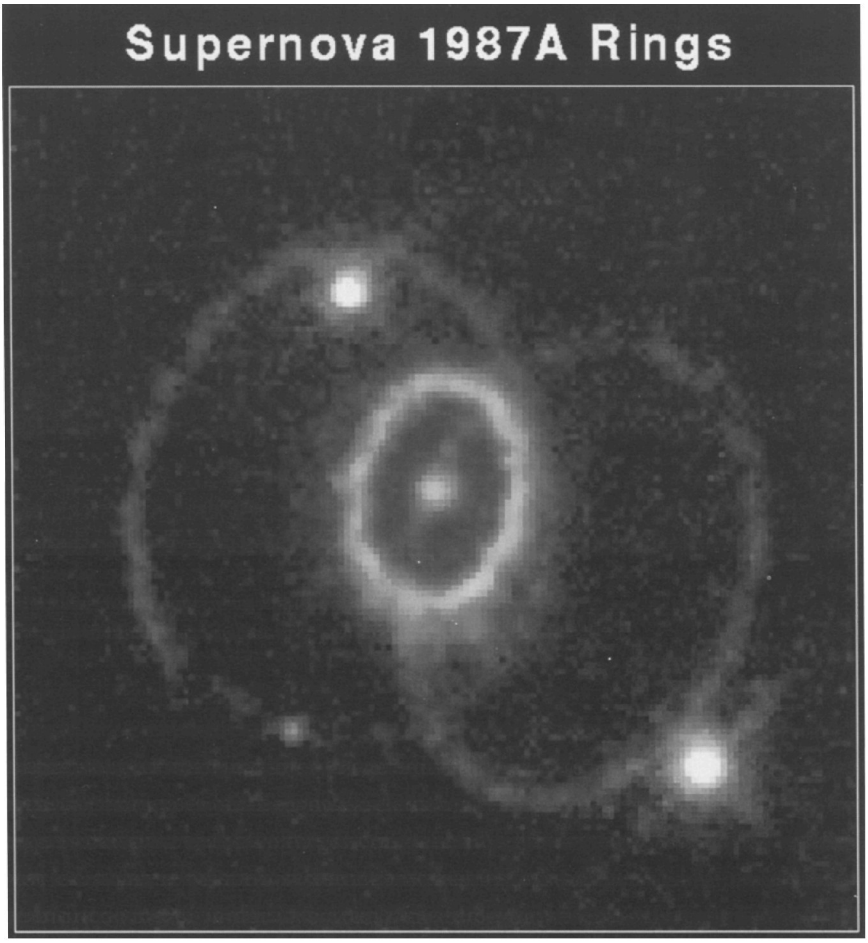


FIGURE 1 The supernova in the Large Magellanic Cloud, at a distance of about 160 000 light years. This picture, taken with the Hubble Space Telescope about five years after the actual explosion was observed, shows strange 'rings', which are thought to result from interaction between the radiation and debris from the explosion and external material that was probably ejected as a slower wind before the supernova occurred.

transmuted into iron) big stars face a crisis. A catastrophic infall squeezes their centres to the density of an atomic nucleus, triggering an explosion that blows off the outer layers. This explosion manifests itself as a supernova of the kind that created the Crab Nebula. The debris contains the outcome of all the

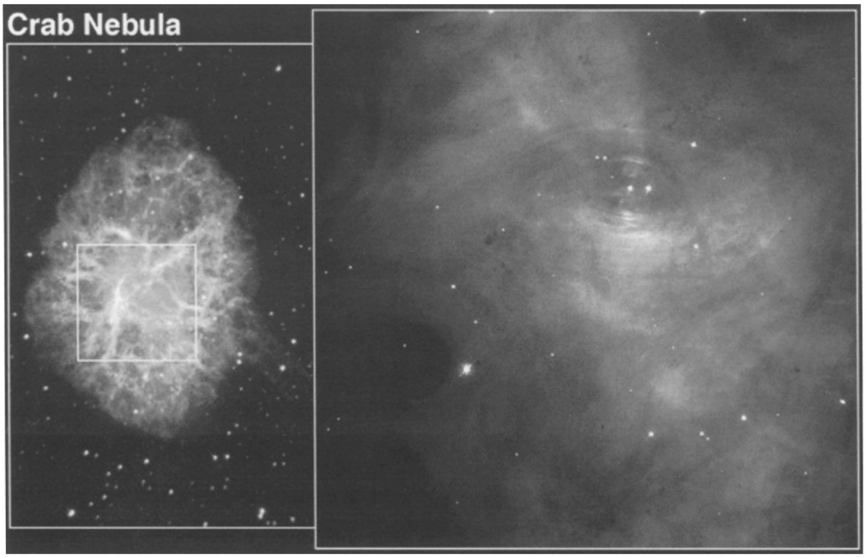


FIGURE 2 The Crab Nebula. The right-hand frame shows continuing activity in the central part of the nebula, induced by the spinning neutron star (pulsar) that was left behind after the explosion.

nuclear alchemy that kept the star shining over its entire lifetime – a great deal of oxygen and carbon, plus traces of many other elements. The calculated 'mix' is gratifyingly close to the proportions now observed in our solar system.

The Milky Way, our home galaxy, resembles a vast ecosystem. Pristine hydrogen is transmuted, inside stars, into the basic building blocks of life: carbon, oxygen, iron and the rest. Some of this material returns to interstellar space, thereafter to be recycled into new generations of stars.

A carbon atom, forged in an early supernova, might wander for hundreds of millions of years in interstellar space. It might then have found itself in a dense interstellar cloud, which collapsed under its own gravity to form stars. It may have entered the core of some new massive star, where it is processed further up the periodic table (into silicon, or into iron), and then flung out in another supernova. Or it may have joined one of the less massive stars, each surrounded by a spinning gaseous disc that condenses into a retinue of planets. One such star could have been our sun. The same carbon atom may have found itself in the newly forming earth, perhaps eventually in a human cell. Each atom has a

https://doi.org/10.1017/CBO9780511629044.009 Published online by Cambridge University Press

pedigree extending back far earlier than our solar system's birth. We are literally the ashes of long-dead stars.

