I. Novikov

# 1: The Universe after the explosion

## *The world we live in*

In our journey to the world of black holes, we came across something seemingly impossible: the growing gravitational field virtually warped the properties of space and time and led to amazing physical processes. Now we are on a path towards quite different boundaries; we will go far into the depths of the Universe where we will again find the absolute reign of gravitation. Moreover, here we perceive a fact of immense significance: the observable Universe was born in the Grand Explosion ('Big Bang') that occurred about 15 billion years ago; the explosion was caused by a mysterious singularity, resembling the one hiding in any black hole.

Of course, the desire to comprehend the world in which we live has characterized man ever since people learned to think. The history of the evolution of man's picture of the Universe is both interesting and instructive. Numerous excellent books were devoted to this history, but our aim in this book is quite different.

Let us return to our time and to today's knowledge. If we sometimes have to refer to history, it is to the recent past; only infrequently will we visit the more remote past of the science of the entire Universe, called *cosmology*.

The first thing we encounter when trying to understand the nature of the Universe, is the distribution of celestial bodies in space.

208

We will be mostly interested in the largest scale accessible to astronomers, and begin with the largest structural units of the Universe: the galaxies.

The reader will remember that our Sun is an element of a huge stellar system that astronomers call the Galaxy with a capital G, and sometimes also *our Galaxy*. The total number of stars in the Galaxy is about a hundred billion.

The predominant part of stars in the Galaxy fill a volume resembling a lens 100 thousand light years in diameter and 12 thousand light years thick. (One light year is the distance covered by light in a year; it equals $10^{13}$km.) The interstellar space contains gas and dust that form large clouds. Their total mass comes to only about 5 per cent of the total mass of stars. In addition to this 'main body' of the Galaxy, it also includes a spherical component 5–10 thousand light years in diameter. As a rule, the stars in this spherical system are fainter and older than those in the flattened component.

The young hot stars of the flat system (sometimes referred to as the disk) form spiral arms. These arms begin in the central part of the Galaxy and spiral out to its farthest outskirts.

The spiral arms of our Galaxy led to the entire system being called the *spiral galaxy*. Spiral arms include vast accumulations of gas, the so-called gaseous clouds where young stars are being formed.

Stars and gas of the disk participate in the orbital motion around the center of the Galaxy on nearly circular orbits. Our Sun moves through the Galaxy at a velocity of about 250 km/s and completes one rotation in 200 million years. The stars of the spherical component also move around the center but their orbits are very elongated and lie at arbitrary angles to the plane of the disk.

Such are the structure and scales of the great stellar city, as our stellar system is sometimes called.

Other stellar cities – galaxies with lower-case g – lie beyond our Galaxy. The diameters of most observable galaxies are only slightly smaller than that of ours: they are tens of thousands of light years across and consist of billions of stars.

All these stellar systems lie at distances greater than millions of light years from us. Only the nearest to us and the largest are seen by the naked eye as blurred smears, others are observed only through powerful telescopes. Distances are so large that the light of the stars in them produces only a faint glow. The largest telescopes resolve individual bright stars only in the galaxies closest to us.

Galaxies differ in shape, composition of star population, and in the type of stellar motion. Astronomers classify galaxies into four main types.



Most galaxies are spiral, similar to our own. Some galaxies, however, have no spiral in the lens-shaped disk. Correspondingly, they are known as lens galaxies.

Third, a considerable number of galaxies are not disk-shaped and consist of only a spherical component. They are known as elliptic galaxies because on photographs and through a telescope they appear as elliptic ovals. As a rule, these stellar systems contain little gas and practically no regions where young stars are being born.

The class of irregular galaxies is the smallest. These resemble spiral galaxies in which bright clouds, where young stars cluster, do not form spiral arms but group into irregularly distributed spots. These galaxies often contain large amounts of gas.

Even this cursory outline shows the diversity of the world of galaxies. This diversity is even more striking when we compare masses and sizes of galaxies.

You may remember that our Galaxy contains about 100 billion stars. The largest (elliptic) galaxies may contain ten thousand billion stars. There exist at the same time 'dwarf' galaxies consisting of only a million stars.

What galaxy can be regarded as typical? A relatively large one, like our Galaxy, or a much smaller one?

This question is as difficult to answer as the following one: what city is typical, one as large as Moscow or one much smaller? Indeed, there are dozens of small towns for each large metropolis. The same picture is found in the world of galaxies. There are a large number of dwarfs in each giant system.

What is the distribution of galaxies in space?

It was found that the distribution is extremely nonuniform. Most galaxies are in clusters. The properties of clusters of galaxies are as diverse as those of galaxies themselves. To arrange them in an acceptable order, astronomers devised several classifications. As always in such cases, no single classification can be regarded as complete. For our purposes here, suffice it to say that clusters can be subsumed under two classes: regular and irregular clusters.

Regular cluster often have tremendous mass.They are spherical and consist of tens of thousands of galaxies. As a rule, all these galaxies are elliptic or lens-shaped. One or two giant elliptic galaxies are at the center. A regular cluster closest to us lies in the direction of the Coma Berenices constellation, at a distance of three hundred

709

million light years; its diameter is more than ten million light years. Inside this cluster, galaxies move relative to one another at velocities of the order of a thousand kilometers per second.

