OPENING THE LAST FRONTIER

Discovery of galaxies, the expansion of the universe, and the edge of the world

by Allan Sandage

After 60 centuries of speculation, man discovered the nature of the universe on its largest scale between 1923 and 1934. Proof that galaxies exist, that they may map the structure of the cosmos and trace its history from the present moment backward nearly to the creation event came from observations made with the large telescopes on Mount Wilson in the first third of the century. The work was concerned with the most ancient of inquiries about the universe as a whole. It opened the way for scientists to enter and explore the previously audacious questions of first origins and evolution.

What made this possible was Edwin Hubble's remarkable discovery that space itself is moving in a pattern of uniform expansion, carrying the galaxies with it. The regularity of the motion is what permits the past to be read. Astronomers, using the methods of science, were able to enter the domain held previously by metaphysics and even less substantially by speculation, and this revolutionary accomplishment changed all thought that followed. It had the same audacity and clarity as the revolution of Copernicus—and later of Kepler, Newton, and Einstein—in changing long-held views about the universe itself.

The development began with Hubble's final proof in 1923 that galaxies exist. It continued through his remarkable discovery of the systematic motion of all galaxies away from each other in 1929, and most importantly through his proof of the homogeneity of the galaxy distribution in 1934. These three fundamental ideas, combined with more recent proofs that stars in our Galaxy are of different ages and can be dated, that other galaxies are no older, that there is an oldest age, and that this age is almost the same as that inferred from the expansion motion, have put astronomy into its present period of extraordinary ferment. The root questions of *creation*, *origin*, and *evolution* are the central theme.

The Discovery of Galaxies

The time was early 1917. The place, the 60-inch reflector, then the largest in the world, at the isolated Mount Wilson Observatory in California. George Ritchey, chief optician at the observatory, was the man responsible for bringing the 60- and 100-inch telescope mirrors to the extraordinarily precise curves needed for the telescopes to work. He was also a fine observer and photographer.

After 1909, Ritchey and other members of the observatory staff regularly photographed white nebulous objects with spiral forms whose nature was then totally unknown. These faint nebulous regions of light, cataloged extensively by the Herschels between 1786 and 1864 (though they were known even before that), were curious objects because they avoided the central plane of the Milky Way. However, they appeared in large numbers near the galactic poles. Were they isolated "island universes" as discussed by such visionaries as Emmanuel Swedenborg, Immanuel Kant, and Thomas Wright, or were they connected with the local system of stars in our very immediate neighborhood?

On July 19, 1917, Ritchey took a photograph of the nebula NGC 6946 using the 60-inch reflector with an exposure of 4 hours 25 minutes. Comparing his plate with an earlier one, Ritchey discovered a new starlike image on the later exposure that was absent on the first. The situation and its consequence were not unlike the discovery and its implication for Tycho Brahe of the new star of 1572, which showed that the heavens themselves were subject to change.

Ritchey verified that the image was real, and the result suggested to him that a new star—a nova—had appeared in a white nebula. (Novae and supernovae are stars that have exploded, either in a minor fashion in a nova with a remnant left for further activity, or exploded "completely" in a supernova where the remnant changes fundamentally,

This year marks the 500th anniversary of the birth of Copernicus and the 50th anniversary of Hubble's discovery of galaxies. This article is a condensation of an essay written for this year's Copernicus celebration.

either into a neutron star—pulsar—or perhaps a "black hole." In either case, the "new star" soon disappears from photographs.) Was it actually a nova, and if so, did it say something about the distance? Yes, and the discovery began the chain of events that finally proved the existence of galaxies in 1923.

Aware of the importance of his discovery, Ritchey searched the plate files of Mount Wilson and found that two new starlike images had appeared and had faded in the Great Nebula in Andromeda (NGC 224) in 1909, which showed that such objects occurred regularly in nebulae.

Heber D. Curtis, a perspicacious man trained as a classical scholar and destined to be a principal player in the drama of finding the universe, then searched old plates at the Lick Observatory taken with the Crossley 36-inch reflector. He discovered another nova in the nebula NGC 4227 and added two more in NGC 4321. From the six new objects known in 1917, Curtis was able to discuss the island universe hypothesis for the white nebulae, as contrasted to Harlow Shapley's belief that a *single* stellar system contained all visible astronomical objects. Curtis cautiously stated that the presence of novae favored the extragalactic nature of the white nebulae.

Shapley disagreed. Both men presented various arguments, which reveal about as much of the state of stellar astronomy in 1920 as they do about the disagreement itself. None of the arguments constituted a direct proof, and the disagreement settled into a peaceful stalemate in the absence of decisive new observations.