But how did our galaxy itself emerge? Where did the basic hydrogen come from? To answer these questions we must broaden our horizons still further in both space and time to the universe of galaxies.

The visible universe

Galaxies are held in equilibrium by a balance between two effects: gravity, which tends to make the stars all gather together; and the countervailing effect of the stellar motions, which if gravity did not act would make a galaxy fly apart. In some galaxies, our own and Andromeda among them, 100 billion stars move in nearly circular orbits in discs. In others, the less photogenic ellipticals, stars are swarming around in more random directions, each feeling the gravitional pull of all the others. Galaxies are not as well understood as stars. Indeed, as I will explain, we do not even know of what they are primarily made.

Galaxies interest cosmologists because they are 'test particles' for probing structure and motions in the large-scale universe. The nearest few thousand galaxies – those closer than about 300 million light years – have been mapped in both hemispheres. They are irregularly distributed into clusters and superclusters. Are there, you may ask, clusters of clusters of clusters *ad infinitum*? Our universe is not like that. If it were, we would see conspicuous clumps in the sky, however deep into space we probed. Although the nearest few thousand galaxies are conspicuously clumped, the brightest *million* galaxies are actually fairly uniform over the sky; as we look at still fainter galaxies, probing still greater distances, clustering becomes less evident and the sky appears smoother.

There is, in other words, a well-defined sense in which the universe is broadly homogeneous. A terrestrial analogy may clarify this. The ocean displays complex patterns – waves (sometimes small riding on large), foam, etc. But once your gaze extends beyond the scale of the longest ocean swells, you see an overall uniformity, stretching to the horizon many miles away. A patch of ocean large enough to be 'typical' must obviously extend several times further than the scale of the longest waves. Our horizon extends far enough to encompass many patches statistically similar one to another, each large enough to constitute a 'fair sample'.

This broad-brush uniformity of *seascapes* is not, however, a general feature of *landscapes*: on land, progressively larger mountain peaks may stretch all the way to the horizon, and a single topographical feature may dominate the entire view. Cosmology is, by definition, the study of the entire universe. We can see only one universe – probably, indeed, only a tiny part of everything there is. Despite these limitations, scientific cosmology *has* progressed, but only because our observable universe (the volume out to the 'horizon' of our observation) resembles a seascape rather than a mountain landscape. Even the biggest superclusters are still small in comparison with the range of powerful telescopes.

The overall motions in our universe are simple too, as Edwin Hubble first realized. Distant galaxies recede from us with a speed proportional to their distance, as though they all started off packed together ten to fifteen billion years ago.

Far away towards the horizon, we see domains whose light set out when the universe was more compressed, more closely packed together (Figure 3). Astronomers can actually *see* the remote past. Telescope images reveal huge numbers of very faint galaxies each so far away that its light set out before our solar system formed. Even more remote are the *quasars* – hyperactive centres of a special class of galaxies, so bright that they vastly outshine the 100 billion stars in their host galaxy. The 'distance record' is held by a quasar so red-shifted that the Lyman-alpha 1216 Å line, the strongest feature in the spectrum of hydrogen and in the far ultra-violet, reaches us in the red part of the spectrum, at around 7200 Å. The ratio of the observed to the emitted wavelength, 5.89, is the factor by which the universe has expanded since the light set out.

Evidence for a Big Bang

Quasars are probes of the era when galaxies were young, and perhaps just forming. But what about still earlier epochs? Did everything really start with a so-called 'Big Bang'? The idea goes back to the Belgian Catholic priest Georges Lemaître in 1931. The phrase itself was introduced by Fred Hoyle, as a derisive description of a theory he never liked.

The name has stuck, and the clinching evidence for the theory came in 1965, when Arno Penzias and Robert Wilson found excess microwave noise, coming equally from all directions and with no obvious source, in their antenna at Bell Laboratories in New York. This has momentous implications: intergalactic space is not completely cold; it is about 3 K above absolute zero (0 K, −273 °C).

FIGURE 3 This picture shows a very deep exposure taken with the Hubble Space Telescope. Although it images only a small patch of sky – about a thousandth the area covered by the full moon – it reveals hundreds of very faint objects. Most of these are galaxies so far away that their light set out when they had only recently formed.

That may not seem much, but there are about a billion quanta of radiation (photons) for every atom in the universe.

This 'cosmic background' causes some 1% of the background 'fuzz' on a television set. It is an 'afterglow' of a pregalactic era when the entire universe was hot and dense and opaque. After the universe had expanded for about half a million years, the temperature fell below 3000 K; the primordial radiation then shifted into the infra-red. The universe literally entered a dark age, which persisted until the first stars in the first galaxies, and maybe also the first quasars, formed and lit up space again. The expansion cooled and diluted the radiation, and stretched its wavelength. But it would still be around – it fills the universe and has nowhere else to go!

But we have got firm grounds for believing that the temperature was once billions of degrees, not just thousands – hot enough for nuclear reactions. The rapid expansion did not allow enough time for everything to be processed into iron, as in hot stars. However, about 25% would turn into helium. The rest would still be hydrogen apart from traces of deuterium and lithium.

What is remarkable is that the proportion of helium in old stars and nebulae, now pinned down with 1% accuracy, turns out to be just about what has been calculated. As a bonus, so are the proportions of lithium and deuterium as well. Moreover, these particular elements were a problem for the stellar nucleogenesis scenario that was so successful for carbon, oxygen, etc. This corroborates an extrapolation right back to when the universe was hot enough for nuclear reactions to occur, i.e. when it was just a *few seconds old.*

One day in 1992, what was then in my opinion the best British daily paper (*The Independent*) heralded a cosmological discovery with this front page (Figure 4); it is a comprehensive depiction of cosmic evolution. (Even dinosaurs are featured; I suppose they are the only scientific topic that matches cosmology in popular appeal.) It goes right back to 10^{-43} second. This is what is called the Planck time, when everything was so dense that quantum fluctuations were important for the entire universe.

