1

Introduction

The discovery of galaxies

In the medieval world picture the stars were regarded as points of light attached to a sphere, whose surface was a long way outside the Solar System but whose volume was thought to be very much smaller than the space which we now know the stars to occupy. Attempts had indeed been made to determine the distance to the stellar sphere based on the possibility that stars might appear to be in different directions when observed from two points on the Earth's surface (fig. 1). Although this method worked for the Sun and Moon and other objects in the Solar System, it failed for the stars, indicating that they were very distant. After the Scientific Revolution in the 16th and 17th centuries, culminating in Newton's explanation of the motion of the planets in terms of a universal law of gravitation, it was realised that the stars were probably also suns or equivalently that the Sun was but one star amongst many and that the *fixed stars* should, in fact, be moving through space and should be influenced by the same law of gravitation. This led to a renewed interest in trying to determine not only their positions but also their motions.

Initially it was thought that there was just one system of stars filling the Universe and this view persisted until the second decade of this century. Earlier than that it had become apparent that the stars are not uniformly distributed in space as, to a first approximation, the stars visible to the naked eye appear to be. At the end of the 18th century William Herschel made the assumption that all stars have the same absolute light output as the Sun and concluded that the visible system of stars had an edge. Although his assumption about stellar luminosities was incorrect, his conclusion was not. By the beginning of the 20th century it was known that the system of stars which we now refer to as our *Galaxy* was highly flattened, most of the stars falling close to a plane which is defined in the sky by the faint band of light known as the *Milky Way*. At the same time it was known that distributed around the sky there were large numbers of fuzzy luminous objects which were clearly not

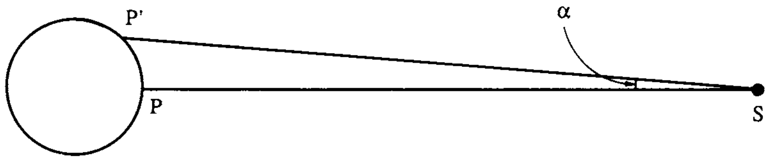

Figure 1. An attempt to measure the distance to a star. If stars were sufficiently close to the Earth, it would be possible to measure the angle α between the direction of a star as seen from two points P, P′ on the Earth's surface. With α and PP′ known, SP could be calculated.

individual stars and which had been given the name *nebulae*. Herschel and his son John were pioneers in cataloguing nebulae and the largest list was the New General Catalogue (NGC) of Dreyer.

Some of these nebulae were subsequently shown to be true nebulae or clouds of gas, which was what their name was meant to indicate, and they are situated in the same region of space as the stars. In contrast, one group of nebulae in particular, known as the *spiral nebulae*, were in time discovered to be composed of faint stars. There was then considerable argument early in the 20th century about whether the spiral nebulae were situated in the outer regions of our Galaxy or whether they were independent stellar systems, possibly of similar status to our Galaxy. If their stars were in fact similar to nearby stars, it was obvious that they must be distant and independent stellar systems, since their stars appeared so faint. This was eventually shown to be true in the early 1920s and these nebulae were subsequently given the name *galaxies*.†

It is now known that the observable Universe contains thousands of millions of galaxies and this number may be a considerable underestimate because of the difficulty of detecting small faint galaxies and any galaxy which has a low surface brightness. The Galaxy is a large galaxy but certainly not one of the very largest. It contains well over 10^{11} stars but the largest galaxies probably contain 10^{12} or 10^{13} stars. Obviously all of these stars have not been counted, although with modern telescopes and instruments which count stars automatically, many millions of stars can be counted in our own Galaxy. Very detailed studies have been made of limited regions of the Galaxy. Although galaxies contain other matter in the form of *interstellar gas* as well as stars, it appears that most of the visible mass in galaxies is in the form of stars, at least at the present time. We can therefore, to a first approximation, regard galaxies simply as systems of stars. It will however become clear later that galaxies also contain much invisible matter and that this matter may be neither stars nor gas. It is generally believed that much of it is composed of elementary particles. These particles, both inside and outside galaxies, may be the main form of mass in the Universe and may play a key rôle in the formation and structure of galaxies. Both stars and galaxies are held together by the force of gravitation and one of the most important questions in astronomy is *why matter in*

† Our Galaxy will always be spelt with a capital G and will be called *the Galaxy* or *our Galaxy*. The name galaxy is derived from the Greek name for the Milky Way.

the Universe is arranged in objects of galactic mass which have inside them sub-units which are objects of stellar mass. Although I shall not be able to answer this question in this book I shall try to present the evidence upon which an answer must be based.