The masses of irregular clusters are much less impressive. The number of galaxies in them is smaller by a factor of several tens than in regular clusters, and these are galaxies of all types. Cluster shapes are irregular, with subclusters lying inside large clusters.

Irregular clusters may be quite small, down to tiny groups consisting of several galaxies.

What is observed on scales still greater than galaxy clusters? Are there clusters of clusters of galaxies, that is, their superclusters?

Recently the American astrophysicists Peebles, Gregory and Thomson, and Einasto, Saar and Jēveer in Estonia in the USSR, discovered that the largest-scale nonuniformities in galactic distribution are observed as 'cellular structure'. The 'cell walls' contain many galaxies and clusters of them, while the volume within the cell is empty. The cell size is about 300 million light years, the walls being about 10 million light years thick. Large clusters of galaxies are found at the nodes of the cellular structure. Superclusters are certain fragments of this cellular system. They are often considerably elongated into 'filaments' or 'noodles'. Superclusters comprise here and there giant voids that are almost free of any luminous material. What is the next, higher echelon in this hierarchy?

Here we find a new phenomenon. So far we have encountered more and more complicated systems: small systems formed a larger one, larger systems formed a still larger one, and so forth. The Universe resembled a Russian matreshka wooden doll: a small doll inside a bigger one, which is enclosed in another, still larger, matreshka. What was found was that the Universe has the largest matreshka of them all! No larger system is formed of the large-scale structure of 'noodles' and 'cells', the Universe being uniformly filled (on the average) by these elements. On the largest scale (greater than three hundred million light years), the Universe is found to be uniform, that is, to have everywhere the same properties. This is a very important property and at the same time one of the riddles that nature holds for us. For some unknown reason, relatively small scales reveal large clumps of matter, namely, celestial bodies and their systems that get more and more complicated, but the structure disappears on very large scale. This resembles sand on a beach. Close to the surface, individual grains are resolved, but if one looks from a



large distance at a large area, the sand is perceived as a uniform mass.

The uniformity of the Universe was established by observations for distances up to ten billion light years!

We will return to the uniformity puzzle later and now take up the question that in all likelihood worries the reader. How was it possible to measure such staggeringly large distances to galaxies and their clusters, and to operate, with such certainty, with the masses and velocities of galaxies?

### *'Measurement scale' and other tools of astronomers*

We will begin with distances. There can be no doubt that the measurement of distances of millions, even billions, of light years has been a miracle performed by modern science.

The measurement of such distances had been out of the question, say, at the beginning of the century. What, then, were the 'measuring sticks' that made it possible to penetrate regions that lie at such unimaginably large distances?

This was an exceptionally hard road for scientists. The progress to gradually greater distances was achieved step by step. Each next step was always based on the success of the earlier one.

The first serious step was made in the middle of the last century. The distances to three nearby stars were measured at practically the same time in Russia, Germany, and Africa. In principle, the basis of these measurements was the same as in measuring distances on the Earth by using a rangefinder. Rangefinders are now even installed in photographic cameras and are thus familiar to everyone. Essentially, the direction to an object sighted through this instrument is somewhat different when viewed through different windows. If the distance between windows (this is called base length) and the angle between directions is known, the distance to the object is easily calculated by using trigonometric formulas. In the rangefinder, this calculation is performed by an elementary mechanical device. The farther the object, the greater the spacing between the windows must be for the measurement to be reliable. This is known as the trigonometric method. When measuring the distances between stars, astronomers use the diameter of the Earth's orbit around the Sun as the base length. The direction to the star is measured at an interval of half a year from diametrically opposite points of the orbit. Even with this enormous base length, the change in the direction to nearby

210

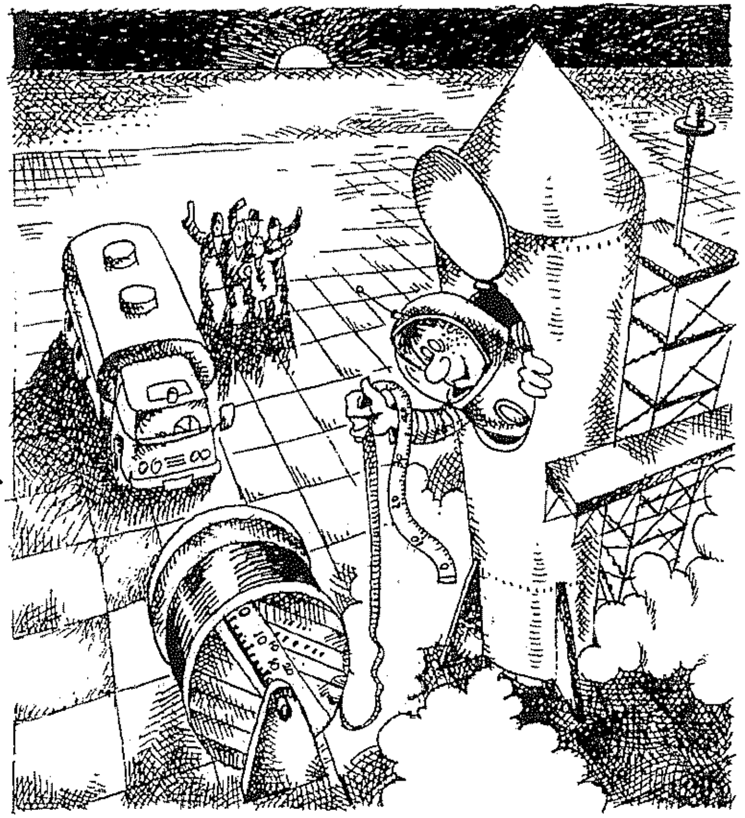
stars does not exceed one second of arc, so that measurements require the highest skill and greatest care.