So, can you believe all the cosmology reported in the newspapers? Is the universe indeed evolving as depicted here? Over the last few years, the case for a 'Big Bang' has had several boosts: the COBE (Cosmic Background Explorer) satellite showed that the background radiation had the expected spectrum, to a precision of a part in 10 000, and there have been better measurements of cosmic helium and deuterium. Moreover, there are several discoveries that *might* have been made, which would have invalidated the hypothesis, and which have *not* been made. The Big Bang has lived dangerously for twenty-five years, and survived.

The grounds for extrapolating back to the stage when the universe had been expanding for *a second* (when the helium formed) deserve to be taken as seriously as, for instance, ideas about the early history of our earth, which are based

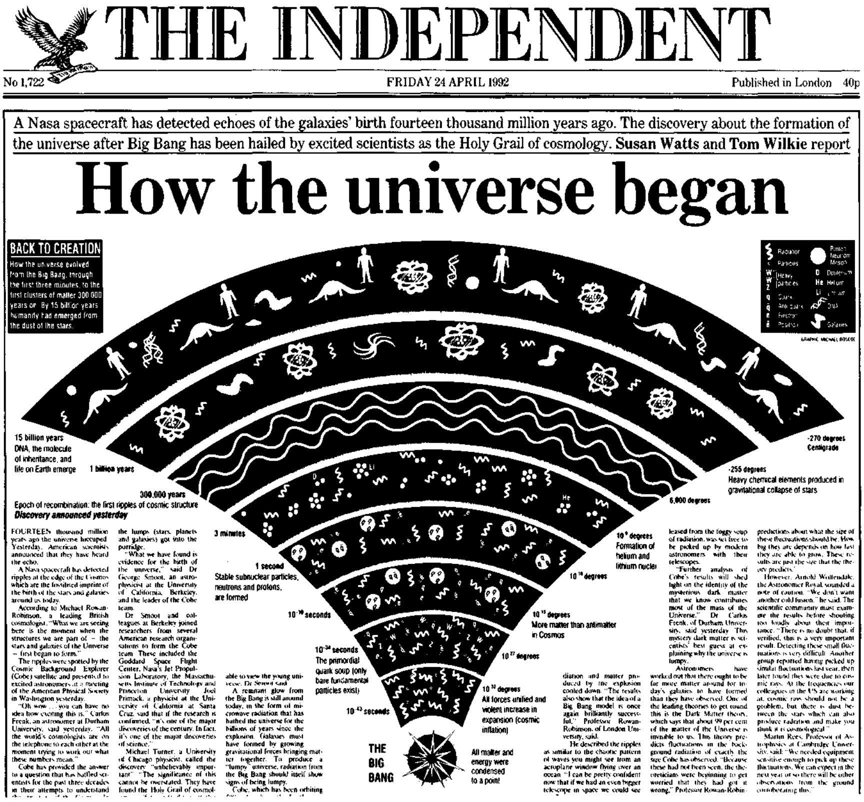


FIGURE 4 The front page of *The Independent* when the detection by the COBE satellite of angular fluctuations in the background radiation was announced. These fluctuations, probably formed in an ultra-early 'inflationary' phase of cosmic expansion, were the precursors of large-scale structure in our present universe.

on inferences by geologists and palaeontologists which are equally indirect (and less quantitative). There are some fervent believers. The great Soviet cosmologist Yakov Zeldovich once claimed that the Big Bang was 'as certain as that the Earth goes round the Sun' (even though he must have known the dictum of his compatriot physicist Lev Landau that cosmologists are 'often in error but never in doubt').

I would bet at least 90% on the general concept – not quite 100%. Consistency does not guarantee truth. Our satisfaction may be as illusory as that of a Ptolomaic astronomer who has just fitted a new epicycle.

Is it absurdly presumptuous to claim to know *anything* about the beginnings of our entire observable universe? Not necessarily. It is complexity, and not sheer size, that makes things hard to understand. In the primordial fireball everything must have been broken down into its simplest constituents. The early universe really could be less baffling – more within our grasp – than the smallest living organism. It is biologists and the Darwinians who face the toughest challenge!

I will return to our universe's hot dense beginnings – and what happened in the first second (the lower part of Figure 4) – but let us now look forward rather than backward, as forecasters rather than fossil hunters.

Futurology

Cosmic time spans extend at least as far into the future as into the past. Suppose America had existed for ever, and you were walking across it, starting on the east coast when the earth formed, and ending up in California ten billion years later, when the sun is about to die. To make this journey, you would have to take *one step every 2000 years.* All recorded history would be three or four steps, just before the half-way stage – somewhere in Kansas perhaps. Not the culmination of the journey!

In this perspective, we are still near the beginning of the evolutionary process. The progression towards diversity has much further to go. Even if life is now unique to the earth, there is time for it to spread from here throughout the entire galaxy, and even beyond.

In about five billion years the sun will die; and the earth with it. At about the same time (give or take a billion years) the Andromeda Galaxy, already falling towards us, will crash into our own Milky Way, merging to form a single amorphous elliptical galaxy. But will the universe expand *forever,* attaining some asymptotic heat death? Or will it, after an immense time, recollapse to the Big Crunch?

The ultra-long-range forecast depends on how much the cosmic expansion is decelerating. The deceleration comes about because everything in the universe exerts a gravitational pull on everything else. It is straightforward to calculate that the expansion will eventually go into reverse if the average cosmic density exceeds about three atoms per cubic metre. Space seems even emptier than that: if the atoms in all the stars and gas in all the galaxies were dispersed uniformly, they would fall short of this 'critical' density by a factor of at least 50.

