CHAPTER ONE

The Nature of the Universe

1. Greek Astronomy

The earliest Greeks seem to have thought the earth to be flat or cylindrical, but, from the fifth century on they knew that the earth was a sphere, and nearly all of them supposed this sphere to be in the centre of the universe and motionless. The stars were supposed to be at the surface of a vast sphere concentric with the earth. The stars rise and set once a day, so this sphere was supposed to rotate on its axis once in 24 hours. Outside this there was nothing, or a sort of unformed chaos. The chief problem of antiquity was to work out how the bodies between the earth and stars moved. These were the moon and sun, and the planets, Mercury, Venus, Mars, Jupiter and Saturn. These rose and set, but not in exact time with the stars. The moon rises about an hour later every day, the sun about four minutes (reckoned by the stars). The planets behave very oddly, for although they generally very slowly fall behind the stars, they sometimes stay still relatively to them and even gain on them for a short time. The Greeks decided that the only *fitting* path for a heavenly body was the 'perfect' figure of a circle; and Aristotle, indeed, laid down that just as on earth bodies, if undisturbed, naturally fell down (i.e. towards the centre), so heavenly bodies *naturally* moved uniformly in circles round the centre.

The problem the Greeks set themselves was to discover a combination of circular motions which would give rise to the curious paths which the planets in fact took. The first important solution was that of concentric spheres. Imagine a 'nest' of spheres one inside the other, turning on axles like wheels. Each sphere has its axle set in the surface of the sphere next larger than itself. The axles incline in various directions and the spheres rotate at very different speeds. The planet, borne on the innermost sphere, will combine the motions of them all, and will describe at intervals a sort of figure-of-eight path which is not unlike its apparent path through the heavens. The correspondence of this system with the observed facts was not at all exact; and another difficulty was that the planets change in brightness and so seem sometimes to be nearer to the earth and sometimes further from it, which could not be the case if they moved on a sphere concentric with it.

The idea of concentric invisible spheres, adopted though not invented by Aristotle, persisted until the seventeenth century;[1] but at the same time another and better system, the Ptolemaic, was used by astronomers from the second to the seventeenth century. The earth, as before, is supposed to be motionless at the centre and the planets to move round it. Each planet moves round a circle, the epicycle, the centre of which moves in another circle[2] which surrounds the earth, though its centre is not at the earth's centre. The centre of the epicycle does not, however, move uniformly, but so as to lie always on the uniformly rotating radius of a third circle.[3] This complicated system, if the size of the circles, the position of their centres and the rate of rotation were properly chosen, could give a very fair account of the motion of the planets. This was important in all ages up till about 1700, because men took much account of astrology, which foretold the events of their lives by the relative positions of the stars and planets. Although by modern standards these systems were far from accurate, they were a great advance on no system at all, and, moreover, were wonderful feats of mathematical reasoning.

The Greeks did not confine themselves to systems in which the earth stands still at the centre of the universe. Aristarchus of

[1] *These were the spheres famous for their music. Many people during the Middle Ages lost sight of the fact that a planet needed a whole nest to itself, and thought of there being just one sphere for each planet, including the sun and moon, with an eighth for the stars and sometimes a ninth outermost one. (Ed.)*

[2] *The deferent. (Ed.)*

[3] *i.e. The centre of the epicycle does not move at a uniform* speed *round the circumference of the deferent; in other words it moves so as to lie not on the uniformly rotating radius of the deferent, but on the uniformly rotating radius or spoke of a third circle, though free, so to speak, to slide up and down along its length. The centre (called the equant) of this third circle and the earth's centre lie equidistant on either side of the centre of the deferent. (Ed.)*

Samos, about 270 B.C., proposed a system identical with the Copernican; it attracted few, if any, followers, however, and there was talk of a charge of impiety being brought against him.

F. SHERWOOD TAYLOR, *Science Past and Present*, 1945

To say that the only 'fitting' path for a heavenly body is the perfect figure of a circle, or that stones have a 'desire' to reach the centre of the earth, or that nature 'abhors' a vacuum – all these strike us nowadays as too quaint, too anthropomorphic, to qualify as scientific statements. But in their time they were serious attempts to express certain observed facts of nature in a generalized form. And when closer observations showed the inadequacy of the first of them in its simplest form, the further attempt to plot planetary movements as combinations of several circles at least imposed some kind of coherence or order on what might otherwise have seemed wholly random scrawlings across the sky.

