

Physical and Observational Evidence for the Expanding Universe

By J. H. REYNOLDS

SO much has been written in recent years on the subject of the expanding universe from the theoretical point of view, that there is no need here to give more than a summary of the position, following Sir Arthur Eddington. The 'Einstein universe' was a conception based on a static solution of Einstein's field equations in which matter was distributed with uniform density, filling a closed space, and in equilibrium owing to the balancing of gravitational attraction and cosmical repulsion. In 1917, Prof. W. de Sitter, of Leyden, put forward his now famous hypothesis, predicting large velocities of recession for distant objects, which was based on a small modification of Einstein's equations. The difficulty about the 'De Sitter universe' was that it was empty, so that the cosmical repulsion acted without hindrance.

The discovery, mainly owing to V. M. Slipher, that the line shifts in the spectra of the spiral nebulae were large and predominantly towards the longer wave-lengths, lent a considerable measure of support to de Sitter's theory, but obviously the universe was not empty of matter, and some other intermediate solution had to be sought. Such a solution was found ultimately by Abbé G. Lemaître in 1927, although its importance was overlooked until de Sitter and Eddington directed attention to it in 1930. They realised at once that a non-static solution of the Einstein equations was possible, so that an expanding or contracting universe was inevitable from Einstein's law of gravitation. The question whether contraction or expansion would best agree with the conditions actually existing was considered settled by the observed displacements of the nebular absorption lines towards the longer wave-lengths, which on the Doppler principle would be interpreted as measures of velocities of recession.

The observational evidence required to establish the hypothesis, and to evaluate the expansion factor, lay in two separate series of data: (1) an adequate number of measured spectra of the extra-galactic nebulae with as wide a range of systems as possible, and (2) a trustworthy scale of distances for these systems.

RADIAL VELOCITIES DERIVED FROM THE LINE DISPLACEMENTS IN THE SPECTRA

From the instrumental point of view, the difficulties of securing satisfactory spectra for measurement were at first considerable. The spectrographs employed for stellar work and for galactic nebulae giving bright line spectra proved quite inadequate to deal with the faint continuous spectrum of the extra-galactic spiral and elliptical nebulae, even with prolonged exposures.

The obvious remedy was to reduce the dispersion, although this made wave-length measures more uncertain, and it was not until a scale of about 100 Å. to the millimetre was adopted that satisfactory densities were obtained, with exposures

from twenty to thirty hours spread over several nights.

The type of spectrum of the extra-galactic nebulae was found in general to be very similar to the solar spectrum, the principle lines being the *H* and *K* lines of calcium, and the *G* group, with some traces of the hydrogen lines, all appearing as wide and shallow absorption lines on a continuous spectrum.

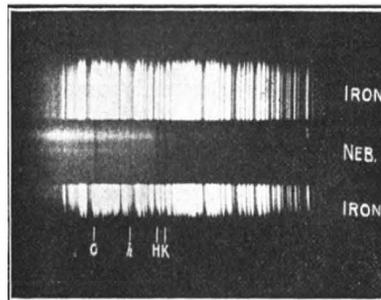


FIG. 1.—Dark-line spectrum of nebula M 81 (N.G.C. 3031) by M. Wolf, Heidelberg.

The first list of forty measured spectra was issued by V. M. Slipher about ten years ago. Among these a small group, including the Andromeda Nebula and M. 33 (the two spirals of largest angular size in the sky and probably the nearest) were found to have large displacements towards the shorter wave-lengths, and therefore considerable velocities of approach. These, with one exception, were all in the south galactic hemisphere in the same region of the sky.

On the face of it, this seemed to raise a serious difficulty, for, according to the hypothesis, it was impossible to have expansion and contraction existing together. The discovery later of the rotation of the galaxy by Dr. Oort, and its subsequent confirmation, furnished a simple explanation of these negative velocities, as the term introduced was a large one, about 280 km./sec., and in the right direction. At the same time, it is as well to state that considerable discrepancies have occurred in the measures of the nearer as well as the farther nebulae. As an example of this, the first measures of the radial velocities of the Andromeda Nebula and M. 33 gave values of -300 km./sec. and -260 km./sec. respectively, while the latest measures of these two spirals give much smaller values, namely, -220 km./sec. and -70 km./sec.