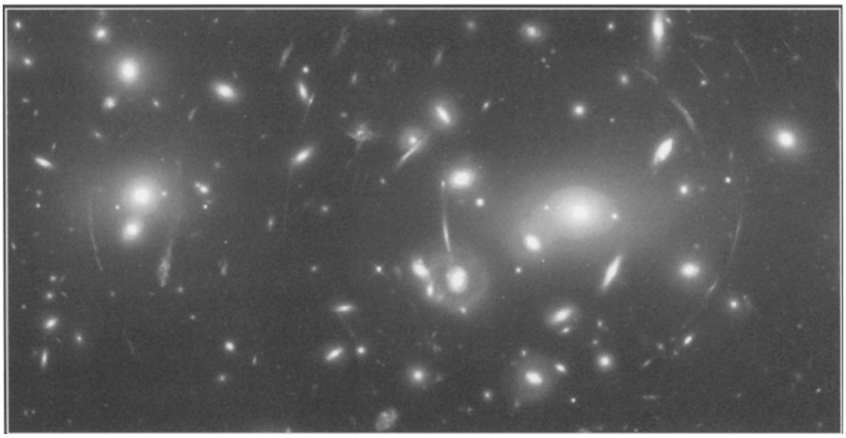


FIGURE 5 The cluster of galaxies Abell 2218. The brighter galaxies belong to the cluster. The faint 'arcs' are remote galaxies, lying far beyond the cluster, whose images are being distorted and magnified by gravitational lensing (a phenomenon whereby light rays are deflected and focused by the gravitational field of a large mass along the line of sight). The strength of the lensing implies that the cluster contains about ten times as much mass in 'dark matter' as in the galaxies we actually see.

At first sight this seems to imply perpetual expansion, by a wide margin. But the case is not so straightforward because there seems to be at least ten times as much material in 'dark' form as we see directly. One line of evidence comes from the discs of galaxies such as our Milky Way or Andromeda. These contain neutral hydrogen gas, which does not itself weigh much, but serves as a tracer of the orbital motion. Radio-astronomers can detect this gas via its emission of the famous 21 cm spectral line. It extends far beyond the limit of the optically detectable disc. The orbital speed is roughly the same all the way out. If the outermost clouds were feeling just the gravitational pull of what we can see, their speeds should fall off roughly as the square root of distance outside the optical limits of the galaxy: the outer gas would move slower, just as Neptune and Pluto orbit the sun more slowly than the earth does. The surprisingly high speed of the outlying gas tells us that an extended invisible halo surrounds these galaxies – just as, if Pluto were moving as fast as the earth, we would have to infer a heavy invisible shell outside the earth's orbit but inside Pluto's.

There is also dark matter pervading entire clusters of galaxies. Figure 5 shows a cluster of galaxies. The faint streaks and arcs are remote galaxies, several times further away than the cluster itself, whose images are, as it were, viewed through a distorting lens. Just as a regular pattern on background wallpaper looks distorted when viewed through a curved sheet of glass, the gravity of the cluster of galaxies deflects the light rays passing through it. This picture would have fascinated Fritz Zwicky, the far-sighted eccentric whom Freeman Dyson extols in Chapter 7. It was Zwicky who first realized, in the 1930s, that clusters of galaxies would fly apart unless they contained more gravitating stuff than is visible; he was also the first person to suggest that gravitational lensing might actually be observable. The visible galaxies in the cluster contain only a tenth as much material as is needed to produce these distorted images – evidence that clusters of, as well as individual, galaxies contain ten times as much mass as we see.

What could this dark matter be? Maybe it is faint stars whose centres are not squeezed hot enough to ignite their nuclear fuel; or black holes, remnants of big stars that were bright when the galaxy was young but have now died. But there are other quite different options. The hot early universe may have contained not just atoms and radiation, but other particles as well. In particular, there should be huge numbers of *neutrinos* – about a billion for every atom in the universe. So even a very tiny individual mass would make the cumulative gravitational effects of neutrinos important. But do neutrinos have any mass at all? Recent experiments at Los Alamos made such a claim but they remain controversial. The results were announced in a paper with thirty-nine authors, but the fortieth member of the group published, in the same issue of the same journal, a paper with a contrary interpretation! So we would be prudent to suspend judgement. If the claimed mass is right, neutrinos contribute more gravitating stuff than do all the stars and gas in the universe.

At least we know neutrinos exist. But particle theorists have a long shopping list of particles that *might* exist, and (if so) could have survived from the early phases of the Big Bang. If such particles pervade our galaxy, there would be 100 000 of them in every cubic metre, most passing straight through the earth without interacting. But their cross-section for colliding with ordinary atoms, though tiny, is not quite zero, and sensitive experiments are being set up to detect the rare events when this happens. The equipment must be placed deep underground, to reduce other types of background signal. A UK group is building such an experiment down a mine in Yorkshire. It is a difficult experiment, but a positive result would not only reveal what 90% of the universe is made of, but also discover new types of particle that could never be detected in other ways.

We should not be surprised that there is dark matter. There is no reason why everything in the universe should shine. The challenge is to decide among many candidates. Its dominance may demote our cosmic status still further. Copernicus dethroned the earth from a central position. Hubble showed that the sun was not in a special place. Now particle chauvinism may have to go. We ourselves, and all the stars and galaxies, would then be trace constituents of a universe whose large-scale structure is controlled by the gravity of dark matter of a quite different kind – we see, as it were, just the white foam on the wavecrests, not the massive waves themselves.

The reliably inferred dark matter in galaxies and clusters is not quite enough to bring cosmic expansion to an eventual halt. However, there is widespread theoretical prejudice (which I will return to later) that the universe has almost exactly the critical density (maybe marginally higher, so that its space–time extent is finite rather than infinite). To those of us who share this prejudice, the burden of proof is on those who contend that there cannot be much still more elusive dark matter *between* clusters of galaxies.

A word now about the emergence of galaxies and clusters. People often wonder how the universe can have started off in thermal equilibrium, a hot dense fireball, and ended up manifestly far from equilibrium: temperatures now range from blazing surfaces of stars (and their even hotter centres) to the night sky only three degrees above absolute zero. Although this seems contrary to thermodynamic intuitions that temperatures tend to equilibrate as things evolve, it is actually a natural outcome of cosmic expansion, and the workings of gravity.

Gravity has the peculiar tendency to drive things further from equilibrium. When gravitating systems lose energy they get *hotter*. A star that loses energy and deflates ends up with a *hotter* centre than before. (To establish a new and more compact equilibrium where pressure can balance a (now stronger) gravitational force, the central temperature must *rise*.)