It was found that even the nearest stars lie at distances greater than one light year.

More than a century has elapsed since the pioneer measurements of distances to stars. Despite the tremendous progress in instruments and measurement techniques, a hundred light years remains the maximum distance reliably measured by the trigonometric method.

One is still unimaginably far from the boundaries of the Galaxy, let alone from other galaxies.

The next significant step (rather, leap) up the staircase that leads far away was made at the beginning of this century; it was essentially based on employing stars whose luminosity varies systematically, that is, variable stars.





It all began with the work of the American astronomer Henrietta S. Leavitt, who studied variable stars of one of the galaxies closest to us: the Small Magellanic Cloud visible in the Southern skies.

Several years after the study started, Leavitt came across an intriguing fact. Twenty five stars were found to be variable, with luminosity being a strictly periodic function of time. Moreover, the greater the luminosity period, the brighter the star was! Leavitt came to a spectacular conclusion: 'Since the variables are probably at nearly the same distance from the Earth†, their periods are apparently associated with their actual emission of light, as determined by their mass, density, and surface brightness.'

It would be difficult to overestimate the importance of this discovery. It became possible to find the luminosity of a star once the period of luminosity variation was known.

We know that the apparent brightness of a star decreases in inverse proportion to the squared distance to the star. A comparison of the true luminosity of a star with the apparent brightness then gives the distance!

Actually, in order to calculate the distance from the period of brightness variations by Leavitt's data, it is necessary to know the true luminosity of at least one such star.

The first attempt to implement this idea was made by Herzsprung. He understood that the stars described by Leavitt in the Small Magellanic Cloud are nothing other than the familiar stars, called cepheids, of our Galaxy. The brightness of cepheids varies because these stars are pulsating. Now it was necessary to determine the true luminosity of at least one cepheid. This is where serious difficulties began. There is not a single cepheid in the neighborhood of the Sun, such that the distance to it could be reliably determined by the trigonometric method and then the true luminosity found for the known brightness and distance.

Numerous attempts were made to find distances to cepheids in our Galaxy. The first estimate was that of Herzsprung himself. We will not describe here the clever indirect techniques he had used. Let me only remark that both the first and many subsequent attempts were so difficult to perform that the results contained considerable errors. These errors were finally ironed out only at the beginning of the 1960s. In fact, this work is so important (it is The Scale of the Universe that we are after!) that elaboration never stops.

† Because all of them are in the same galaxy, i.e., in the Small Magellanic Cloud.

211

Once the true luminosity of at least one cepheid with known period of brightness variation had been evaluated, it became possible to determine the distance to any cepheid. Indeed, now the function 'period versus true luminosity' for cepheids had been established. To find the distance to any cepheid, it is sufficient to determine its brightness variation period by observation, obtain the corresponding true luminosity, compare it with the apparent brightness, and calculate the distance. If the cepheid is in a cluster or a galaxy, the result is the distance to this system.

Here cepheids are employed as 'standard candles' whose true brightness is known. The technique as described above was accordingly called the 'standard candle' method.

The role of cepheids in measuring distances is so outstanding that the well-known American astronomer Shapley called them 'the most important' stars.

The true luminosity of cepheids is very high: they are a thousand times brighter than our Sun. Therefore, they are observable from sufficiently large distances, up to 15 million light years. They are thus suitable for measuring distances to the nearest galaxies.

However, we need to measure still greater distances!

Another step has to be made for further progress. It is desirable to identify 'standard candles' that are brighter than cepheids and thus visible from much farther off. Such 'candles' have been found. Galaxies are usually surrounded by stellar clusters that are called globular because of their shape.

When cepheids made it possible for us to find distances to the nearest galaxies, true luminosities of globular clusters around different galaxies were compared. It was found that if one selects the brightest globular cluster at each galaxy, the true luminosity of these brightest clusters is practically the same for all galaxies.

Hence it was possible to use globular clusters as the 'standard candle' that is brighter than cepheids.

This method evaluates distances up to 60 million light years. This is sufficient for 'reaching' the nearest clusters of galaxies. Alas, so far we cannot resolve globular clusters at distances above this limit.

The next step involves a still brighter 'standard candle'. It was discovered that the brightest galaxies in different clusters of galaxies have very nearly the same luminosity: about ten times that of our Galaxy.

These brightest 'standard candles' take us out to billions of light years.



Such is the 'hierarchy of scales' that astronomers use on the way to the depths of the Universe.

How are the velocities of the farthest objects measured?

No doubt, any displacement of stars and other objects, from which the velocity at right angles to the line of sight could be calculated, is immeasurably small at the distances that separate us from the nearest galaxies and even more so from those farther away.

The only velocity that can be measured, and measured simply and reliably, is the velocity of accession or recession of celestial bodies with respect to us. This measurement is based upon taking into account the Doppler effect that we have already mentioned in the book. When a celestial body moves towards us, the light it emits is shifted to the blue end of the spectrum, and when it moves away from us, to the red end. The spectroscopic measurement of these shifts makes it possible to calculate the velocity, or rather its component, along the 'line of sight'. Astronomers refer to velocities found from the Doppler effect as 'line-of-sight velocities'.