Much of the detailed knowledge which we possess about stars has been obtained because we are fortunate enough to be close to one particular star – the Sun. In the same way we might hope to learn much about galaxies by a study of the Galaxy. There is, however, one very important distinction; whereas we are *near* to the Sun, we are *inside* the Galaxy. It proves quite difficult to discover the structure of an object from the inside. As an example, for a long time it was believed that the Sun was very near to the centre of the Galaxy because the apparently bright stars are approximately symmetrically placed around the Sun. It was subsequently realised that, as will be described further below, the interstellar gas in the Galaxy absorbs starlight and behaves as an interstellar fog, which gives a totally wrong impression about distances inside the Galaxy.

Differences between stars and galaxies

Although both stars and galaxies are objects which are held together by the attractive force of gravitation, they differ in many important respects both qualitative and quantitative. One simple observational fact is that, whereas the majority of stars are spherical or depart only slightly from spherical shape, galaxies exist in many shapes from essentially spherical to ones which are highly flattened; some of the latter are spheroidal but others have a much less symmetrical structure. Many highly flattened galaxies are observed to rotate rapidly. The great variety in galactic shape indicates that the classification of galaxies may be much more complicated than the classification of stars. Those properties of the classification of stars which are required in this book will be introduced as they are needed; the classification of galaxies is discussed in Chapter 3.

Another very important difference between stars and galaxies is that there is at present no clear evidence that there exist galaxies of significantly different ages; they may almost all have been formed between 10^{10} and 2×10^{10} years ago, with the actual spread of ages being much less than either of these figures. We shall, however, also see that the formation process of a galaxy may be very prolonged so that there is not a sharp distinction between formation and evolution. This is totally different from the case of stars in the Galaxy, where we know that some are essentially as old as the Galaxy while others are no more than a few million years old and star formation is certainly still continuing in the Galaxy today. It has proved possible to study stellar evolution not by observing the changes with time in the properties of an individual star but by investigating the properties of stars of similar masses but of different ages in our neighbourhood of the Galaxy. As all the galaxies near to us are about equally old, it appears that our only hope of a direct study of galactic evolution is to look at the most distant galaxies in the Universe and to compare their properties with those of nearby galaxies. Because light has taken a considerable time to reach us from them, we see them as they were in the remote past and, if the average properties of galaxies change as they age, we might hope to obtain some information about this change. At present the information that has been obtained is not very clear. One reason is that the distant galaxies appear very faint and are difficult to study. Another reason is that the assumption that galactic properties do *not* change significantly with *age* tends itself to be used in deriving the *distance scale of the Universe,* as will be discussed later in this chapter. What is clear is that the timescale of significant galactic evolution appears to be comparable with the believed *age of the Universe.*

The evolution of galaxies

There are two independent reasons which suggest that the Galaxy (as well as most other galaxies) may have been much more luminous early in its life history than it is now. If the Galaxy was formed by condensation from intergalactic gas, as is generally thought likely, it must initially have had a much greater total energy than it at present possesses. Thus a collapsing cloud has sufficient energy to expand again to its initial size assuming no dissipation of energy occurs and, if it is to settle down as a relatively compact object compared with its initial size, a considerable amount of energy must be lost from the system. However the energy is lost, it must be lost during the initial collapse phase which may last only a very short time compared to the total life of the Galaxy. This problem of galactic formation will be discussed further in Chapters 7 & 8. The second reason, which relates specifically to our own Galaxy but which almost certainly applies to other galaxies as well, is concerned with its chemical composition and this will be discussed in Chapter 7. If, as seems plausible, the original chemical composition of the Universe was a mixture of hydrogen and helium and if the heavier elements have all been produced by nuclear reactions in stars or more massive objects in the galactic lifetime, the observed variation of stellar chemical composition with stellar age suggests that there must have been a rapid burst of production of heavy elements on a galactic scale early in galactic history. This also suggests a very high early galactic luminosity. We shall, however, see in Chapter 7 that there are some suggestions that these conclusions may not be correct and that the initial collapse phase must be much longer than was originally believed.