Gravity does something else. It renders the expanding universe unstable to the growth of structure, in the sense that even very slight initial irregularities would evolve into conspicuous density contrasts. Theorists are now carrying out increasingly elaborate computer simulations of how this happened. Slight fluctuations are 'fed in' at the start of the simulation: exactly how they are prescribed depends on the cosmological assumptions. Figure 6 shows three 'frames' from a simulation of a region containing a few thousand galaxies, large enough to be a fair sample of our universe. As the expansion proceeds, regions slightly denser than average lag further and further behind. Eventually they stop expanding and condense into gaseous protogalaxies which fragment into stars. The same process on larger scales leads to clusters and superclusters. The aim is to make different assumptions about the initial fluctuations, the dark matter, etc., and see which leads to a pattern of structure closest to a typical sample of the real universe.

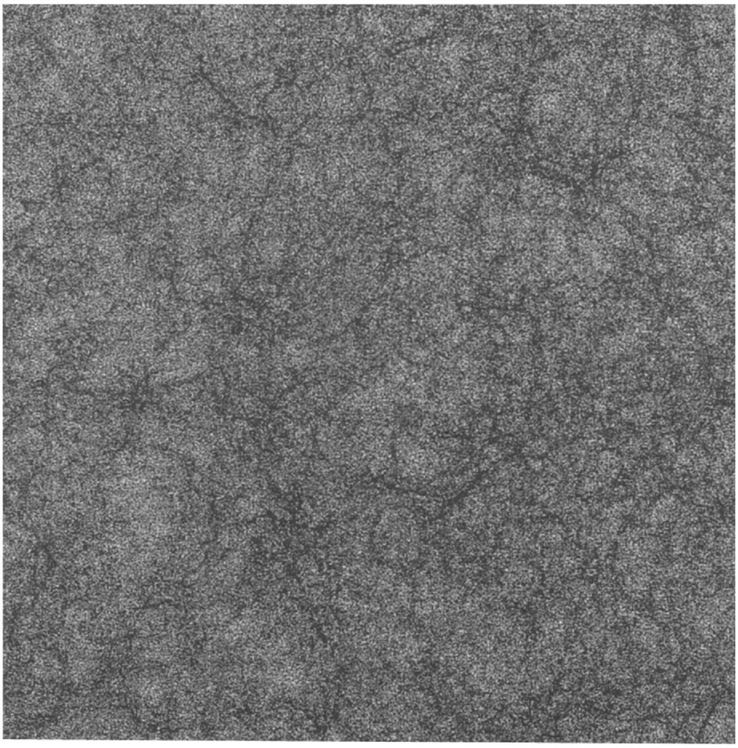
The microwave background, a relic of a pregalactic universe that was not perfectly smooth, should bear the imprint of the initial fluctuations. This radiation in effect comes from a very distant surface or horizon, far beyond the quasars; it has propagated freely since a time long before the clusters had fully formed. Radiation from an incipient cluster on that surface would appear slightly cooler, because it loses extra energy climbing out of the gravitational pull of an overdense region. Conversely, radiation from the direction of an incipient void would be slightly hotter. The fractional differences in the temperature involve this same small ratio – they are only about one part in 100 000.

Temperature non-uniformities with about this amplitude were first detected by NASA's COBE satellite. To measure such small effects was a technical triumph. But the fluctuations were not unexpected. It would have been far more baffling if they had not been there. That would have implied an early universe so smooth that it would not have been compatible with the conspicuous clustering we see in the present universe – we would have had to postulate some process more efficient than gravity for pulling these structures together.

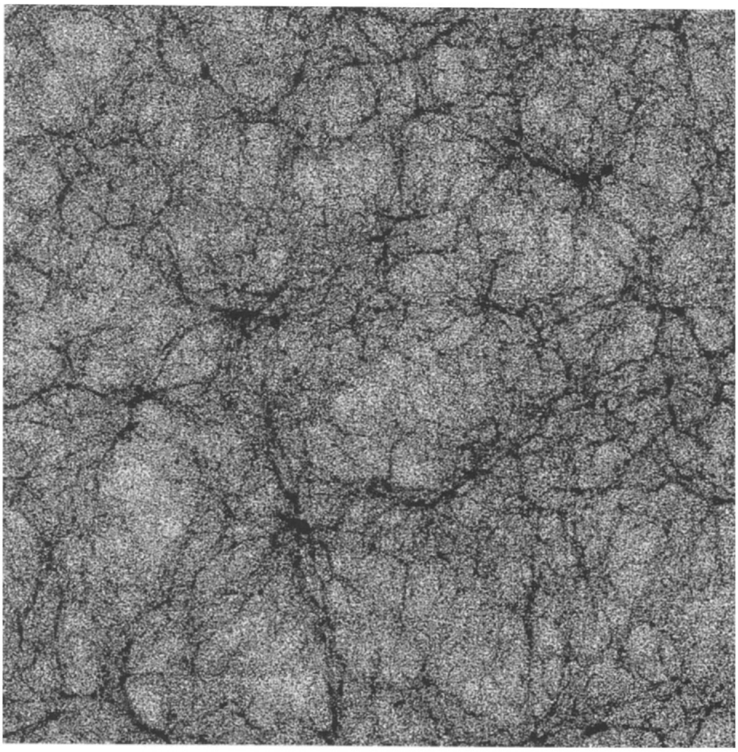
COBE got the first positive results, but already these are being complemented and extended by ground-based and balloon experiments and two ambitious new space experiments are now planned. The embryonic precursors of galaxies and larger cosmic structures are no longer just hypothetical entities but can actually be measured.