Finally, let us look at the masses of galaxies and clusters of galaxies. These can be evaluated by using the law of universal gravitation.

Assume that we observe, say, an elliptic galaxy. Stars in it move at certain velocities with respect to one another. They would scatter all over space were it not for gravitational forces. Gravitation due to the entire mass of the galaxy prevents them from running away. Having measured the relative velocities of stars in a galaxy (this can be done by Doppler techniques) and knowing the size of the galaxy, we can evaluate gravitational forces and, hence, the mass generating them.

In measuring the mass of galactic clusters, a similar approach is used, but instead of the motions of individual stars, one operates with the motion of galaxies in a cluster.

Now the reader realizes, in general, how astronomers obtained the numbers that describe the arrangement of the Universe on a large scale.

Another question arises. What is the motion of galaxy clusters and individual remote galaxies?

The answer to this question was the greatest discovery of twentieth century science. It was found that we live in an expanding Universe. Clusters of galaxies recede from one another, and all the matter in the Universe had been set in the state of expansion by a mysterious Big Bang that had occurred in the very remote past.

212

## The Universe has to evolve

The Universe in which we live has to either expand or contract: this was predicted theoretically by the famous Soviet scientist Aleksandr Freidmann in 1922–4. Friedmann's work was rigorously mathematical, based on Einstein's theory of gravitation. Actually, we need not employ rigorous mathematics to understand the essential point of his discovery. Like all truly great discoveries, it is basically very simple.

You remember why an ordinary star neither contracts nor expands. Gravitational forces are balanced by the force produced by the pressure drop between the dense core of the star and its loose surface layer. In contrast to this, the Universe is uniform on the largest scale, so that no pressure drops are possible. Therefore, gravitation is the only significant force there.

As a result, if the vast masses of the Universe are imagined to be at rest with respect to one another at some moment and to be uniformly distributed, gravitation will immediately set them in motion and the very next moment the matter will start to contract. Gravitation can be balanced out in relatively small systems by the circular motion of bodies on orbits (as in the Solar System) or by random motion of bodies on very elongated orbits (as in elliptic galaxies). However, this is impossible in the all-embracing Universe: some bodies would have to be assigned velocities greater than the velocity of light, in violation of nature's laws.

Friedmann's conclusion was that a stationary universe is impossible. Actually, the Universe does not necessarily have to contract under the influence of the gravitational force. If all masses are first given outward velocities, the Universe expands and the gravitation only slows down the receding bodies. It is thus the result of initial conditions, or rather of the physics of the processes that dictated the initial velocities of masses, whether the Universe will contract or expand. Thus the inevitability of the global evolution of the Universe was theoretically discovered.

This idea was absolutely novel and extremely unusual. A number of different schemes describing the structure of the Universe reigned in science, replacing one another in the history of mankind. One feature was common to all (or nearly all) of them: these were the structural schemes, presenting the eternally unchanging 'mechanism of the Universe's clockwork' but not the development, the evolution, or the maturing of this Universe. The notion of stationarity of the Universe seemed self-evident. Most complex properties could unfold in the Universe but why, from what state, and in what direction was the Universe to evolve?



The idea of the evolution of the entire Universe appeared to be nonsensical, or hardly acceptable to even first-class scientific minds. The great Einstein is one example. The creator of general relativity realized how important his theory was for cosmology. Immediately after he completed the development of general relativity, he tried to find out whether the equations of the theory, when applied to the entire Universe, had a static solution, that is, a solution describing a time-independent state. It appeared obvious to Einstein that he had to look for a static, not evolutionary, model of the Universe. But static solutions could not be found when general relativity equations were applied to the Universe. The idea of a static world was so compelling that Einstein lost faith in his equations and even tried to modify them so as to make them have stationary solutions. We will discuss this attempt later.

What made this idea of static Universe so attractive?

In all likelihood, it agreed with the apparent stationarity, and invariable nature of astronomical bodies, be it the Solar System, stars, clusters of stars, or galaxies. The observed constancy of astronomical phenomena on all scales familiar to people was willy-nilly extended to the Universe as a whole. Aristotle had formulated this attitude very clearly in his treatise *On the Heavens*: 'Throughout all past time, according to the records handed down from generation to generation, we find no trace of change either in the outermost heaven or in any one of its proper parts.'

Today, at the end of the twentieth century, we have grown accustomed to the idea of a changing Universe. We realize that stars and other celestial bodies and their systems only seem to stay unchanged. Man observes them for too short a time to be able to notice their evolution. In fact, stars are born, they live, and then die. Their life spans often reach billions of years. The energy radiated by stars is supplied by nuclear reactions burning in their cores. No store of energy is unlimited; the store of nuclear energy is also finite. Hence, the Sun and the stars appeared in a finite past and had a certain history.

Today we are able to observe violent processes of explosion and evolution in such giant systems as galaxies. The matter in galaxies is gradually transformed in nuclear processes within stars. Hydrogen is converted into helium, and then into heavier chemical elements.

A stationary scenario is thus unacceptable for an astronomical system, provided we consider sufficiently long time intervals. If the model of the Universe had to be built anew, we would have to require that the model would be evolutionary, and that it would specify the epoch in which stars, galaxies, etc., began to form.