Ptolemy lived in the second century A.D., Aristarchus in the third B.C. Not until Copernicus (1473–1543) did the heleocentric hypothesis find another champion. And even then, Copernicus merely put it forward as another possible way of accounting for appearances, which was why he could number bishops among his patrons. It was Galileo's tactless insistence, not that this might be so but that this was *so, which made it impossible for the Pope to ignore his challenge to the authority of church and scripture.*

However, it was left to Newton to establish finally not only that this was so but how (and why?) this was so.

11. The Achievement of Newton

Men have known for several thousand years that the sun and the planets move in regular ways against a background of stars which seem to be still. These regularities can be used to look forward as well as back: the Babylonians were able to use them to forecast eclipses of the sun. The sun, the moon and the planets can be pictured as being carried round the earth on these regular paths in great shells or spheres. Or the paths, which seen from the earth are curiously looped, can be thought of as the rolling of wheels upon wheels; it was in this way that Ptolemy and other Greeks in Alexandria patterned them on the night sky eighteen hundred years ago. Ptolemy's picture does not claim to explain the movement of the planets, if indeed we could make him understand this meaning of the word 'explain' which has become natural for us. It gives an order to their movements by describing them, and so tells us where we may expect to see them next.

Two things happened in the sixteenth century to make astronomy ill at ease with this description; and they are both of interest, because they remind us that science is compounded of fact and logic. The Danish astronomer Tycho Brahe took better and more regular observations of the positions of the planets, and they showed that Ptolemy's paths, charming though they looked as mathematical curves, were really only rather crude guides to where the planets rolled. And even earlier, Copernicus showed that these paths were much simpler if they were looked at not from the earth but from the sun. Early in the seventeenth century, these two findings were combined by Kepler, who had worked for Brahe. Kepler used the measurements of Brahe and the speculations of Copernicus to frame general descriptions of the orbits of the planets: for example, he showed that, seen from the sun as focus, a planet sweeps out equal areas of its ellipse in each equal interval of time.

It was these empirical generalisations of Kepler which Newton and his contemporaries worked from when they began to look for a deeper order below the movements of the planets. They had also a new weapon of theory. For while Kepler had been at work in the north, Galileo in Italy had at last overthrown the physical conceptions in the works of Aristotle, which had long been attacked in Paris. By the time the Royal Society was founded, the complicated Greek ideas of motion with their conflict of earth and air, of impact and vacuum were out of the way. There were no clear new laws of motion yet; it was left to Newton to set these out; but there were fair descriptions of where and how masses in fact move, and no interest at all in where they ought to want to move.[1]

What was the nature of Newton's insight? How did he exercise those great gifts, and seize the great opportunity which I have described?

If we put what he did most baldly, it is this: that he carried on the simplification which Kepler had begun, but carried it beyond geometry into physics. Ptolemy, Copernicus, Tycho Brahe and Kepler, at bottom all looked no further than to plot the paths of the planets. Kepler found likenesses between these paths deeper than anything in the traditional astronomy, for his were likenesses of motion as well as shape. Nevertheless his paths remained descriptions, more accurate and more concise than Ptolemy's, but no more universal. For even when Kepler speculated about an attraction of the planets to the sun he had no principle to link it to the movement of earthly masses. Galileo had the first glimpse of that; and there were others as the seventeenth century marched on, who knew what kind of principle they were looking for; but it was Newton who formulated it, sudden and entire. He said that change of motion is produced by force; that the motion between masses, whether apple, moon and earth, or planet and sun, is produced by gravitational forces which attract them to one another. And he alone of his contemporaries had the mathematical power to show that, if these forces are postulated in the right way, then they keep the planets spinning like a clockwork, they keep the moon on its orbit, and the tides moving under the moon; and they hold the universe together. These achievements are so great that they out-top astronomy; and they are only a part of Newton's whole achievement. But more than the achievement, it is the thought within which deserves our

[1] *Aristotle was more interested in final than in efficient causes.* (*The final cause of the stone's hitting Goliath's forehead was David's wish to kill him; the efficient cause was the impetus given the stone by the movement of certain muscles in David's arm.*) *There had to be a motive, as it were, for a moving object. So, stones and rain* (*earth and water*) *fell in order to rejoin and take their rightful place in the two innermost spheres of the universe, whereas bubbles and flames* (*air and fire*) *rose in order to become part of the next two spheres.* (*Ed.*)

study. There is the searching conception of the universe as a machine; not a pattern but a clockwork. There is the conception of the moving forces within the machine: the single spring of action in gravitation. There is the brilliant compromise between the description of the astronomers and the First Cause of the theologians, in which Newton shaped once for all the notion of cause as it has remained ever since. Newton indeed has taken over just enough of the Aristotelean nature of things to make the world work by giving all matter a single nature – that it seeks to join with all other matter.