If one had to summarize, in just one sentence, 'What's been happening since the Big Bang?' the best answer might be to take a deep breath and say 'Ever since the beginning, gravity's "anti-thermodynamic" effects have been amplify-

(a)



(b)



https://doi.org/10.1017/CBO9780511629044.009 Published online by Cambridge University Press

(c)

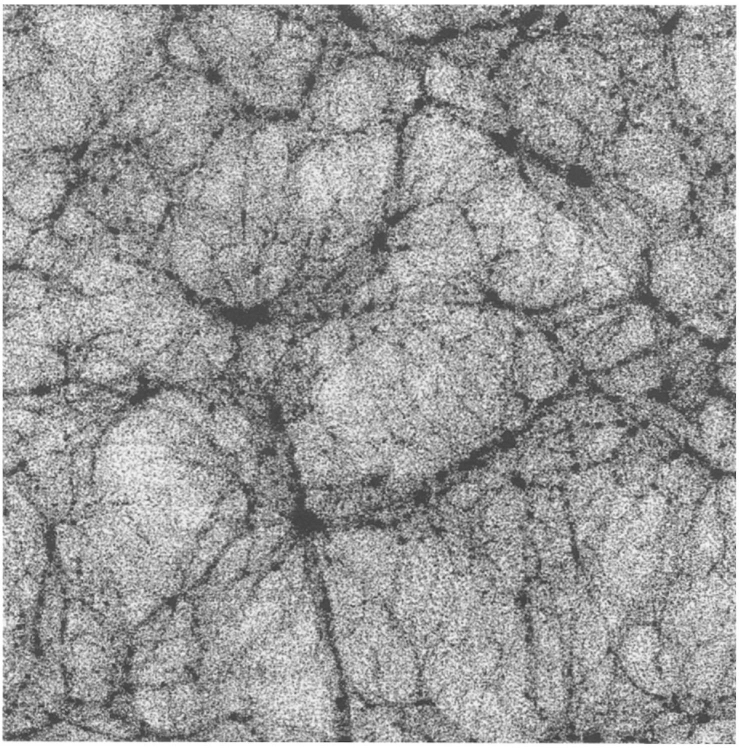


FIGURE 6 These three 'frames' show three stages in the development of clustering in the expanding universe (computed by the Virgo consortium). The scale is adjusted so that each picture shows the same amount of mass (which would, of course, have been more compressed at earlier stages in the expansion). The region shown would be about 300 million light years across.

ing inhomogeneities, and creating progressively steeper temperature gradients – a prerequisite for emergence of the complexity that lies around us ten billion years later, and of which we are part.'

The way cosmic structures evolve is in principle as predictable as the orbits of the planets, which have been understood since Newton's time. But to Newton, some features of the solar sytem were a mystery. He showed why the planets traced out ellipses. It was a mystery to him, however, why they were 'set up' with their orbits almost in the same plane, all circling the sun in the same way. In his *Opticks* he writes:

> blind fate could never make all the planets move one and the same way in

> orbits concentrick . . . Such a wonderful uniformity in the planetary system must be allowed the effect of choice.

This co-planarity is now understood: it is a natural outcome of the solar system's origin as a spinning protostellar disc.

The demarcation between phenomena that are the manifestations or working out of known laws, and those which are mysterious 'initial conditions' still exists, as sharply as it did for Newton. We are still, at some stage, reduced to saying 'Things are as they are because they were as they were'. The progress has been to push the barrier back from the beginning of the solar system to the first second of the Big Bang.

Just as Newton had to specify the initial trajectories of each planet, our calculations of cosmic structure need to specify a few numbers:

(1) The expansion rate.
(2) The proportions of ordinary atoms (or their constituent quarks), dark matter and radiation.
(3) The character of the fluctuations – large enough to evolve into structures, but not to invalidate the overall uniformity.

Can we take a further step, and explain these numbers in terms of some processes that happened still earlier than the starting point of these simulations?

The trouble is that, the further we extrapolate back, the less confidence we have that known physics is either adequate or applicable. For the first millisecond, everything would have been squeezed denser than an atomic nucleus. For the first 10^{-14} second the energy of every particle would surpass what even the new accelerator at CERN (Centre Européene pour la Recherche Nucleáire) will reach.

Not even the boldest theorists can extrapolate back beyond the stage when quantum effects become important for the entire universe. The two great foundations of twentieth-century physics are, on the one hand, Einstein's theory of gravity (general relativity) and, on the other, the quantum uncertainty principle. But there is generally no overlap between these two great concepts. Gravity is so weak that it is negligible on the scale of single molecules, where quantum effects are crucial. Conversely, gravitating systems such as planets and stars are so large that quantum effects can be ignored in studying how they move. Right back at the beginning of the universe, the densities could have been so high that quantum effects were important for the whole universe. This happens at the Planck time, 10^{-43} second.

It is helpful to divide cosmic history into three parts. Part 1 is the first millisecond, a brief but eventful era spanning forty decades of logarithmic time, starting at the Planck time. This is the intellectual habitat of the high-energy theorist and the quantum cosmologist. Part 2 runs from a millisecond to about a million years. It is an era where cautious empiricists like myself feel more at home. The densities are far below nuclear density, but everything is still expanding in an almost homogeneous fashion. The relevant physics is firmly based on laboratory tests, and theory is corroborated by good quantitative evidence: the cosmic helium abundance, the background radiation, etc. Part 2 of cosmic history, though it lies in the remote past, is the easiest to understand. The tractability lasts only so long as the universe remains amorphous and structureless. When the first gravitationally bound structures condense out – when the first stars, galaxies and quasars have formed and lit up – the era studied by traditional astronomers (Part 3) begins. We then witness complex manifestations of well-known basic laws. Gravity, gas dynamics and feedback effects from early stars combine to initiate the complexities we see around us and of which we are part. Part 3 of cosmic history is difficult for the same reason that all environmental sciences, from meterology to ecology, are difficult.

Then we realize that the few basic numbers that determine how the universe has evolved are all legacies of the uncertain physics of Part 1. I would now like to discuss some ideas about this, and about how even the physical laws themselves may have been imprinted in the ultra-early universe. Be warned that I am now entering speculative territory, where even Zeldovich would harbour some doubts.

First, what about the initial expansion rate? This has to be very precisely tuned. The two eschatologies – perpetual expansion or recollapse to a 'crunch' – seem very different. But our universe is still expanding after ten billion years. Had it recollapsed sooner, there would not have been time for stars to evolve; indeed, if it had collapsed after less than a million years it would have remained opaque, precluding any thermodynamic disequilibrium. On the other hand, the expansion cannot be too much faster than the critical rate. Otherwise gravity would have been overwhelmed by kinetic energy and the clouds that developed into galaxies would have been unable to condense out.