If the galactic rotation is responsible for these velocities of approach in the south galactic hemisphere, one would expect a similar effect on the nearer spirals of large apparent angular diameter in the north galactic hemisphere, so long as they are similarly situated in galactic latitude and longitude. There are two objects which conform to these conditions, M. 81 (N.G.C. 3031) and N.G.C. 2403. The velocity shift of the first has been measured and is of the right order, -30 km./sec. N.G.C. 2403,

which should show a much larger negative velocity from its position, has not yet been included in the lists of radial velocities.

Since 1928 the question of line displacements has been taken up at Mount Wilson by Hubble and Humason, and the number of known displacements very greatly extended by the inclusion of fainter, smaller, and more distant objects. Special spectro-

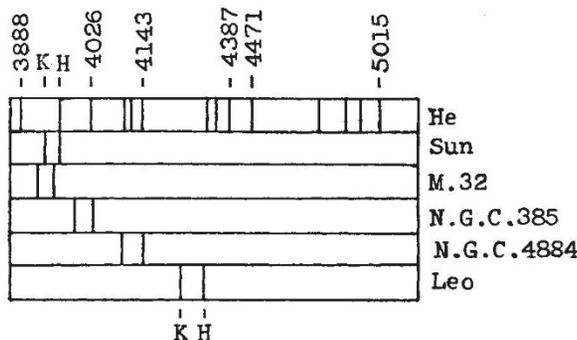


FIG. 2.—Measured displacements of the H and K lines.

graphic apparatus had to be contrived for dealing with these faint nebular images in the 100-in. reflector. For some of the previous work a dispersion of 170 A. per millimetre was found to be concentrated enough, but this was replaced by a dispersion of 340 A. per millimetre.

In its turn, this also was discarded for an even smaller dispersion for very faint nebulae. With a spectrograph objective of F/0.6 designed by Dr. W. B. Rayton, two prisms gave a dispersion of about 418 A. per millimetre at 4500, and one prism 875 A. per millimetre. The uncertainty of the result with the last-named is claimed to be not greater than 300 km./sec., which is small compared with the line displacements. The highest velocity yet measured was found in a cluster of small elliptical nebulae in Leo, where the displacement in the brightest member was equivalent to the stupendous rate of +19,700 km./sec.

The small scale of the spectrum is naturally a disadvantage so far as measurement is concerned, but a well-defined comparison spectrum of helium above and below the nebular spectrum gave a series of wave-length positions quite accurate enough for the purpose.

THE DISTANCES OF THE EXTRA-GALACTIC NEBULAE

The great difficulty in correlating the radial velocities derived from the spectra with distance has been the lack of a definite scale of distances. Until a few years ago, even the order of distance was only a matter of conjecture, but the discovery by Hubble of Cepheid variables in the outer regions of the Andromeda Nebula gave at last a basis for a definite estimate of the distance of this system.

The previous work of Shapley and others had shown that a relationship existed between the absolute magnitude of these stars and their period. Those of short periods of a day or less are of considerably less intrinsic brilliancy than those of much longer period. All the Cepheids found by

Hubble were of the longer periods, corresponding to the stars of greater absolute magnitude, those of shorter period being presumably too faint to be detected. As a first approximation the distance of the Andromeda Nebula was determined as a million light-years, and a similar investigation of M. 33, a neighbouring formation in the sky containing well-defined masses of stars, led to a slightly less distance. The Cepheid variables are undoubtedly the most reliable guide to distances, but there are other types of stars to which absolute magnitudes can be assigned within wider limits. Such are O type stars of very high surface temperatures, B type of temperatures of about 15,000 C., novae which have occurred from time to time in the nuclear regions, and others. These were also found to exist in the two nebulae mentioned. Linked up with the Cepheids they gave a reasonably consistent result. Estimates of these and other spiral and elliptical nebulae were also made by Lundmark of Uppsala and agreed in general with those obtained by Hubble.