As late as 1922 there was still no agreement among astronomers as to whether galaxies existed. Shapley, indeed, had shown earlier that the galactic system was probably finite and that its center was in Sagittarius some 30,000 light years distant, but he did not make the necessary generalization outward to the galaxies.

Hubble did. He heeded the signs, made new observations, and discovered the universe.

Cataloging the Galaxies

Galaxies are always referred to by M or NGC numbers. These designations came from two catalogs: one prepared by the comet-seeking French astronomer Charles Messier and published in three installments between 1774 and 1784 by the Paris Academy; the other was compiled by J. L. E. Dryer in 1888, published by the Royal Astronomical Society, and called the *New General Catalogue.*

The brightest galaxies have Messier (M) numbers, but there are only about 100 objects set out in his three lists. Messier compiled this useful tabulation only as an aid for his principal activity of comet hunting, and, since nebulae resemble comets in the eyepiece of a telescope, Messier could keep track of the unwanted stationary nebulae during these searches.

The much more extensive NGC catalog grew out of the major search for diverse astronomical objects conducted from both hemispheres by the Herschels, father and son. The first major survey was that of Sir William Herschel (1738-1822) who systematically observed the northern sky from England with telescopes of his own design and manufacture beginning in the 1780's, long before the invention of photography.

Herschel, a professional musician in his early years and later the first president of the Royal Astronomical Society and the private astronomer to King George III, was a most industrious observer and one of the great figures in the history of astronomy. Herschel presented a copy of the preliminary catalog of 1,000 nebulae and clusters to the Royal Society in 1786; this was followed in 1789 by a second edition containing 1,000 additional entries, and in 1802 by a third list of 500 nebulae, all found by visual methods.

John Herschel continued the work by taking his father's telescopes to Capetown, South Africa, where he finished *The General Catalogue of Nebulae* in 1864. The list contains 5,079 objects, of which 4,630 were discovered by the two Herschels and 449 by others. This catalog is of enormous historical interest and forms the base upon which the *New General Catalogue,* used today, was published by Dreyer in 1888. All bright galaxies are known by their NGC numbers.

Hubble's discovery had the same audacity and clarity as the revolution of Copernicus in changing long-held views about the universe

John C. Duncan, professor of astronomy at Wellesley College in Massachusetts from 1916 to 1950, came west nearly every summer to help with the observing at Mount Wilson. He soon became an expert in astrophotography and was regularly assigned summer time on the Mount Wilson reflectors. In the summer of 1920 Duncan took a series of plates of M33 with the newly completed Hooker 100-inch telescope. By comparing the plates, he found three variable stars, presumably related to the nebula itself.

Edwin Hubble

At the same time, Hubble, following up Ritchey's continuing discoveries of novae, had been keeping M31 under surveillance. An object marked by him as a nova on a 100-inch plate taken on October 5, 1923, was, within weeks, found to be a periodic variable star—the first recognized Cepheid type in a white nebula. Further, one of Duncan's three variables later proved to be a Cepheid. Within a very few months, Hubble had located many similar periodic variables in the large nebulae NGC 6822, M33, and M31. On January 1, 1925, in a paper read in his absence to the American Association for the Advancement of Science in Washington, D.C., he publicly announced that Cepheid variables were present in white nebulae. And, because Hubble had found the characteristic relation between the period of the light variation and the luminosity of these special variables, it was clear that the distances to the parent nebulae were large, that they were clearly outside our own Milky Way, and that they were systems of stars similar in every way to our own Galaxy.

Important as this discovery was, the true Copernican aspect of the work came only later when Hubble showed in 1931 and 1934 that galaxies were the principal constituents of the large-scale structure of the universe itself. The demonstration was made by counting galaxies to successive limits of distance and showing that they are distributed in depth in a generally homogeneous and isotropic manner to the limit of the largest telescopes, providing that sufficiently large volumes are surveyed. The concept has not been seriously challenged since that time. Hubble's conclusions were strongly verified by independent work by Nicholas Mayall at Lick Observatory in 1934, and later by others. There are no indications that galaxies and their clusterings are themselves only subunits of a still higher ordering. Galaxies and their clusters appear to represent the markers in space that define the large-scale features of the cosmos.

Hubble, by this demonstration, breached the last frontier and found the universe itself. The change of thought brought about by these revelations was of the same kind as the Copernican message of 1543. A previously deeply held view of the nature and content of astronomical space was shown to be incorrect, and a large segment of truth was revealed by the new work.