But let us return to the story of Friedmann's discovery. The first paper proving that the Universe must evolve was received by the editor of *Zeitschrift für Physik* at the end of June 1922. Einstein was so certain that the equations describing the state of the Universe had to have a static solution that he judged Friedmann's paper to be erroneous. In mid-September, *Zeitschrift für Physik* received Einstein's short note. In Fock's words, Einstein remarked in it,

> somewhat haughtily, that Friedmann's results seemed suspicious and that he found there an error; after correction, Friedmann's solution reduces to the stationary one.

Friedmann was informed of Einstein's attitude by his Petrograd colleague Krutkov, who was at the time on a mission abroad. In December 1922, Friedmann wrote a letter to Einstein, where he outlined details of his calculations and presented conclusive proof of correctness of his results. The letter ended with these words:

> Should you consider the calculation described in my letter correct, may I ask you to inform the editor of *Zeitschrift für Physik* of your opinion; in this case you may deem it possible to publish an addendum to your earlier note or to copy in the journal a suitable excerpt from my present letter.

The letter reached Einstein and has survived in his archive; it seems, however, that he failed to read it at the time or simply overlooked it, being quite sure that Friedmann was wrong.

It May 1923, Krutkov met Einstein in Leiden, in the house of the well-known Dutch physicist Paul Ehrenfest, and persuaded Einstein in a number of discussions that Friedmann was right. In a letter to his sister in May 1923, Krutkov wrote: 'Victory over Einstein in the argument concerning Friedmann. Petrograd's honor has been saved!'

Immediately after the meeting with Krutkov, Einstein sent the following note to *Zeitschrift für Physik* (I quote here the complete text):

> On Freidmann's paper 'On the curvature of space'
> My preceding note criticized the above-mentioned paper. Freidmann's letter communicated to me by Mr Krutkov made it clear to me that my critique reflected a miscalculation. I believe that Freidmann's results are correct and shed new light on the problem. It is found that the field equations allow, in addition to static solutions, also dynamic (i.e. variable in time) centrally symmetric solutions for spatial structure.



Later Einstein continued to emphasize the importance of Friedmann's work for the development of modern cosmology. He wrote in 1931: 'Friedmann . . . was the first to start on this road.'

### The discovery of the expansion of the Universe

Distant systems of stars (galaxies and clusters of galaxies) are the largest structural units of the Universe known to astronomers. These systems are observed over tremendous distances; it was the study of their motions that produced the observational basis for an analysis of kinematics of the Universe.

The measurement of radial velocities of galaxies was pioneered at the beginning of this century by the American astrophysicist Slipher. Distances to galaxies were not known at the time and heated debates raged among astrophysicists: do these objects lie within our Galaxy or far beyond its bounds? Slipher established that most galaxies (36 out of 41 he observed) are receding, at velocities up to two thousand kilometers per second. Only several galaxies were moving towards us. Later it was found that the Sun revolves around the center of our Galaxy at a speed of 250 km/s, so that the 'accession velocities' of these nearby galaxies simply reflect the fact that the Sun is currently moving towards these objects.

Slipher thus established that galaxies are receding from us. Lines in their spectra were shifted towards the red end. This phenomenon is known as the 'red shift'.

Distances to galaxies were determined in the 1920s.

In 1923, the American astronomer Edwin P. Hubble discovered the first cepheid in one of the closest galaxies in the Andromeda constellation. A year later he discovered more than ten cepheids in this galaxy and twenty-two cepheids in another galaxy in the Triangulum constellation.

Cepheids were discovered in other galaxies as well. Distances to these cepheids and, hence, to the galaxies in which they were found, proved to be much larger than the diameter of our own Galaxy. It had thus been finally established that these galaxies are remote stellar systems similar to ours.

Even the pioneering work on establishing distances to galaxies employed other methods in addition to cepheids. One of the methods involved using the brightest stars of a galaxy as distance indicators.

214

Apparently, the brightest stars have identical luminosity in our Galaxy and in other galaxies, so that they can serve as 'standard candles' for the determination of distance. Being brighter than cepheids, the brightest stars are seen from greater distances and thus offer a more powerful distance indicator.

A comparison of distances to galaxies with recession velocities (these were determined by Slipher and other astronomers and were only corrected for the motion of the Sun within the Galaxy) made it possible for Hubble to establish in 1929 a spectacular relation: the farther a galaxy lies from us, the greater its recession velocity. It was found that the recession velocities of galaxies and distances to them are very simply related: velocity is directly proportional to distance. The proportionality coefficient is now known as the Hubble constant.

According to Hubble's measurements, the galaxies at a distance of a million light years fly away at a velocity of 170 km/s.

Fifty years have elapsed since Hubble's discovery. The power of astronomical tools has increased dramatically, and new studies have confirmed Hubble's law, that is, the proportionality of the recession velocity of galaxies to the distance separating them from us. It was found, however, that Hubble greatly overestimated the proportionality coefficient.

It happened because there was an error in his evaluation of distances to galaxies. They happened to be reduced by a factor of six to ten. This is not surprising because we have already seen that in order to evaluate very large distances, one has to go up the steps of a long ladder, with errors possible at each step.

The main sources of errors were identified only after 1950 when the largest (at the time) 200-inch telescope of the Mount Palomar observatory was put in operation. In 1952, the American astrophysicist Baade established that cepheids of the type used by Hubble were actually about four times brighter than had been previously believed. This conclusion signified that the distances to the nearest galaxies, evaluated via cepheids, were in fact almost twice as large as earlier calculations suggested. Additional corrections tripled the distances to nearby galaxies. An error at this step of the ladder implied errors at the subsequent steps as well. All measured distances to more distant galaxies also had to be tripled.