Later a list of the eight nearest nebulae was compiled by Hubble, based on magnitude observations



FIG. 3.—The Great Nebula (M 31) in Andromeda.

of involved stars of known types and including the two Magellanic Clouds which were regarded as irregular extra-galactic nebulae.

For more distant nebulae in which faint stars were actually detected but the types of which were unknown, Hubble adopted the principle that there is an upper limit to intrinsic stellar luminosity, and that observed differences in magnitude of these faint stars in various nebulae give a rough guide to differences in distance.

Unfortunately, discrete star discs can be more easily detected in the condensed or later type of spiral than those having an amorphous bright nucleus and a varying amount of outer condensation. Out of a total of 31 nebulae of the type of the Andromeda Nebula, 19 gave negative results, while out of 30 of the condensed type of M. 101, 22 gave positive results. This means that for two objects of different types, comparable so far as diameter and distance are concerned, one would probably show star discs and the other would not.

Some other criterion, therefore, had to be adopted. The most obvious was the angular diameter, which on the average would vary directly in inverse ratio to the distance for any particular class of nebulae.

The difficulty of getting a reliable scale of distances, either by the integrated light of the nebulae or the diameters, is very well exemplified by taking the two spirals of largest angular diameter in the sky—M. 33 and the Andromeda Nebula (M. 31). The distances of both have been well determined



FIG. 4.—Spiral nebula M 33 photographed with 60 in. reflector at Mount Wilson.

by Hubble from the Cepheid variables involved, and come out as follows :

	Distance in 10^4 parsecs.	Diameter.	Estimated Visual Integrated Magnitude.
M 33	2.36	80'	7
M 31	2.47	140'	5

(One parsec = 3.259 light years.)

M. 31 is thus more than half as large again in angular diameter, and its total integrated light six

No. 3282, Vol. 130]

times that of M 33, yet it is actually more distant although the difference is comparatively small.

Of course, if a great number of nebulae of either the spiral or elliptical type are taken, the diameter and integrated luminosity together would give a rough scale of distances, but the distances of particular objects might be out by as much as 40 per cent.

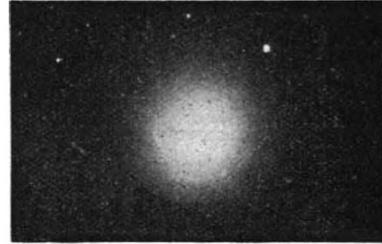


FIG. 5.—N.G.C. 3379, a typical elliptical nebula.

Hubble and Humason prefer to adopt total luminosity alone as a criterion for the nebulae in which no stars are visible, on the ground that while increased exposure systematically extends the linear dimensions on the plate up to a superior limit, the percentage of increased diameter is much greater than the percentage of total luminosity, as there is normally a great concentration of luminosity in the nuclear regions. But in any event, it is obvious that the great trouble in arriving at a satisfactory scale of distances is the lack of homogeneity in the observational material. Even for the same objects the integrated luminosity obtained by photographic out-of-focus images differed by as much as three magnitudes in the results obtained at Mount Wilson and Harvard respectively. In the clusters of elliptical nebulae, which must be of the same order of distance, there is a range of four magnitudes between the brightest and the faintest, so here integrated luminosities would be anything but a sure guide.

CORRELATION OF RADIAL VELOCITIES AND DISTANCES

Notwithstanding the unsatisfactory nature of the scale of distances, Hubble established in 1929 that there is a general correlation between radial velocities and distances of something like 500 km./sec. per million parsecs. The following lists give some typical examples :

ESTIMATED DISTANCES AND EQUIVALENT RADIAL VELOCITIES OF THE EIGHT NEAREST KNOWN SYSTEMS

System.	R.	V.	L.	B.
Nubecula Major .	0.262	+170	245°	-35°
Nubecula Minor .	0.290	+290	265	-40
N.G.C. 6822 .	1.92	-130	350	-19
M. 33 .	2.36	-70	101	-31
M. 31 .	2.47	-185	87	-22
M. 101 .	4.0	+200	116	+58
N.G.C. 2403 .	6.3	..	62	+28
M. 81 .	7.3	-30	107	+39

R. = distance in parsecs $\times 10^5$.

V. = equivalent velocity displacement.

L. = Galactic Longitude.

B. = Galactic Latitude.

The distances given here are based on the Cepheid variables, except the two last, where stars of other types are used as gauges.