The Expansion of the Universe

Remarkable as these researches were, they were but a prelude to further revelations in an area new to science—an area that had been the province of speculators, philosophers, and mystics for 6,000 years—*creation* and *eschatology*. The story did not begin in a classical astronomical environment, but rather was first foretold as a curious theoretical possibility shortly after the appearance of Einstein's gravitational theory in 1916. The details are too complex to describe completely, but the circumstances can be understood without them.

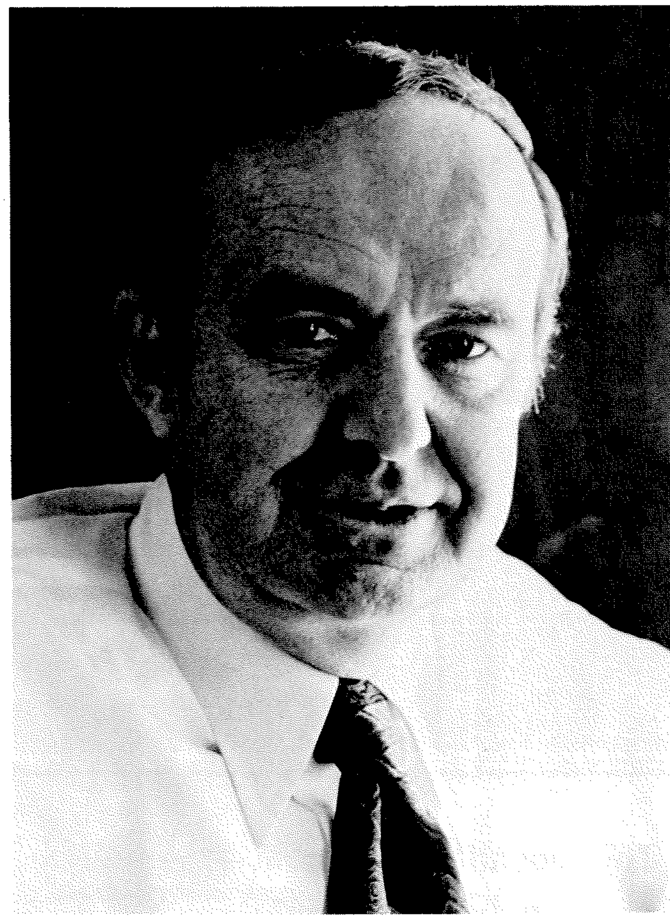
The general theory of relativity is a description of gravitational forces by the effect of mass on the shape of space and time. The presence of mass distorts the flatness of space-time, and this, in turn, controls the motion of particles placed near the mass.

In 1916 static or nonchanging solutions were sought for the Einstein equations that connect space, time, and energy because it was then believed that no large-scale systematic motions existed among astronomical objects. Three static solutions were in fact found. Two of them described ordinary geometry where the space and the time coordinates of an event agreed more or less with intuition. But the third, discovered by the Dutch astronomer Willem de Sitter, was scandalous. The measure of time depended on the position in space, meaning that time intervals were longer the further a clock was from an observer (us). Distant objects such as galaxies might be expected to exhibit an apparent redshift in the color of their light because all of their atoms, which are clocks, would have appeared to slow down. De Sitter predicted that distant galaxies should have large redshifts.

The de Sitter effect does *not* represent a true expansion of space where the distance between galaxies actually changes with time, but before the discovery of the true expansion effect in the Einstein equations in 1922 by Alexander Friedmann, and by George Lemaitre in 1927, many astronomers looked for the de Sitter effect in the astronomical data.

In 1925 no observer knew, or at least understood, the dynamical solutions of Friedmann and Lemaitre in which space itself is in a state of uniform motion compatible with Einstein's equations—a motion that enlarges all distances to galaxies at a velocity proportional to their distance from any given observer.