Before the distance scale was thus reconsidered, all the nearest galaxies seemed to be substantially smaller than ours. This was strange. After re-evaluation, it became clear that quite a few galaxies are as large as ours and some are even larger. This result supported the belief that scale re-evaluation was correct.



At the end of the 1950s, it was found that further rungs of the ladder taking us deeper into the Universe also contained serious errors. Hubble also made mistakes in determining distances to far-away galaxies in which cepheids are not seen. There were two reasons for this. First, to calculate the visible brightness of very faint stars in other galaxies, comparison must be made with known standard sources. This is a very difficult problem, and errors were indeed revealed in the standard procedure.

The second cause of error was that Hubble had mistaken very bright clouds of ionized hydrogen in distant galaxies for the brightest stars (and employed them as 'standard candles'). Viewed from such distances, these clouds looked like bright star-like points. As a result, the distance scale to far-away galaxies was enlarged further by about 2.2.

When all these factors are taken into account, we find that all distances to the farthest galaxies are greater by a factor of six to ten in comparison with Hubble's estimates. A better evaluation cannot yet be given. The Hubble constant was also reduced by the same factor. According to current data, galaxies at a distance of a million light years recede at velocities of about 25 kilometers per second.

Having made these clarifying remarks, we can return to the principal importance of Hubble's discovery for our understanding of the structure of the Universe.

This discovery demonstrated that galaxies fly away from us in every direction and that the velocity of recession is proportional to the distance.

This fact is likely to be met with disbelief. Why is it that galaxies recede from no other object but our Galaxy? Are we really at the center of the Universe?

This conclusion is wrong. Actually, galaxies recede not only from our Galaxy but from one another as well. If we watched the world from a different galaxy, we would observe just the same picture of recession that we see from our stellar system.

To understand this, imagine two galaxies that move away from us in the same direction, and that the second galaxy lies at twice the distance between us and the first galaxy; correspondingly, the second galaxy recedes at twice the velocity of recession of the first. Imagine that we set our observation post on the second galaxy. It moves away from our original galaxy, and an observer here (who naturally considers himself to be at rest) regards our Galaxy as moving away from him, in the opposite direction, at the same velocity. The first galaxy, halfway between our Galaxy and the second galaxy, moves more slowly, so that an observer on the second galaxy decides that the first one moves in the same direction as our Galaxy but that its velocity is lower. This argument applies to any three galaxies.

Therefore an observer in any galaxy perceives the same pattern of galaxies running away from the point where he is stationed.

We can suggest another model to illustrate the situation. Imagine a homogeneous sphere and enlarge its dimensions, say, to twice the original size, preserving the homogeneity of the sphere. Obviously, the distances between any two points inside the sphere will also be doubled, regardless of the choice of these two points. Hence, wherever we place an observer inside the expanding sphere, he will see the same picture of points receding from him inside the sphere. If the sphere is taken to be of infinite radius, we obtain the situation described above, which is independent of the position of the observer.

The fundamental fact is that galaxies do fly outwards: the Universe is expanding. This is a spectacular confirmation of the conclusion of Freidmann's theory on the nonstationarity of the Universe.

The following question is sometimes posed. Assume that the entire Universe is uniformly filled (on average) with clusters of galaxies. The question would be: 'where' and 'into what' is the Universe expanding?

This question is essentially wrong. The Universe is the totality of everything that exists. There is nothing 'outside' the Universe: neither galaxies, nor any other matter, nor anything at all, be it space or time. There is no vacuum into which to expand. Actually, the Universe needs nothing outside it to allow its expansion. The following example will clearly illustrate this statement.

Let there be an infinite plane with points, representing galaxies, spread uniformly on it. Now we stretch this plane uniformly in all directions, so as to enlarge the distances between the points. The plane was originally infinite; where did it expand to? Obviously, these are the properties of the infinite. If we double it, we again get the same infinity.

Let us forget galaxies and the Universe for a while and devote some time to the notion of infinity, since it is crucial for our concept of the Universe.

The infinity is the subject of the branch of mathematics called *set theory*. Typically, people not looking into this field professionally



have a very fuzzy (and naïve) picture of infinity. Intuitively, one is inclined to think that infinity is what we get by indefinitely continuing the count 1, 2, 3 . . . . If this is so, do we need a theory of the infinite?

Actually, the properties of the infinite go far beyond the indefinite continuation of the sequence 1, 2, 3 . . . . Moreover, these properties are infinitely more varied and overwhelming than any properties of finite numbers or their groups.

We will describe some of them. Let us begin with a story ascribed to the famous mathematician David Hilbert (as rephrased by a contemporary mathematician).

Imagine a hotel with an infinite number of rooms 'enumerated' by the natural numbers

1, 2, 3, . . . .

All rooms are occupied. A new guest arrives late at night. 'We are full,' says the receptionist. 'That is of no importance,' objects the hotel manager. 'We will move the guest in room 1 to room 2, the guest in room 2 to room 3, the guest in room 3 to room 4, and so on, and offer room 1 to the new guest.'