The solar galactic rotation effect, found by Hubble to be about 280 km./sec. with its apex in galactic longitude 32° and latitude $+18^\circ$, introduces large corrections in most of the above. Taking this into account, all the negative velocities disappear, and we are left with a quantity which represents a combination of the expansion rate and the peculiar motion of the system.

In the next table the solar galactic rotation effect is smaller compared with the velocities. Representative systems only are given, as the total number now known altogether is over eighty.

Here the radial velocities have already been corrected for the galactic rotation.

The distances have been estimated from involved stars of the 18th magnitude and fainter, the types of which are unknown. These stars are taken as representatives of maximum intrinsic luminosities.

System.	<i>R.</i>	<i>V_s</i>	<i>M_n</i>	<i>M_s</i>
N.C.C. 5236 .	0.9	+500	10.4	18.5
3627 .	0.9	+850	9.1	18.5
1068 .	1.0	+920	9.1	18.7
7331 .	1.1	+500	10.4	19.0
4258 .	1.4	+500	8.7	19.5
4151 .	1.7	+960	12.0	20.0

M_n = integrated nebular luminosities.
M_s = stellar magnitudes.

Past this point stars are no longer available as criteria of the more distant systems, so reliance has been placed by Hubble on the integrated visual or photographic magnitudes of the nebulae themselves. If we take a hypothetical nebula of average total intrinsic luminosity and imagine it removed to twice the distance, its diameter would of course be halved and its total luminosity reduced to a quarter.

As the ratio between one stellar magnitude and the next is 2.512, it follows that a decrease in luminosity of one magnitude would be equal to an increase of distance by 1.58. This is therefore the equivalent ratio in the following table, where integrated magnitudes take the place of estimated distances. Large discrepancies are inevitable in such a procedure, as the results show.

As a final test of the validity of the expansion factor for very distant objects, Hubble and Humason examined certain of the brighter nebulae which appear in isolated groups. The great majority of these are of the small elliptical type, showing no spiral structure. Some nebulae in the Virgo group, which would perhaps be better described as a concentration in the line of sight, gave largely

different velocities, but the general results were such as to give ample support to the expansion hypothesis.

From the data given, Hubble and Humason adopt a round figure of 560 km./sec. per million parsecs as the velocity-distance relation. Although the



FIG. 6.—Field of elliptical nebulae by 100-in. reflector at Mount Wilson. R.A. 3 h. 12 m. Decl. $+41^\circ 6'$ (Perseus).

whole investigation was necessarily based on indeterminate and to some extent arbitrarily selected

System.	<i>V_s</i> (corrected for solar motion).	Mag. Vis. or Phot.
N.G.C. 4051	740	11.9
3227	1090	12.0
7217	1250	12.3
2859	1450	11.1
2950	1560	11.6
3010	1950	11.8
6703	2280	13.6
6702	2530	14.6
6359	3250	14.3
6661	4170	14
6710	5380	15

material, the relation seems to be a reality. The velocity ratio found can scarcely be regarded yet as

Group.	Number.	Range of <i>V_s</i> km./sec.	Mean Photr. Integrated Magnitude.
Virgo . . .	7	600-1170	12.5
Pegasus . . .	5	3500-4000	15.5
Pisces . . .	4	4460-4960	15.4
Cancer . . .	2	4670-4970	16.0
Perseus . . .	4	4800-5800	16.4
Coma . . .	4	5100-8000	17.0
Ursa Major (Baade 24)	1	11800	18.0
Leo . . .	1	19600	19.0

more than a first approximation, but it is of the right order.

A question has been raised as to the proper interpretation of the observed displacements of the

spectral lines. Is the Doppler effect the only possible interpretation? If the slowing down of light over vast distances is a possibility, shifts towards the red would be expected—indeed, de Sitter's original cosmology suggested this as alternative or complementary.

