

The Age of the Universe

The following is a memorandum written by the late Richard Chace Tolman, Professor of Physical Chemistry and Mathematical Physics at the California Institute of Technology. The occasion for this investigation was an informal request from his friend of long standing, Dr. Warren Weaver of the Rockefeller Foundation, for his opinion on the current scientific status of the problem of the age of the universe. To answer this Tolman plunged, with characteristic energy and thoroughness, into a study of the relevant findings, and synthesized them in the memorandum here reproduced. We are greatly indebted to Professor H. P. Robertson who has taken care of the final preparations for publication of the manuscript reproduced here.—Editor.

March 22, 1948

MEMORANDUM

TO: WARREN WEAVER
FROM: R. C. TOLMAN
SUBJECT: AGE OF THE UNIVERSE

FIRST of all I think that we have to begin by putting the phrase "age of the universe" in quotation marks, since I see at present no evidence against the assumption that the material universe has always existed. For me all that such a phrase could mean is the estimated time back to some important large scale event, for which we think we have evidence, e.g., the occurrence of exceedingly high densities and temperatures or the beginning of nebular recession. This carries for me no implication that the universe was created without previous past history at the time of any such event. For example, if at some time in the past there was an occurrence of high densities and temperatures followed by recessional expansion for the matter within the range of our present telescopes, it would seem reasonable to consider this as the probable outcome of a preceding stage of contraction, for that matter.

For our purposes, I think that there are at present only three methods of estimating dates for past events that need concern us. These methods are based on considering (1) the consequences of stellar interaction in our own galaxy, (2) the consequences of radioactive decay in terrestrial or meteoric matter, and (3) the consequences of nebular recession. Let us take a look at the figures provided by the first two of these methods and then consider their connection with nebular recession.

The consequences of stellar interaction provide four of ten items, having a bearing on time scale, which were considered by Bok¹ in his recent report prepared for the Royal Astronomical Society. These items are concerned (a) with the dynamics of loose galactic clusters, (b) with the dynamics of dense star clusters, (c) with the statistics of wide binaries, and (d) with the equipartition of stellar kinetic energies. Bok regards consideration of items (a), (b), and (c) as providing, rightly I think, strong

evidence in favor of his final figure $(3 \text{ to } 5) \times 10^9$ years for "the time-scale of the universe." Under item (d) he regards the old arguments for a much longer time scale, based on the 10^{12} to 10^{13} years needed for establishing equipartition of kinetic energy among the stars, as controverted by the later accumulation of evidence that such equipartition has not yet been attained. Of the further items considered by Bok, one is concerned with radioactive disintegration, one is concerned with nebular recession, and the remainder lead to no very definite conclusions. As an outcome, it seems to me reasonable to take Bok's figure $(3 \text{ to } 5) \times 10^9$ years as a sensible estimate for the past time during which gravitational interaction has been taking place among stars having roughly their present large scale distribution, i.e., as the time back to the last major disturbance which affected our own galaxy as a whole.

The consequences of radioactive disintegration provide three important kinds of past date. In the case of rock samples, ratios of radioactive substance to decay product, e.g., of uranium to radiogenic lead, presumably give information as to the date when the rock was laid down in its present form without subsequent disintegration. In the case of the earth as a whole, Holmes' treatment of Nier's isotopic analysis of samples of lead from minerals of known geologic age presumably gives information as to the date when "the isotopic constitution of the earth's primeval lead began to be modified by additions of lead isotopes generated from uranium I, actinium U, and thorium." In the case of meteorites, the ratios, of uranium (+thorium) to helium, presumably give information as to the date when the meteoric matter solidified.

As might be expected a wide range is found for the dates of deposition of different kinds of rock, with the two earliest dates being 1.765×10^9 years ago for a pre-Cambrian sample of uraninite found in Karelia, Russia, and 1.985×10^9 years ago for another such sample of uraninite taken from the Huron Claim in Manitoba.² Such figures for dates of deposition must, I think, be taken very seriously.