J. BRONOWSKI, *The Common Sense of Science*, 1951

How fair is it to argue that where Greek astronomy described Newton explained? Certainly Newton's account is more satisfactory, in that it incorporates, and establishes a connexion between, observations of movements of many different kinds besides planetary ones. Equally certainly it is a limited one in that, although it is completely based on the idea of gravity, Newton 'evades the hypnosis of the insoluble problem'[1] *and refuses to put forward any hypothesis as to the nature of this force which can operate across or through such empty distances. Does his use of this concept of gravity amount then, in the last analysis, to an explanation, or merely a very much more detailed description (a physical or mechanical description, if you will, rather than a geometric one) of the behaviour not just of planets but of apples and galaxies as well? When, in fact, do descriptions cease to be descriptions and become explanations?*

As for whether Copernicus and Galileo were right in maintaining that the earth goes round the sun . . .

III. Relatively Speaking

Before Copernicus, people thought that the earth stood still and the heavens revolved about it once a day. Copernicus taught that 'really' the earth rotates once a day, and the daily revolution of

[1] *H. T. Pledge,* Science Since 1500, *London 1939.*

sun and stars is only 'apparent'. Galileo and Newton endorsed this view, and many things were thought to prove it – for example the flattening of the earth at the poles, and the fact that bodies are heavier there than at the equator. But in the modern theory the question between Copernicus and his predecessors is merely one of convenience; all motion is relative, and there is no difference between the two statements: 'the earth rotates once a day' and 'the heavens revolve about the earth once a day'. The two mean exactly the same thing, just as it means the same thing if I say that a certain length is six feet or two yards. Astronomy is easier if we take the sun as fixed than if we take the earth, just as accounts are easier in a decimal coinage. But to say more for Copernicus is to assume absolute motion, which is a fiction. All motion is relative, and it is a mere convention to take one body as at rest. All such conventions are equally legitimate, though not all are equally convenient.

BERTRAND RUSSELL, *The ABC of Relativity*, 1925

The next passage, though perhaps conceding premature victory to one of the two sides to an argument which still goes on, is an admirable summary of much recent astronomic thought.

It also gives a good idea of how scientists are sometimes able to attack a problem as seemingly insoluble as how the universe began, and of the peculiar difficulties which face astronomers more than most scientists. For, although geologists and evolutionary biologists, for instance, are also obliged to speculate about events which happened long ago and which may in many respects have been unlike anything happening now, and although, at the other end of the scale, the behaviour of electrons or the complex and all-important internal architecture of protein molecules must seem as exasperatingly inaccessible to the clumsy and inadequate means of perception at man's disposal as is the nature of those galaxies we can only hear, and at that faintly across vast stretches of both time and space, the obstacles to an advance of astronomical knowledge are still probably uniquely intimidating.

IV. The Origin of the Universe

Just as it was natural to think of the earth as the centre of the solar system, so for a long time there was a tendency to think of the sun as the centre of the starry system. But during this century the sun has been steadily demoted. It is now seen as a rather modest star situated towards the edge of a disc-shaped galaxy or array of stars. This galaxy is itself one member of a 'local' cluster of galaxies – and thousands of such clusters are scattered through the universe.

At the same time, our scale of cosmic distance has stretched vertiginously. By the turn of the century, the nearest star was known to be four 'light years' away, or 2,352,000 million miles. By 1930 it was agreed that the diameter of our galaxy is about 100,000 light years, while a neighbouring galaxy in Andromeda is nearly one million light years away. By 1936 the American, Edwin Hubble, was speaking of galaxies 500 million light years away, but in 1952 Walter Baade reinterpreted the observations and blithely multiplied the size of Hubble's universe by two. Subsequent corrections have again tripled this, and distances of several thousand million light years are now almost commonplace.

But the most startling discovery has been that all these distant galaxies appear to be hurtling away from us – and the farther away they are, the faster they hurtle. This is the starting point for all modern cosmologies.