A letter to NATURE of Jan. 16, 1932, from Prof. W. D. MacMillan, suggests that the red shift is due

to loss of energy in the photon in course of time, due either to inherent instability or collisions with other photons. He concludes: "Such an interpretation of the extraordinary shifts that are observed will be more acceptable to many than an interpretation which makes our galaxy a centre from which all others are fleeing with speeds that are proportional to the distances".

Canons of Archæological Theory *

By Dr. DAVID RANDALL-MACIVER

FROM a vast and intricate subject I will select for discussion only two of the principal problems of archæology—namely, the application of a time-scale and the proof of the dissemination of a culture. First, then, as to the time-scale. A series of culture periods has been well established, so that there is a reliable system of what is called 'relative chronology' from the earliest Stone Age down to the time of full documentary history. But it is a very different matter when we attempt to translate these culture periods into centuries and thousands of years. We are wholly dependent for our absolute chronology upon the dates recorded or obtained by immediate inference from ancient writings or traditions. The fragmentary relics of Mesopotamian and Egyptian official chronology furnish a time-scale, liable to much uncertainty in minor details, but trustworthy in all its main lines. As archæological discovery proceeds in the coming years, we may reasonably hope to arrive at a completely graduated scale of chronological dating in actual years for every part of the ancient world after 3500 B.C. But if it is asked what means we have for establishing a chronological as well as a typological scheme behind 3500 or possibly 4000 B.C., I answer unhesitatingly that we have none, and that unless earlier written records or traditions come to light it is probable that we shall never have any.

One very crude method of attempting to avoid this impasse is so illogical that I need spend little time in discussing it. Below the strata in which definitely datable objects are found—whether at Knossos, Ur, Susa, Mohendjodaro, or any other very ancient site—there are generally strata of a certain thickness in which other and obviously earlier forms occur. Now, it is sometimes suggested, even by skilled explorers in their less discreet moments, that the mere thickness of these undated layers may give an indication of the length of time which it took to form them. Yet a very slight amount of reflection, not to speak of actual experience in the field, will show that this reasoning is as childish as it is simple. I have myself seen in Egypt deposits many feet deep which can, nevertheless, be proved, by well-dated objects at the top and bottom, to have been formed within a single century; and I have also seen a concentrated stratum of not more than four feet which contained the products of many centuries closely pressed together.

* From the presidential address to Section H (Anthropology) of the British Association, entitled "The Place of Archæology as a Science, and Some Practical Problems in its Development", delivered at York on Sept. 2.

In a less crude but not very different form, the same error appears in the attempt made by several justly admired writers to establish a chronological scale for the typological series preceding the historical in a country like Egypt. The system of sequence-dating based on typology is now familiar to all students. It was established for Egypt by Sir Flinders Petrie, and for Europe in general by Montelius. As a scheme of relative chronology it sometimes creaks a little, but on the whole it works well and has justified itself, though it may need occasional emendation. But the recurring attempts made by one author after another to translate this relative system into an absolute chronology of years have no logical justification whatsoever, and only encourage self-deception.

The argument is really based on an assumption which can easily be shown to be fallacious. This is the assumption that the rate of progress in civilisation is always uniform. If we know the rate of development in types which took place during the First and Second Dynasties and know also from inscriptions the length of these dynasties, then, it is argued, we have a yard-stick which can be applied to the period preceding the First and Second Dynasties. It is as though a policeman, having timed a speeding motor car over a measured mile, and found that it was going at sixty miles an hour, should appear before the magistrate and state that it was evident the defendant had been proceeding all day at sixty miles an hour.

If, however, we must abandon such illegitimate methods, it is not quite impossible that properly directed ingenuity may find some others which will give a rough scale, less accurate indeed than the chronological, but, nevertheless, valuable. The recent success of Miss Caton-Thompson in settling the very difficult dating of Badarian culture by truly logical methods based on geology is very encouraging; and thirty years ago I myself made a suggestion which I still think has some value. If, I suggested, we could discover the village corresponding to an ancient cemetery, and also ascertain the total number of burials in that cemetery, then we should be able by calculating the presumable death-rate to arrive at a rough estimate of the number of generations. It is evident that several factors in this equation can never be established more than approximately, but it would be worth attempting, if ever a suitable site could be found.

Next we may briefly consider the problem of the dissemination of cultures. This is really the cardina