Determination of the structure of the Galaxy

I now consider in slightly more detail how the present structure of the Universe has been deduced. As I have already mentioned, the discovery of Newton's law of universal gravitation and its application to the motion of bodies in the solar system led to the realisation that stars should be both distributed at different distances through space and moving. This stimulated attempts both to determine the distances of stars and to study their motions. It was not very long before Halley discovered the motion of stars across the sky through the background of (presumably) more distant stars (fig. 2). This effect, which was given the name *proper motion*, was discovered by comparing the positions of stars in the sky

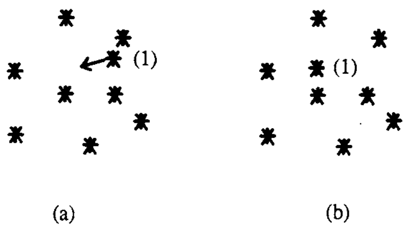

Figure 2. Proper motion. In (a) the star marked (1) is moving in the direction of the arrow. As a result, at a later time in (b), its position relative to more distant or more slowly moving stars has changed. Its angular displacement is called its proper motion.

according to the observations of the 17th century with those made by the Greeks many centuries earlier. It is now realised that there was an error in Halley's procedure which vitiated his results but reliable proper motions were subsequently measured. The discovery of proper motions will become very much easier as observations accumulate over a very long period. There is some advantage if successive observations are made with the same telescope, so that photographic plates can be compared directly, and the need for such observations has ensured that some very old telescopes have remained in use long after they would otherwise have been scrapped. However, very accurate modern measuring techniques are making it easier than it was to compare results obtained with different telescopes. Both proper motions and parallaxes (defined below) are being measured by the HIPPARCOS satellite launched in 1990. Unfortunately the satellite did not get into its planned orbit and as a result its lifetime will be reduced. Nevertheless it appears that very valuable results are being obtained.

It took very much longer to measure the distance to any stars. Both Newton and Herschel estimated the distances to stars by assuming that they were all really just as bright as the Sun, so that their apparent brightness was a measure of their distance. Because the Sun is a fairly average type of star, these estimates turn out to be very good for many stars although they are very bad indeed for some very bright or faint stars. The estimates were good enough to show that stars are extremely distant in comparison with the size of the solar system. However, the first direct measures of stellar distance were not made until three astronomers obtained distances to different stars in 1838/39 by the method of *stellar parallax*† (fig. 3); the parallax is the angle subtended at the star by the radius of the Earth's orbit around the Sun. For all stars that have so far been discovered it is less than 1″.

Once the distance of a star is known as well as its proper motion, the angular velocity across the sky can be converted into a *tangential velocity,* usually expressed in km s^{-1}. Finally the development of the science of spectroscopy in the second half of the 19th century led to the measurement of stellar *radial velocities* by use of the Doppler effect (fig. 4). By the early years of the 20th century it had

† The three stars were 61 Cygni measured by F.W.Bessel, α Lyrae (Vega) by F.G.W. Struve and α Centauri by T. Henderson.

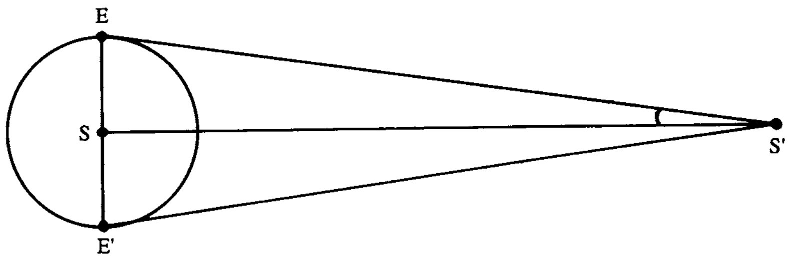

Figure 3. Parallax. If a star S′ is observed from opposite sides E,E′ of the Earth's orbit around the Sun S, the angle ES′S is called its parallax. If the angle is measured, the distance to the star can be determined.

become apparent that the stars in the neighbourhood of the Sun were typically a few light years† apart and that they had velocities relative to the Sun of a few tens or in some cases a few hundreds of kilometres per second.