The observation was based on a phenomenon which may be experienced on a railway platform. The whistle of an approaching express train sounds higher than that of a receding one – the pitch drops suddenly as the train roars through the station. This is because the sound waves are crowded together, so to speak, as the train approaches, and stretched out as it recedes.

The same effect applies to light waves. An approaching galaxy, blowing its luminous train whistle, should shift its light to a higher 'pitch' – it looks slightly more blue in colour. A receding galaxy looks redder. The effect is very slight, but it can be detected by measuring a shift in certain dark lines which appear when the light from the galaxy is analysed with a spectroscope.

By 1930 it had become clear that most of the galaxies showed a 'red shift', and were thus receding. But even more startling, the farther the galaxy, the greater the shift. Last summer, an object was photographed whose red shift is so great that it must be receding at nearly half the speed of light, and is about 4,500 million light years away.

If this relation between speed and distance holds still farther out, there could be galaxies receding from us so fast that their light never reaches us, and they can never be observed. Thus the speed of light now defines the ultimate frontier of the observable universe. It is within this frontier that the cosmologist must go to work.

The discovery of the receding galaxies led straight to the 'big bang' theory of the universe. If the galaxies are getting farther apart, they must originally have been closer together. By measuring the speed of expansion and reversing it, one can calculate a time in the past when all the galaxies were packed close together.

At this time, it was supposed, the universe may have consisted of a 'primeval atom' of incredibly concentrated energy. This atom then blew apart, and the universe has been expanding like a bomb ever since. In the future, it may simply disperse and run down like a clock – or it might start to contract again into another primeval atom.

Some scientists found this picture profoundly unsatisfactory. It raises unanswerable questions about the origin of the universe, the nature of the primeval atom, or the cause of the big bang. Why not simply abolish these questions?

The 'steady state' theory, first discussed half as a joke by Bondi and Gold, asks us to imagine a universe which has always existed and always will, and does not evolve in any way. It would look broadly the same anywhere in space and anywhere in time. This implies that the average number of galaxies in any part of space must stay the same. But astronomy shows that the galaxies are all rushing away from each other. Therefore, said the steady state theory, we must postulate a continuous creation of matter to fill up the gaps.

This, it was suggested, simply appears continuously in empty space in the form of hydrogen. It condenses to form new galaxies, which develop just fast enough to balance the dispersion of existing ones, thus maintaining the average galaxy population for every part of the universe at the same time. Fred Hoyle showed that this daring idea, apparently contradicting classical laws of physics about the conservation of matter and energy, could be firmly based on a development of the relativity theory.

These ideas have gained momentum. Their great attraction was that they could be tested by experiment. If the universe is evolving, the galaxies would have been closer together in the past than they are to-day; but if the universe is in a steady state, it would look the same however far back in time we went.

Now astronomy is in the curious situation that it can undertake a kind of time travel and test these alternatives directly. The reason is that the stars and galaxies are so far away that light takes an enormous time to reach us. Light reaching us now left the more distant galaxies thousands of millions of years ago. We are thus not seeing them *as they are now*, but *as they were*. What is more, the stars that are farthest off in space are farthest off in time. Probing out into space with telescopes is like digging down a geological stratum of fossilised light. And a study of the most distant galaxies can indicate what the universe was like long ago.

If the big bang theory is right, the most distant galaxies we see should be closer together – since we are seeing them as they were nearer to the time of the bang. But if the 'steady state' theory is right, the distant galaxies should be distributed in the same way as the near ones. This was the most direct and obvious test suggested by the theory.

But a snag cropped up. It turned out that optical telescopes could not hope to see quite far enough out into space (or, if you prefer, back into time). The 200-inch telescope at Mount Palomar, California, the largest in the world, can just begin to probe the regions where there should be a detectable difference between an evolving and a 'steady state' universe. But the results are too marginal to be conclusive.

At this point, radio astronomy came into its own. Surveys of the sky with radio telescopes revealed a considerable number of 'radio stars'. These were point-like sources of radio waves in the sky, only a few of which seemed associated with objects visible through telescopes. But in 1952, Walter Baade turned the 200-inch telescope on the position of an intense radio star in the constellation of Cygnus which had been accurately pin-pointed by F. G. Smith at Cambridge. He succeeded in photographing a most peculiar object; it appeared to be a pair of galaxies in collision about 500 million light years away. Another theory holds that it is one galaxy splitting in two, but the important thing is that it emits extremely powerful radio waves. Optically it is almost invisible, but to the radio telescope it is as 'bright' as the sun.