² See Bulletins of the National Research Council on the Age of the Earth, Number 80 (1931), and later. Also see Endeavour article by A. Holmes.

¹ B. J. Bok, M.N.R.A.S. 106, 61 (1946).

The treatment of Holmes³ gives a very significant concentration of values around 3.35×10^9 years ago for the large number of solutions which he has computed. This provides, I think, strong evidence for assuming that the earth has existed for approximately 3.35×10^9 years, sufficiently separate from other bodies so that its contents have been protected from admixture, and sufficiently "cooled down" so that "normal rates" of radioactive decay have since prevailed.

From their analyses of the uranium and helium content of meteorites, Paneth⁴ and his co-workers have obtained a surprisingly wide range of values from 6×10^7 to 7×10^9 for the dates of solidification of meteoric matter, calculated on the assumption that the helium present has resulted from radioactive decay. Bauer,⁵ however, has recently called attention to the extraordinary fact that the *smallest* meteorites are the ones showing the *greatest* helium production, and has shown the possibility of accounting for this on the basis of α -particle production in the surface layers of meteoroids by cosmic-ray irradiation. There is, I think, nothing in the present findings contradicting, either the current belief that most meteors belong to the solar system, or the natural assumption that they were separated from the sun during the same era as the earth and other planets.

Taking the foregoing figures as significant, we must now consider their connection with nebular recession. In the first place, if we take $(3 \text{ to } 5) \times 10^9$ years as the time since the last major disturbance of the galaxy during which undisturbed gravitational interactions have been taking place among the stars of our own system, we shall also want to take this as the time when our galaxy became disengaged from other nebulae as a consequence of nebular recession, i.e., as giving an approximate date for the beginning of recession. In the second place, if we take 3.35×10^9 years ago as the time when the material of the earth became protected against admixture and its radioactive components began decaying at their present normal rates, we shall also be inclined to take this as the time when the general density of matter had fallen to such an extent that the earth could be separated off from other matter and fall to a temperature much lower than that responsible for the preceding synthesis of radioactive elements, and hence, as again giving an approximate date for the beginning of nebular recession. The correspondence between the figures, $(3 \text{ to } 5) \times 10^9$ and 3.35×10^9 years, obtained from two such different methods of dating the beginning of recession, is sufficient to increase our confidence in the reality of the conceptual picture that we are employing.

We must now ask whether such a date for the beginning of recession—in particular, such a large figure for the time back to that event—can be regarded as

constant with the results of giving a sensible theoretical treatment to Hubble's observations on nebular magnitudes, nebular red shifts, and nebular counts to various limiting magnitudes. We shall find it helpful to look at this problem from successively more sophisticated points of view.

Let us first regard the recession of the nebulae as a purely kinematic phenomenon, taking place in ordinary Euclidean space, with the motions of the nebulae unaffected by gravitational action. We may then treat the nebulae as receding from us with velocities, which have remained constant during past time, but which—in accordance with observations on red shift and magnitude—increase approximately linearly with distance as we go to fainter and fainter nebulae. Taking Hubble's figures⁶ for fractional red shift $\delta\lambda/\lambda$ as a function of nebular distance d , and interpreting red shift as an ordinary Doppler effect, we may then write for the ratio v/c of the velocity of nebular recession to the velocity of light

$$v/c = \delta\lambda/\lambda = 5.37 \times 10^{-10} d, \quad (1)$$

where d is the distance to the nebula in light years and c has the value unity in light years per year. Solving for the past time during which the nebula has been moving away from us we then obtain

$$\Delta t = d/v = 1.86 \times 10^9 \text{ years}, \quad (2)$$

as a figure for the time back to the beginning of recession. It is of interest to note that this figure is actually slightly less than the quite certain figures given above for the dates of deposition of the oldest rocks. It is perhaps of more significance, however, to point out that even this very much oversimplified treatment leads to a figure of the same order of magnitude as those derived on quite different grounds.