By that time it had become clear that the Sun was one member of a large dynamical system of stars and it was believed that it was very close to the centre of the system. The reason for that belief is as follows. I have already remarked that the stars visible to the naked eye are approximately symmetrically placed about the Sun in space. In addition the Milky Way is a belt of fainter and more distant stars. This led to the picture of the Galaxy shown in fig. 5. A central, approximately spherical, distribution of stars is surrounded by a belt of more distant stars in the form of a torus; that is, shaped like a car tyre inner tube. This view of the Galaxy (or the Universe as it was then supposed to be) was put forward by Eddington in his book *Stellar Movements and the Structure of the Universe* (1912). Only ten years later a very different view of the Galaxy was gaining acceptance; this is shown in fig. 6. This shows an edge-on view of the Galaxy with the Sun very far from the central nuclear bulge. How could such a dramatic change of view have come about in such a short time?

The distribution of globular clusters

One clue to the change of view is given by the *globular star clusters* which are shown in fig. 6. These are compact groupings of stars which contain perhaps 10^5 to 10^6 stars each. The American astronomer Shapley pointed out that they do not appear in all directions in the sky, as is obvious from the position of the Sun and Solar System in fig. 6, but that they do appear to be part of the Galaxy. He suggested that the globular clusters form an approximately spherical system centred on the centre of our Galaxy. He then found that he could make sense of the observations only if the Sun were a large distance from the centre of the Galaxy (approximately 15 kpc) comparable with the radius of the system of globular clusters. This is shown in fig. 6.

† Throughout this book we shall use one or other of two measurements of distance. The *light year*, the distance travelled by light in one year, is 9×10^{15} m and the *parsec*, the distance at which a stellar parallax is one second, is 3×10^{16} m. Inside galaxies the useful unit is 10^3 pc(kpc), between galaxies 10^6 pc(Mpc) and for the whole observable Universe 10^9 pc(Gpc).

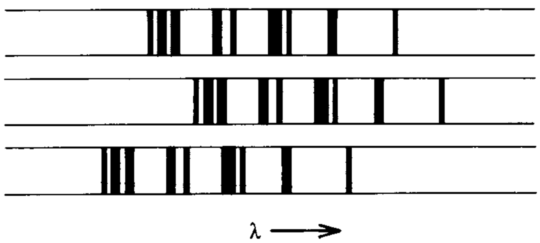

Figure 4. The Doppler shift of stellar spectra. The three spectra represent respectively that of a star at rest relative to the Sun, one receding and one approaching the Sun. λ is wavelength.

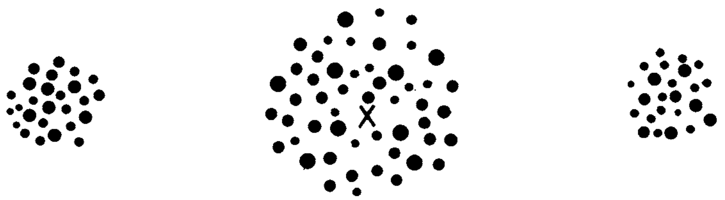

Figure 5. Eddington's (1912) picture of the Galaxy. The Sun is marked **x**.

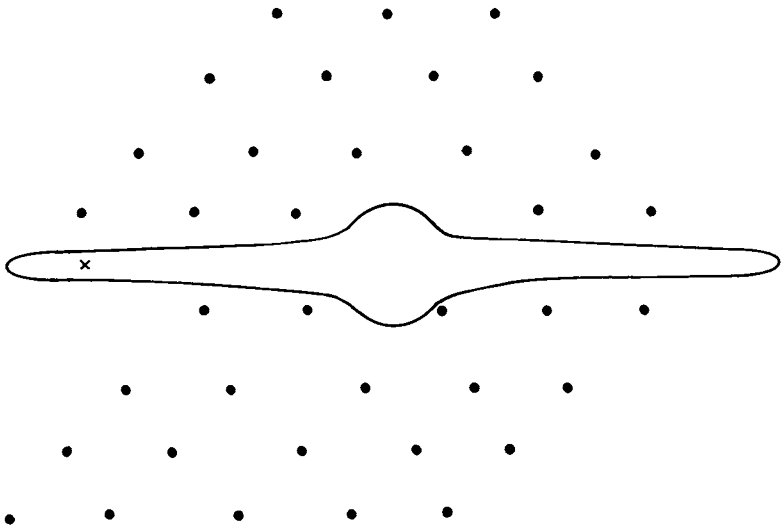

Figure 6. A schematic view of the Galaxy from the side showing the thin disk and the central nuclear bulge. The filled circles are globular star clusters and the Sun is marked **x**.