In Newtonian terms the initial potential and kinetic energies were very closely matched. How did this come about? Why does the universe have the large-scale uniformity which is a prerequisite for progress in cosmology?

The answer may lie in something remarkable that happened during the first 10^{-36} second, when our entire observable universe was a few centimetres across. Ever since that time, the cosmic expansion has been *decelerating*, because of the gravitational pull that each part of the universe exerts on everything else. Theoretical physicists have come up with serious (though still, of course, tentative) reasons why, at the colossal densities before that time, a new kind of 'cosmical repulsion' might come into play and overwhelm 'ordinary' gravity. The expansion of the ultra-early universe would then have been exponentially *accelerated*, so that an embryo universe could have inflated, homogenized and established the 'fine-tuned' balance between gravitational and kinetic energy when it was only 10^{-36} second old.

This generic idea that the universe went through a so-called inflationary phase is compellingly attractive. The fluctuations from which clusters and superclusters form, and the even vaster ones whose imprint on the background radiation spreads right across the sky, may be the outcome of microscopic quantum phenomena from an ultra-ancient epoch when the universe was squeezed smaller than a golfball. We do not, of course, know the physics that prevailed at this ultra-early time, but there is a real prospect of discovering something about it. Specific models of how the inflation is driven make distinctive predictions about things we can observe: large-scale clustering, and small non-uniformities in the background radiation over the sky. We shall soon be confronting the inflationary era of cosmic expansion with *real empirical tests*, just as we can already, by measuring the abundances of helium and deuterium, learn about physical conditions during the first few seconds.

Some other remarkable fossils of the ultra-early universe, conjectured by theorists, are being looked for. Among these are magnetic monopoles, and small black holes the size of an atom but weighing as much as a mountain. Even more astonishing are cosmic strings – elastic loops of concentrated energy, thinner than an elementary particle, but long enough to stretch across the universe, flailing around at nearly the speed of light, and heavy enough for their gravity to affect entire galaxies. These would be crucial links between the cosmos and the microworld.

The inflationary idea also, incidentally, strongly suggests that the mean cosmic density is very close to the 'critical' value that demarcates the boundary between perpetual expansion and eventual recollapse. That is the basis of the prejudice I mentioned earlier in favour of the critical density.

The universe has, then, in a sense, zero net energy. Every atom has a rest mass energy: Einstein's mc^2. It also has a negative potential energy due to the gravitational field of everything else, and this exactly balances its rest mass. Thus, it 'costs nothing', as it were, to expand the mass and energy in our universe.

Physicists sometimes loosely express such ideas by saying that the universe can essentially arise 'from nothing'. They should watch their language, especially when talking to philosophers. The physicist's vacuum has all particles and forces latent in it; it is a far richer construct than the philosopher's 'nothing'.

Any such theory would, of course, be hard to check, and may never be taken too seriously unless it has a compelling inevitability about it, a resounding ring of truth that compels assent. In any case it would not tell us *why there was* a universe. To quote Stephen Hawking in *A Brief History of Time* (1988): 'What is it that breathes fire into the questions? Why does the Universe go to all the bother of existing?'

The character of the universe and everything in it depends, of course, on the strengths of the basic physical forces – gravity, electromagnetism, etc. These are also part of our 'initial conditions'.

I already mentioned gravity's counterthermodynamic tendencies. Gravity has a second crucial feature, its feebleness. The gravitational pull between two protons is 36 powers of 10 weaker than the electrical repulsion between them. On large scales, gravity wins because everything has, as it were, the same sign of 'gravitational charge' – there is no cancellation of positive and negative, as in electricity.

Imagine a series of lumps containing successively 10, 100, 1000 atoms and so on. The 24th, containing 10^{24} atoms, would be about the size of a sugar lump. The 40th would be a mountain or small asteroid. The gravitational energy of each atom due to the rest of the lump it belongs to is proportional to the mass/radius (M/R). The radius of a lump is proportional to the cube root of the mass, so gravity gains in importance as the 2/3 power of the number of particles. Gravity is handicapped by 36 powers of 10 to start with. So it becomes competitive only for the 54th lump, containing 10^{54} atoms, because 36 is 2/3 of 54. That lump would be as big as Jupiter. Anything still larger is crushed by gravity, and would become a star. It is because gravity is so weak that stars have to be so massive. In any lesser aggregate, gravity could not squeeze the material to high enough central densities and pressures for nuclear fusion to occur.

Consider now a hypothetical universe where gravity was 10^{10} times stronger than in ours – 'only' 26 rather than 36 powers of 10 weaker than the electrical forces in atoms – but where the microphysics was unchanged. Atoms and molecules would behave just as in the actual universe, but objects would not need to be so large before gravity became competitive. In this imagined universe, mini suns with 10^{-15} times the sun's mass would live for about one year.

The (literally) crushing effect of strong gravity would hamper complex evolution on this hypothetical world. No animals on any planet large enough to retain an atmosphere could be any bigger than insects, and they would need thick legs to support them. More constraining still is the limited time. Chemical and metabolic processes depend on microphysics, and would not be speeded up. But the mini sun would have exhausted its energy before even the first steps in organic evolution had got under way.

If gravity were stronger, there would be fewer powers of 10 between *astrophysical* timescales and the basic microphysical timescales for physical or chemical reactions. Paradoxically, the weaker gravity is – provided it is not exactly zero – the grander and more complex can be its consequences.

Our cosmic environment is exceedingly sensitive to other physical quantities. For instance, if protons had electrical charges just a small percentage greater, then no atoms other than hydrogen could exist: chemistry would be a simple subject. Indeed, if you imagine 'turning a set of knobs' to adjust the physical constants, most choices would lead to 'still-born' universes in which the prevailing laws would not allow *any* complexity to emerge. These universes might, for instance, never deviate from thermodynamic equilibrium, or (perhaps because gravity was very strong) they might exists for too short a time, or (more radically) have only two spatial dimensions. How should we respond to this line of thought? Its implications depend very starkly on what the final theory (if there is one) is like. There are two contrasting possibilities.