To obtain a more valid theoretical treatment of nebular recession, it is clearly evident that we must allow for the effects of gravitation on nebular motions. For this purpose I feel sure that we must take the relativity theory of gravitation as the appropriate one to apply at the present time. The results that we obtain from such application will of course depend not only on our theory of gravitation but also on the cosmological model that we choose for the investigation. It will be natural to begin by choosing homogeneous expanding models for study, since such models are mathematically simple enough to permit relatively expeditious treatment, are flexible enough to exhibit according to circumstances a wide variety of types of temporal behavior, and have the property that all observers at rest with respect to matter at their own locality would find the same behavior for surrounding matter as at first sight might seem desirable.

For a second treatment, let us then regard nebular recession as a phenomenon, which accompanies the behavior of a homogeneous model of the universe, when it

³ A. Holmes, *Endeavour* (July, 1947); *Nature* **157**, 680 (1946); *Nature* **159**, 127 (1947).

⁴ F. Paneth *et al.*, *Nature* **149**, 235 (1942).

⁵ C. A. Bauer, *Phys. Rev.* **72**, 354 (1947).

⁶ E. P. Hubble, *Astrophys. J.* **84**, 479 (1936).

expands in accordance with the relativistic theory of gravitation. For the line element for such a model we can write⁷

$$ds^2 = -e^{\theta(t)} \left(\frac{d\tilde{r}^2}{1-\tilde{r}^2/R_0^2} + \tilde{r}^2 d\theta^2 + \tilde{r}^2 \sin^2\theta d\phi^2 \right) + dt^2, \quad (3)$$

and for the local pressure and density of fluid in the model we can write

$$8\pi p = -(1/R_0^2)e^{-\theta} - d^2g/dt^2 - \frac{3}{4}\dot{g}^2 + \Lambda, \quad (4)$$

$$8\pi\rho = (3/R_0^2)e^{-\theta} + \frac{3}{4}\dot{g}^2 - \Lambda, \quad (5)$$

where R_0 is a constant determining spatial curvature, Λ is the so-called cosmological constant, and the terms containing $g(t)$ and its derivatives give the dependence of the pressure and density of the fluid on time.

For the temporal behavior of such models, there is in general a wide range of possibilities,⁸ depending on boundary conditions, on the nature of the fluid filling the model, and on the values of R_0 and Λ . The possibilities include models which at a given rate of expansion could have had very short or very long past time scales since the beginning of their expansion. We must interest ourselves, however, in the past behavior of a model specifically chosen to agree with present observations on the actual universe. In this connection it will be convenient to express $g(t)$ as a power series

$$g(t) = 2(kt + lt^2 + mt^3 + \dots), \quad (6)$$

developed around $t=0$ taken as the present time.

In terms of the foregoing formulation, we can now obtain an appropriate model for consideration by using the results found by Hubble.⁶ From observations on red shift as a function of nebular magnitude, he is able to determine

$$k = 5.37 \times 10^{-10} \text{ yr}^{-1}, \quad (7)$$

$$l = -2.54 \times 10^{-19} \text{ yr}^{-2}, \quad (8)$$

as the values of the first two coefficients in (6) with m and higher coefficients undetermined but small. And from counts of the numbers of nebulae out to different limiting magnitudes, he is able to determine

$$1/R_0^2 = 4.52 \times 10^{-18} \text{ yr}^{-2} \quad (9)$$

as giving the spatial curvature which has to be introduced if the nebulae are really uniformly distributed in an expanding universe. To carry out our contemplated treatment, we shall also need information as to the value of the cosmological constant Λ . For this purpose we may solve (4) for Λ in terms of the values of p , g , \dot{g} , and d^2g/dt^2 at the present time $t=0$. Doing so with the

help of (6), (7), and (8), we obtain

$$\begin{aligned} \Lambda &= (1/R_0^2) + 8\pi p + 4l + 3k^2, \\ &= (1/R_0^2) + 8\pi p - 1.51 \times 10^{-19} \text{ yr}^{-2}, \end{aligned} \quad (10)$$

where p is the pressure at the present time of the fluid by which we represent the contents of the universe. As an approximation, p might be taken as zero, as is often done in treating cosmological behavior, and in any case can hardly have a value as large as $1.51 \times 10^{-19} \text{ yr}^{-2}$, which would correspond to intergalactic radiation at a temperature greater than 100°K . Hence, we may assign to Λ the limits

$$(1/R_0^2) - 1.51 \times 10^{-19} < \Lambda < 1/R_0^2 \text{ yr}^{-2}. \quad (11)$$