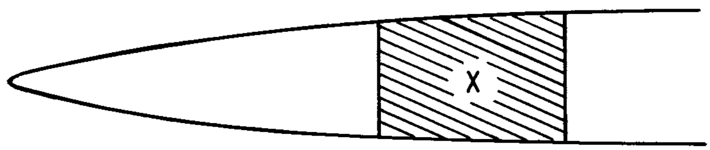

Figure 7. Effect of interstellar obscuration. Because of the interstellar obscuration it is only easy to see disk stars in the small hatched region around the Sun (**x**).

Interstellar gas and dust

Although this explains the observed distribution of globular clusters, it is now necessary to understand why this eccentric position of the Sun is not immediately apparent in our observations of stars. In particular, why is the sky not overwhelmingly more bright in the direction of the galactic centre than elsewhere? The answer to this question soon became apparent with confirmation of the existence of significant amounts of *interstellar gas and dust* which had been suspected much earlier. The way in which this gas and dust have been discovered will be described in Chapter 2; for the present it is sufficient to say that the interstellar matter absorbs and scatters starlight. The interstellar dust produces an interstellar fog, which makes it very difficult for us to see distant stars, so that the stars which we see with the naked eye or a small telescope are essentially only those near to the Sun (fig. 7). The gas and dust lie in the thin disk of the Galaxy, which also contains most of the stars. As many of the globular clusters are well out of the disk, as is indicated in fig. 6, there is not much interference with the study of the system of globular clusters. There is, however, *some* absorption of starlight in the direction of the globular clusters and, partly as a result, the present best estimate of the distance to the galactic centre (8.5 kpc) is significantly less than the value obtained by Shapley.

External galaxies

At the same time that the outline structure of the Galaxy was being resolved, there was dispute about whether the Galaxy was the whole Universe or whether there were other galaxies. As I have already mentioned, the dispute concerned objects known as nebulae. It had been known for a very long time that in addition to stars the Universe contained nebulous looking objects. The nebulae were studied by astronomers such as William Herschel and were catalogued by the French astronomer Messier towards the end of the 18th century. A much more complete list was the *New General Catalogue* produced by Dreyer almost a hundred years later. Even today many galaxies are known by a number given in one or other of these catalogues. The Andromeda galaxy is M31 and one of its satellite galaxies is NGC205. The nebulous objects were ultimately shown to fall into essentially four classes; star clusters (such as globular clusters), true nebulae (gas clouds), planetary nebulae (a cloud of gas surrounding a star, which has ejected it) and the group which we now know to be galaxies. The principal

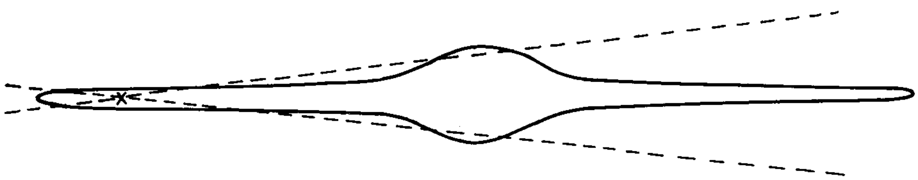

Figure 8. The zone of avoidance. Because of absorption by interstellar matter near the galactic plane, very few galaxies can be seen from the Sun (x) in directions between the two dashed lines.

examples of this group were known as *spiral nebulae* because of their spiral appearance, first discovered by the Earl of Rosse. Already, long before the spiral structure was known, an Englishman, Thomas Wright, and a German, Immanuel Kant, had suggested that some of the nebulae might be *island universes*.