A thousand more guests arrive the same night. 'We are full,' says the receptionist. 'No problem,' objects the manager. 'The guest in room 1 goes to room 1001, the guest in room 2 to room 1002, and so on, and the newly arrived guests are to be given the vacated rooms from 1 to 1000.'

The guests barely have time to settle in their rooms when a new wave of arrivals floods the hotel. This time the number of new guests is infinite; we denote them by $A_1$, $A_2$, $A_3$, . . . . 'We are full', says the receptionist. 'It's all right,' says the hotel manager. 'Move the guest in room 1 to room 2, the guest in room 2 to room 4, the guest in room 3 to room 6, that is, each guest will be asked to move to the room with twice the former number. Now we can offer rooms 1, 3, 5, . . . to the guests $A_1$, $A_2$, $A_3$, . . . .'

This story clearly shows that part of the infinite may be equal to the whole. Indeed, let us write the infinite sequence of even numbers as an infinite row, and place the sequence of guests' numbers as a row below the first:

2, 4, 6, 8, . . .
1, 2, 3, 4, . . . .

Each even number corresponds to one guest's number, and vice

216

versa. Hence, the number of even numbers equals the number of all elements of the natural number series. At a first glance, this conclusion disagrees with our intuition. Indeed, even numbers are only one half of the entire set of natural numbers. This is indeed so for any finite set of numbers. But all this changes when we turn to infinite sets, and a part may be found to equal the whole, as we can clearly see by comparing the two sequences written above.

Other examples given in Hilbert's amusing story point to a number of similar properties.

The above examples may give the impression that all infinities are 'identical', that is, that any infinite set of elements can be enumerated by the infinite set of natural numbers, as we have done with the set of even numbers.

This is definitely not so!

The famous mathematician Georg Cantor proved in the last century that the number of points on a segment of a straight line cannot be counted in any way. They cannot be enumerated by the infinite sequence of natural numbers, assigning one number to each point, regardless of the order in which the points are considered. There will always be at least one point left without a number!

This is not very difficult to understand. Indeed, imagine that we take a segment of unit length and characterize the position of each point by its distance from the left end (chosen as the origin). We write these distances as decimal fractions. In fact, the position of each point is written, generally, as an infinite decimal fraction (a fraction with an infinite number of decimal places after the decimal point). Of course, all decimal places beginning with some place may be zeros in some exceptional cases.

Now imagine that someone has succeeded, in contrast to our statement, in enumerating the points of this segment. We then arrange the decimal fractions that characterize the positions of these points on the segment into a table, in the order of their numbers. The first row has the infinite fraction for the position of the point given number 1; the second row has the infinite decimal fraction for the point given number 2, and so on. Our table may look, for example, like this:

0.328 697 008 33 ...
0.919 671 384 52 ...
0.000 637 011  4 ...
........



Let us show that there is at least one point on the segment that is absent from this list, so that the list is definitely incomplete.

To write the decimal fraction characterizing the position of this point on the segment, we do as follows. As the first decimal place after the decimal point we take any number that differs from the first decimal place in the first row of our table (in our example, this is not 3 but, say, 5). For the second decimal place of our fraction we choose any number not equal to that in the second decimal place of the second row of the table (in our example, not 1); we follow this procedure indefinitely. Clearly we get a fraction that is absent from our list. Indeed, it does not coincide with the first row because it definitely differs from it in the first decimal place, not with the second row because it definitely differs from it in the second decimal place, etc.

The point whose distance is given by this fraction is absent from our infinite list, and, hence, has no number.

One might think that enumeration should begin with this point, and other points are to be enumerated after it. As the Dutch mathematician Hans Freudenthal jokingly remarked, this was the strategy of the man who bet that he would eat 20 potatos. Having eaten 19 of them and feeling unable to chew the last one, the man sighed and complained: 'I should have started with this potato.'

Obviously, if we begin the enumeration with the particular point that originally had no number, we can find in just the same way another point that has no number in the new enumeration.

The reader may be rather tired from the need to follow this unusual construction but it is so important that I wanted to outline it in order to convey at least some feeling of how extraordinary are the properties that we find in the realm of the infinite.

The number of points on a unit-length segment of the line is thus definitely greater than that of natural numbers. Mathematicians say that the *cardinality* of the infinite set of points on a segment of a line is *higher* than that of the set of natural numbers.

Infinities are thus not all the same. Some have greater cardinality, that is, are richer in elements, and some lower.

It may seem that the number of points on the entire straight line is certainly greater than on a unit-length segment, since a segment is only a part of the line. But we are now prudent because we remember that the rule 'a part is smaller than the whole' does not work in the realm of the infinite. Indeed, the infinite sets of points on

217

a line and of its segment have the same cardinality. These are indistinguishable infinities!

Furthermore, the infinite set of points on the entire plane and even in the entire three-dimensional space has the same cardinality as the set of points of a segment of a line. All these infinities are identical. A suspicion may arise that since the set of points of the entire infinite space is not richer than the set of points of a line segment, a set with greater cardinality simply does not exist, that is, that this infinity is the greatest.

This guess is incorrect. Mathematicians are able to construct sets of progressively greater cardinality, that is, richer and richer infinities. In other words, the greatest infinity does not exist; this sequence is also infinite.

But let us proceed no farther than this first step into the world of the infinite. A journey into it may prove to be no less captivating than a voyage through the world of black holes or through the far reaches of the Universe, but it would nevertheless be a path to a different field of human knowledge.