It was immediately realised that if such a distant object could generate such powerful radio waves, similar objects at far greater distances would still be detectable by radio telescopes. In other words, the radio astronomers could probe much further into space and time, and settle the cosmological controversy.

This, in effect, is what the Cambridge team under Professor Martin Ryle are now confident that they have done. Since 1958, the team have been using a new and extremely powerful radio telescope. With this, they have made the most detailed survey and analysis of parts of the radio sky yet undertaken. On it, they have based the elaborate studies presented to the Royal Astronomical Society last week. The work is intricate, and the argument involved – but the gist of it is this:

If, like the radio telescope, we could perceive the radio waves from the sky, we would see many bright radio stars on a dimly luminous background of general radiation. We would ask, straight away: what are the radio stars, how far away are they, and what causes the background?

The first move was to direct optical telescopes at radio stars, and see if they are associated with any visible objects. Most of them, it turns out, are not. A few radio stars have been identified with visible objects inside our galaxy; a few more represent radio emission from nearby galaxies; and a rather larger number have been identified with extremely faint objects, some of which are probably colliding (or splitting) galaxies like the Cygnus source. These are a million times more powerful radio emitters than our own or nearby galaxies.

But there remain a very large number of radio stars for which no corresponding visible objects can be found. They could be nearby objects, which for some reason give out radio waves but little or no light. Or they could be very distant objects like the Cygnus source – in which case they might be used to settle the cosmological question. How are the two possibilities to be distinguished?

Suppose first, the Cambridge team said, that most of the radio stars are not far away, but are associated with our own galaxy. Suppose, too, that they are all powerful enough to be detected by the radio telescope. In this case, since we are situated towards the edge of our galaxy, we would expect to see more radio stars towards the galactic centre than towards the edge.

One Cambridge study, therefore, was to analyse the distribution of the radio stars in the sky – and it was shown conclusively that they are distributed evenly.

Does this prove that the radio stars are beyond the galaxy? No – there is another possibility: they might be very weak radio emitters, so that the radio telescope detects only the nearest ones, *well inside* the galaxy. These would then represent a small sample of the total radio star population of the galaxy. The rest of this population, too faint to be distinguished individually, would contribute to the general luminous background.

If we assume the detectable radio stars are all very weak, they must all be very near, and hence packed rather densely round us. But the more distant, undetectable ones must be equally closely packed – so the galaxy as a whole must be very densely populated with these weak, radio emitting objects. This large population would contribute to the luminous radio background – and would make it very bright. In fact, measurements of the background show that it is nowhere near bright enough, and it turns out that the only way of reconciling the observed background with the observed number of radio stars is to suppose that most of these stars are *outside* the galaxy.

This argument was then extended beyond our galaxy, and it was proved that most radio stars are not only outside the galaxy, but are very far away. The majority of them, in fact, must be rare, very powerful sources, comparable to the Cygnus source, but much more distant. The sources now being detected at Cambridge, in fact, are providing a picture of the universe as it was some 8,000 million years ago – and the sources themselves are receding at something approaching nine-tenths of the speed of light.

The final step at Cambridge was to work out the cosmological implications of this. On the big bang theory, the radio star population should get denser as you go back in time – there should be more of the faintest, most distant sources. On the 'steady state' theory, the population density should stay the same. The Cambridge observations have shown now unmistakably that the density of weaker sources is at least three times, and probably ten times, higher than the 'steady state' theory predicts. Therefore, the universe must be evolving.

JOHN DAVY, from an article in the *Observer*, 12 February 1961[1]

[1] *The* Sunday Times *of the same date reported Professor Fred Hoyle as being 'more confident than ever' in the theory of continuous creation. One possible alternative interpretation of the Cambridge results, according to Hoyle, was that a higher proportion of the very old, or very distant, galaxies might well be of the peculiar type which acts as a powerful source of radio waves. However, by 1964 he was postulating as a more likely source of these radio waves the immensely distant super stars, quasi-stars or quasars as they are variously called, which are central to his latest theories. These are condensations of cosmic dust such as would, in the normal run, break up into separate blobs or stars and form galaxies, but which for some reason or other continue to contract as a whole, and so form gigantic single stars which are imploding under the enormous force of gravity they generate. These implosions might, thinks Hoyle, squeeze out great jets of matter which could be the immensely powerful sources of radio waves detected by radio telescopes.* (*Ed.*)