There were two main reasons why the discovery that the spiral nebulae were independent galaxies was delayed. The first was that they are so distant that they could not be shown to be systems of stars until very large optical telescopes were available. This observational situation was transformed when the 100 inch Mount Wilson telescope came into operation in 1919. The second reason was concerned with the arrangement of the nebulae in space. The nebulae, unlike the bright stars, did not appear to be completely uniformly distributed around the sky. Instead they appeared to be largely absent in the direction of the plane of our Galaxy, giving rise to what was called the *zone of avoidance* (fig. 8). Very few nebulae could be seen in directions in and close to the plane of the Milky Way. Because through this observation the system of spiral nebulae appeared to be related to the structure of our Galaxy, there were very strong arguments that they must be part of the Galaxy or at most nearby satellites of it. These arguments were realised to be invalid when the discovery of the absorption of starlight by interstellar matter indicated that galaxies in the direction of the zone of avoidance could not be seen even if they were there. At about the same time, it became possible to see, on photographs taken with the 100 inch telescope, that the nearby spiral galaxies were composed of stars. If these stars were assumed to be similar to stars in our Galaxy, it was possible to estimate that nearby galaxies were typically several million light years away and consequently well outside our Galaxy. Thus, the idea of a Universe of galaxies became established.

The expanding Universe

Possibly the most compelling evidence for the independence of the external galaxies was obtained in the middle 1920s largely through the work of Hubble and his collaborators. I have already shown that the radial velocities of stars in the Galaxy can be measured by means of the Doppler shift of spectral lines. Hubble measured the Doppler shifts for many galaxies instead of stars. In the case of nearby galaxies, the Doppler shift could be either a redshift or a blueshift indicating that there are random motions in the system of galaxies similar to the random velocities of stars in the solar neighbourhood. In contrast the spectral lines of all of the more distant galaxies were shifted to the red. If the redshift is interpreted as being due to the Doppler effect, this means that all of the distant galaxies have a relative velocity directed away from us. This observation led to the idea of the *expansion of the Universe.* A discussion of the expanding Universe is the subject matter of *cosmology*, which I shall not study in detail in this book. It is not, however, possible to interpret the observations of distant galaxies or to discuss the structure and evolution of galaxies without taking some notice of cosmological problems, as I shall explain in both Chapters 3 and 8. Other suggestions have from time to time been advanced to explain the redshift. These include the *tired light* hypothesis that light loses energy and is redshifted simply by travelling large distances. Neither this nor any other suggestion appears to be as well-founded as the Doppler explanation and I shall assume without further discussion that the Universe *is* expanding. This assumption will only significantly affect the contents of Chapter 8.

The general picture of the Universe which was established in the 1920s is still generally accepted today. Many new types of object have been discovered inside galaxies, particularly by the use of observational techniques in new branches of astronomy such as *radio astronomy, infrared astronomy, ultraviolet astronomy* and *X-ray astronomy*. In fact the experience of the present century has been that, each time a new technique has become available, a whole class of exciting new objects has been discovered. Most of the galaxies in our neighbourhood have fairly regular shapes but recently many so-called peculiar galaxies have been discovered. These include galaxies which are apparently undergoing an explosion and others which have been distorted by the gravitational fields of their neighbours. In addition the *quasars* have been discovered. They have the largest redshifts of any objects known, they are extremely luminous and there is not as yet a full understanding of the source of the luminosity. The properties of peculiar galaxies and quasars will be described briefly in Chapter 3.

The Hubble constant

When Hubble discovered the expansion of the Universe, he also provided an estimate of the distance to the galaxies which he observed. It is now believed that his estimates are almost a factor of ten too low. I shall shortly explain how this has happened. This estimate can be summed up in one parameter, the *Hubble constant.* Hubble showed that the velocity of recession of distant galaxies appeared to be related to their distance from us by a law of the form

$$v = H_0 r, \tag{1.1}$$

where H_0 is known as Hubble's constant. With v usually measured in km s^{-1} and r in Mpc, the units of H_0 are km s^{-1} Mpc^{-1}; clearly H_0^{-1} has the dimensions of time and it (t_H) is in some sense *the age of the Universe*. I shall make this relation more precise in Chapter 8. An early value of H_0 obtained by Hubble was

$$H_0 \approx 550 \text{ km s}^{-1} \text{ Mpc}^{-1}, \tag{1.2}$$