Hubble's remarkable discovery in 1929 that the universe

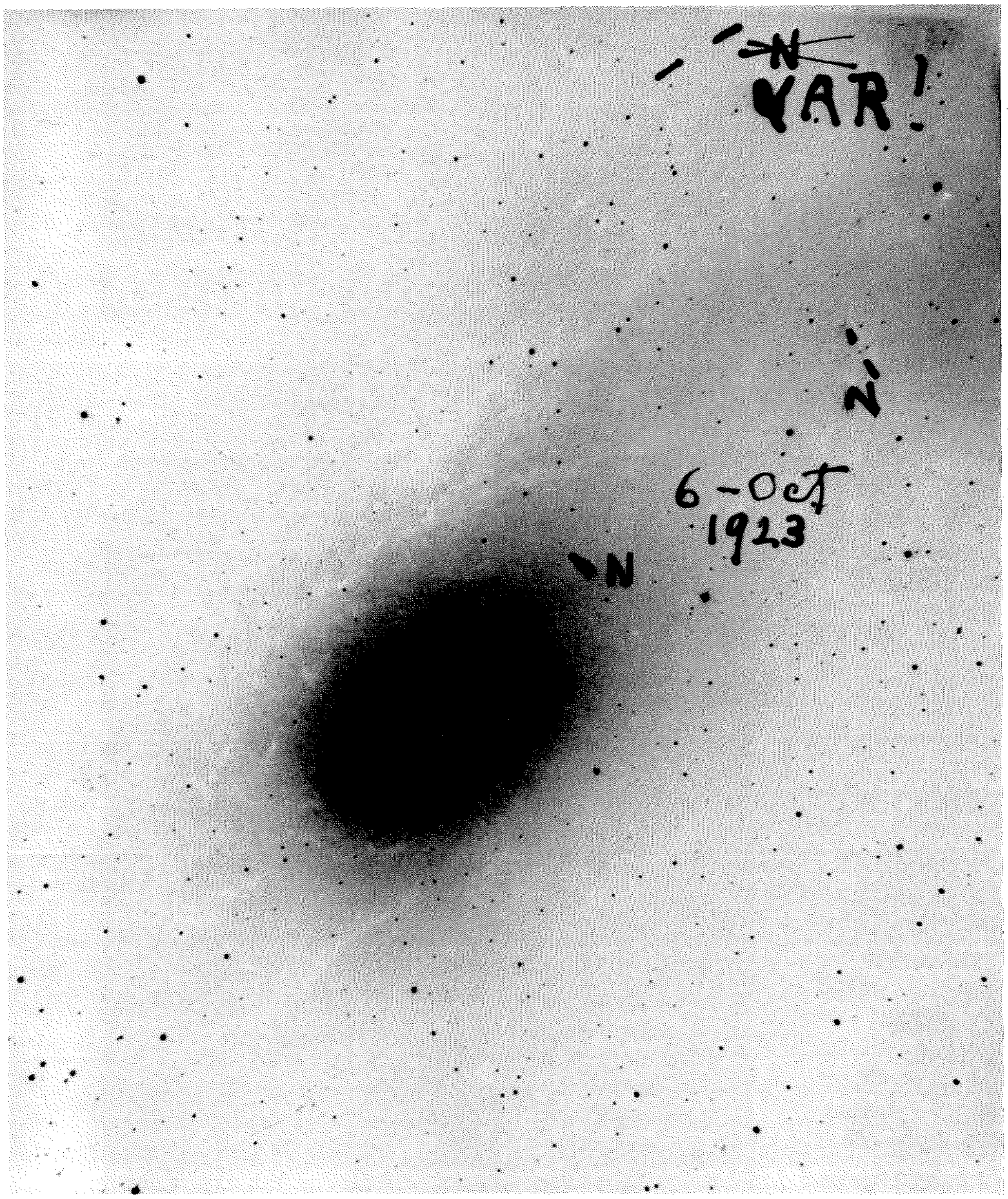


Allan Sandage

The redshift-distance effect is the vehicle by which we can travel backward into history

expands came only after inconclusive discussions in 1925 and 1926. It was based on new redshift observations of Milton Humason at Mount Wilson that could be added to the older material of Vesto Slipher of the Lowell Observatory in Arizona. But most importantly, Hubble's solution used his own improved data on the distances to galaxies. A linear relation between velocity and distance was announced by Hubble in 1929, with a tentative conclusion that the effect might be that predicted by de Sitter.

Events then proceeded very rapidly, both in and out of the Mount Wilson Observatory in the early months of 1930. Sir Arthur Eddington and George McVittie had rediscovered the literature papers of Friedmann and Lemaitre that gave genuine nonstatic solutions of a real expansion; de Sitter had acknowledged that these, rather than his curious static solution, apparently represented the true physical situation; and Hubble and Humason had quickly made new observations of velocities of very



Edwin Hubble's final proof that galaxies exist appears on a historic photograph taken with the 100-inch telescope on Mount Wilson. On October 6, 1923, Hubble marked with an N what he thought was a nova in the Great Nebula in Andromeda. He crossed the mark out a few weeks later when the object proved to be a periodic variable star —the first Cepheid type to be found in a white nebula. Using this star, then several other Cepheids, as measuring rods. Hubble found that the nebula was nearly a million light years away, far outside our own Milky Way, and was a system of stars similar in every way to our own Galaxy.

distant galaxies, with results that showed again the linear velocity-distance effect. From observations made at Palomar as late as 1972, Hubble's law is known to apply at redshifts that correspond to at least 1/3 the speed of light.