One, option A, is that some such final theory fixes all the physical constants uniquely so that they are all calculable from some fundamental equation. The physics governing our universe *could not* then logically have been otherwise. It would then just be a brute fact that the uniquely specified physical constants happened to lie in the narrowly restricted range that allowed such complexity to evolve in the low-energy world we inhabit. The intricate consequences implicit in the fundamental equations may astonish us, but our amazement would be no less subjective than that of a mathematician who is surprised at the intricately interrelated consequences of a simple algorithm. (Take, for instance, the 'Mandelbrot set'. The recipe or algorithm for constructing this astonishing pattern can be written in just a few lines, but it encodes an intricate variety of new structures, however much we magnify it.) Any apparent fine-tuning would have to be accepted as just coincidental.

But there is an alternative, option B. The numbers we call the constants of physics may not be uniquely fixed by the fundamental theory. If an 'ensemble of universes' existed, each one governed by different physics, there would be some in which the conditions were tuned propitiously for complexity to emerge.

This second option would be fulfilled by some variants of inflationary cosmology. According to the Russian cosmologist Andrei Linde, our universe, itself extending far beyond the ten billion light years we can so far see, rather than being 'everything there is' is just one bubble linked to other space–times in an infinite eternally replicating ensemble – the metauniverse. To recall my earlier analogy: the ocean may extend far beyond our horizon, but that does not mean it extends uniformly to infinity.

The physical forces, and the masses of the elementary particles, are the outcome of some kind of phase transition, connected to the force that drives inflation. The imprint left by these phase transitions (the relative strengths of the present-day forces) may then be somewhat arbitrary or 'accidental', like the patterns of ice on a pond, or the way a magnet behaves when cooled. The 'constants of nature' would have quite different values in other universes in the ensemble.

If this (very schematic!) picture had anything in it, the apparent 'fine-tuning' need not surprise us at all. The fundamental constants would be the outcome of random accidents (option B), so in a sufficiently capacious metauniverse the physics would inevitably turn out propitious for complexity to evolve in some members of the ensemble, and we are obviously not surprised to find ourselves in one of these.

This strays dangerously close to what is called anthropic reasoning. Fortunately, there is no space here to stray further, but I do not think anthropic reasoning is quite as silly or vacuous as it is sometimes made to sound. Indeed, it gives us extra grounds for suspecting that any final theory will have the permissive character of option B: there is then nothing surprising about the existence of a universe with specific 'coincidental' features.

I earlier gave an implicit 'health warning', urging that you should not necessarily believe all the cosmology you read in the newspapers. Let me conclude by trying to assess the state of play – where we can lay confident bets and where we should not (or not yet).

Cosmology has made quite amazing progress since the years when the now-abandoned steady-state theory was being boisterously debated, especially here in Cambridge. That theory, then, seemed specially attractive because, if it were correct, every evolutionary process (from atoms to galaxies) had to be going on somewhere now, and so should be observable and could be investigated. At that time, it was felt that a Big Bang theory could never be made scientific, because the key processes would be shrouded in the remote past. They are, indeed, in the remote past, but what is remarkable is that they are not inaccessible to study. Telescopes can view directly 90% of cosmic history; other techniques can probe still earlier phases. We are confident about cosmic history, at least in broad outline, back to one second, when the first elements were made – back to the start of what I have called Part 2 of cosmic history. The challenge is now to delineate more fully how an almost featureless fireball evolved into the cosmos of which we are a part ten billion years later.

The last illustration (Figure 7) shows Einstein: the young Einstein, not the benign and unkempt sage of poster and T-shirt. One of his best-known remarks is that 'The most incomprehensible thing about the universe is that it is comprehensible'. Cosmology has progressed because the laws of physics we study in the laboratory apply in the remotest quasar, and back to the first few seconds of the 'Big Bang'. When there is no firm link with laboratory science, cosmological inferences are more fragile. That is why we are on shakier ground when we venture back into the first millisecond, and we should not disguise this. Our methodology is no longer like that of a geologist, or practitioner of other historical sciences: new basic physics has to be discovered, rather than established physics being applied.

Ideas about the ultra-early universe are often presented in popular books in the same tone as descriptions of the Hubble Law, the microwave background, etc. This makes me somewhat uneasy: there is a risk that overcredulous readers may accept tentative speculations; on the other hand, more sceptical readers

FIGURE 7 The young Einstein.

may overlook the strengthening range of observations that buttress claims about the later stages of cosmic evolution.

Some previously speculative questions are now coming within the scope of serious science. In the ultra-early universe, the mysteries of the cosmos and the microworld overlap. Processes as early as 10^{-36} second may have imprinted the excess of matter over anti-matter, the ripples in the fabric of space–time, and perhaps the physical laws themselves.

Modern cosmology has been moulded by cultural climate, and given an impetus by the influx of scientists – particle physicists, for instance – with different expertise and style. It has been moulded further by the opportunities and constraints of the available techniques: experimental, observational and computational. These sociological dimensions are in themselves fascinating. However, such studies should not obscure what to us 'in the zoo' seems the essence of our science: that it is a collective and cumulative enterprise which, albeit fitfully, is bringing the workings of the cosmos into a sharper and 'truer' focus.

FURTHER READING

Audouze, J. and Israel, G. (eds.) *The Cambridge Atlas of Astronomy*, 3rd edition. Cambridge: Cambridge University Press, 1994.

Barrow, J. *The Origin of the Universe*, London: Weidenfeld & Nicolson, 1995.

Begelman, M. and Rees, M. *Gravity's Fatal Attraction: Black Holes in the Universe*, New York: W. H. Freeman, 1995.

Rees, M. *Perspectives in Astrophysical Cosmology*, Cambridge: Cambridge University Press, 1995.

Rees, M. *Before the Beginning: Our Universe and Others*, London: Simon & Schuster; New York: Addison Wesley, 1997.

Silk, J. *A Short History of the Universe*, New York: W. H. Freeman, 1995.