Let us return to the expansion of the Universe. After the infinity stories above, we are no longer surprised that an infinite Universe can expand infinitely, needing for this exposition nothing that lies 'outside' the Universe, nothing of what is not the Universe.

As the infinite number of hotel guests in Hilbert's story could be distributed among even-number rooms only, thus doubling the distance between them, so the distance between galaxies in the Universe can be, say, doubled without going 'outside' the Universe.

Another important question arises, however: what is the reason for the Universe's expansion? What gave galaxies their velocities? The reader remembers that the theory of gravitation cannot answer this question. Galaxies are now moving out by inertia and their velocities are slowed down by gravitation.

We shall return to what caused the Universe to expand in the last chapter of this book.

One concluding remark is necessary here. It is sometimes said that owing to the expansion of the Universe, everything that exists is also expanding: galaxies not only recede from one another but they themselves expand, and so do individual stars, our Earth, and all other bodies. This is patently wrong. The recession of galaxies does not affect in the least the dimensions of individual bodies. Cosmological expansion does not act on gravitationally bound bodies, such as galaxies, stars, or the Earth, just as the expansion of a gas cloud does



not affect individual gas molecules. There is no doubt that cosmic objects may expand or contract but these changes are caused by internal factors, that is, processes occurring inside these bodies.

### *Is the Universe indeed expanding?*

The conclusion about the expansion of the Universe did not immediately gain general recognition. The idea of the evolution of the entire world was too grandiose. This idea leads to a number of awe-inspiring and far-reaching consequencies. It implies that in the very remote past, when the expansion began, the Universe looked very different from what we observe today. I mentioned at the beginning of this chapter that this idea met with numerous objections stemming partly from the inertia of scientific thinking and partly from preconceived pseudophilosophical arguments. The notion of an unchanging, static Universe seemed to be so much more habitual and nondisturbing. All this stimulated a number of attempts to give an alternative explanation, not based on the Doppler effect, to the observed 'red shift' in the spectra of remote galaxies. If these explanations were true, one could stop regarding galaxies as receding and the Universe as expanding.

This attitude is excellently reflected in the short witty piece of the Canadian humorist and economist Stephen Leacock, *Common Sense and the Universe*:

> For some twenty-five years past, indeed ever since the promulgation of this terrific idea . . . we had lived as best we could in an expanding universe, one in which everything, at terrific speed kept getting further away from everything else. It suggested to us the disappointed lover in the romance who leaped on his horse and rode madly off in all directions. The idea was majestic in its sheer size, but it somehow gave an uncomfortable sensation.

Attempts to 'defend' the stationarity of the Universe are sometimes made even nowadays. But what is the physical reality? Is there a physical process that causes photons to become 'redder' but does not involve the recession of galaxies?

In principle, we know such processes.

For a quantum to 'redden', it must lose some of its energy. This may happen when a quantum collides with electrons of the interstellar medium in its long voyage through the cosmos; in some versions of the hypothesis, it collides with other photons in intergalactic space. In all such hypotheses, an interacting photon not only loses some energy but also changes the direction of propagation. As a result, light quanta propagating right in our direction will gradually start moving on diverging trajectories. The images of remote galaxies would be slightly blurred.

No such effect is observed, and for this reason it cannot explain the 'red shift'. Besides, it would require the intergalactic medium to have fantastically high density and would produce a number of other observable effects.

As another mechanism of photon 'red-shifting', a hypothetical decay of photons was suggested, accompanied by the creation of yet unknown particles. The Soviet physicist Bronstein demonstrated as early as in the 1930s that if such a process did exist (it was later shown that it does not), the probability of photon decay would be inversely proportional to frequency. Hence, the longer the wavelength of a photon, the further it would shift to the red end. Quanta of radiowaves would be expected to shift to the red quicker than visible-light quanta. Precise measurements were carried out in the 1960s of the shift of the radiofrequency line at the wavelength 21 cm. This line is clearly visible in the spectra of cold interstellar gas in many other galaxies.

The red shift of all 30 galaxies observed in the radiofrequency range was found to be the same as the red shift of the visible light emitted by these galaxies.

Consequently, the assumption of red-shifting of quanta as a result of their ageing has to be completely rejected.

The only possible explanation of the cosmological red shift is the Doppler effect caused by the expansion of the Universe.

It must be emphasized again that theorists predicted the nonstationarity of the Universe before it was observed by astronomers, and the discovery of red shift confirmed this prediction. We must be amazed not so much by the phenomena of the red shift and the expansion of the Universe (its nonstatic behavior is a direct corollary of the fundamental laws of physics) as by the tenacity and longevity of conservative beliefs.

If observations found no systematic shift in the spectral lines of galaxies, that is, found no evidence of nonstationarity, this would mean that the laws of gravitation need correction and that some yet-unknown universal force prevents the gravitation from making the Universe nonstatic.

In fact, an attempt to introduce such a force had already been made at the birth of modern cosmology by Einstein, before Friedmann's and Hubble's work. This idea is described in the next chapter.



I must add that some astrophysicists point out that some quasars and galaxies may also have, in addition to the cosmological red shift due to the expansion of the Universe, a red shift due to other causes, such as a strong gravitational field or even other yet unknown processes. In principle, such processes are not impossible. I think, however, that the observational data cited to support these hypotheses are ambiguous and inconclusive and can be explained in a conventional manner.