The Journey Backward Toward Creation

The incredible significance of the discovery is simply this: Interpreted in terms of the Friedmann-Lemaitre solutions of the Einstein equations, the data require that the universe is an *evolving* thing. It changes with time in a highly regular way. The observed linear velocity-distance relation is the only form of the motion in which the universe can remain similar to itself always, while changing its scale. It is also the only pattern that admits a creation event in the datable past. The fact that the observed form of the law has these properties leads us naturally to suppose that there was a time when the universe was totally different than it is now—that it was, in some sense, created.

The theory in its simplest form requires mutual distances between galaxies to increase with time. The distances were smaller in the past, and were exceedingly small at a particular instant (called the Friedmann singula time) when all space was together. Were the distances, in fact, ever so small as to lead one to believe that all energy, and hence all matter, came out of the Friedmann singularity to begin the expansion?

Some astronomers believe the evidence for change and aging in the universe that leads to this conclusion is overwhelming. Others are not yet convinced. Despite the debate, what *is* clear is that the existence of the redshift-distance effect among galaxies is one of the most profound facts in natural science.

The effect is general for all galaxies. It is the same in all directions. It is the vehicle by which we can travel backward into history. The regular motion, once mapped and calibrated, gives the time when the cosmos began as light, and hence turned into everything we now observe.

The expansion of the universe, through its possible connection with the earliest events in the world, has naturally dominated thinking about universal evolution during the past 25 years. It enormously stimulated the search for other evidence of systematic change and evolution of astronomical bodies.

A new development began along these lines in the 1940's with the realization that stars themselves have a finite life; the oldest of them live and die on time scales that are about the same as the expansion age. The discovery in the mid-1950's that there is an oldest age to stars in our Galaxy, and that this age is closely the same as the Friedmann time itself, was not only crucial to the case for an evolving universe (the big bang), but it opened the way to find the clues as to how our Galaxy formed at its own birthday.

The *oldest* stars were found to move in highly elliptical orbits about the center of our Galaxy rather than in nearly circular orbits like the younger stars such as the sun. The in-and-out plunging motion of these first-formed galactic stars betrays the early motions of matter in the Milky Way at the time of *its* formation, and shows that our stellar system perhaps formed by collapse of interstellar gas toward the galactic center about a Friedmann time (13 billion years) ago.

The important generalization of these results is the revolutionary change in thought and attitude forced by the discoveries. The universe has not always been the way it looks today; and the lights are either going out all over the world as the expansion proceeds or, if the expansion should stop and contraction start, a new cycle could begin.

The Edge of the World

What Hubble did in 1925, in 1929, and again in 1934 was to open the last frontier by his discovery of galaxies, of their generally homogeneous distribution, and the expansion of the universe. The boundary of this frontier clearly can never be reached, either because the universe is infinite, or—if finite—is unbounded, like the surface of a sphere.

However, in a real sense there must be an observational boundary due to the finite speed of light. We look back in time as we look out in space. If the universe has indeed evolved from an earlier time when galaxies did not exist, and if we could look back to that time by looking far enough away, the galaxies beyond a certain distance would not appear on photographic plates. They did not exist before the time of first-galaxy formation. If this edge of the world, given by the time horizon, could be found, it would be a unique proof for a big-bang universe.

Can we see this far? Yes. Have astronomers already done so? Perhaps. Quasars are the nuclei of some galaxies that, for unknown reasons, generate enormous amounts of energy by some semi-explosive process. They are luminous enough to be seen at distances so large that light left them when the universe was only 10 percent of its present age. Said differently, with quasars we see the conditions in the cosmos as they were close to creation.

Now the curiosity of this almost unbelievable situation is even more startling. Quasars have been found with all redshifts up to a certain critical *limiting* value. None are known with larger velocities, although they should have been observed if they existed. The upper redshift limit corresponds to a look-back-time of about 12 billion years—only about a billion years short of the Friedmann creation event.

Does the absence of larger redshifts mean that the time horizon has been breached, and that we look back further than the time of first-galaxy formation? If so, we are now observing not only a matter-horizon in space, but the edge of the universe of galaxies in *time* as well. That we can, in principle, see the edge of the world is amazing. That we may have done so already would be unique. Observations planned for Palomar during the next few years are expected to illuminate this possibility.