We are now ready to investigate the past time scale of our model. Solving (4) for d^2g/dt^2 , we have, in general,

$$-d^2g/dt^2 = \frac{3}{4}\dot{g}^2 + 8\pi p + (1/R_0^2)e^{-\theta} - \Lambda. \quad (12)$$

Since p can at no time be negative on physical grounds, since $\Lambda < 1/R_0^2$ by (11), and since $g(t)$ will be negative in the past, we can then write

$$-d^2g/dt^2/\dot{g}^2 = d(1/\dot{g})/dt \geq \frac{3}{4}, \quad (13)$$

as an inequality which will hold for any time back to the beginning of recession. Integrating, this gives us

$$\Delta t \leq (4/3)(1/\dot{g}_{t=0} - 1/\dot{g}_{t=t_i}) \leq 4/6k, \quad (14)$$

$$\Delta t \leq 1.24 \times 10^9 \text{ years},$$

as a condition on the time interval back to any initial time t_i in the past at which the model was already expanding, with \dot{g} already positive. We thus find that this presumably improved treatment leads to an even shorter time back to the beginning of recession than that given by Eq. (2) as the result of our simplified kinematic treatment.

In inquiring into the cause for this unsatisfactory finding, it might first be suggested that Hubble's results may not be accurate enough to justify the conclusion drawn. It is certain of course that further and more accurate observations will lead to improved results. Nevertheless, it is, I think, highly improbable that the changes could be sufficient to lead to a satisfactory time scale for a homogeneous expanding model. In addition, there are other unsatisfactory features of the model, since in order to make a homogeneous model give a good representation of actual nebular counts out to various limiting magnitudes, it was necessary to introduce a large value for its spatial curvature $1/R_0$. As already emphasized by Hubble, this means a much larger value for the smoothed out density of matter in space (around 10^{-26} g/cm^3) than seems reasonable, and means a closed universe of much smaller dimensions than seem probable. It is my opinion that homogeneous relativistic models can give insight into the possibilities for various kinds of gravitational behavior, but cannot give a very close approximation for the actual gravitational behavior of the nebulae.

⁷ R. C. Tolman, *Relativity, Thermodynamics and Cosmology* (Oxford University Press, New York, 1934), Eq. (149.5).

⁸ Reference 7. See §§157, 158.

Since a homogeneous distribution of matter, expanding in accordance with the relativity theory of gravitation, does not give a good representation of actual nebular observations, it might next be suggested that the trouble lies with the general theory of relativity. This suggestion has found a certain popularity partly, I think, because of the difficulties involved in understanding and applying the general theory of relativity, and partly because it is an open-ended kind of suggestion which allows free rein to unbridled fancy. Thus it sets us free to entertain a variety of exciting speculations: (a) that the nebulae actually stay put in space and the red shifts result not from recession but from some unknown and doubtless extremely important physical principle in accordance with which the frequency of a photon would change with time (Zwicky), (b) that the actually correct laws of gravity could themselves be derived from the homogeneity of the universe (Milne), (c) that there are two mysterious kinds of time, a "kinematical time" and a "dynamical time" which are logarithmically interconnected (Milne), and (d) that the constants of nature are not really constant but have values which change with time (Dirac). Some of these possibilities must be regarded as interesting. Furthermore, it is reasonable to regard general relativity as a development which like others before it will sometimes find its place in some broader theoretical structure. Nevertheless, general relativity provides our present best theory of gravitation—and a very good one at that—and it is my opinion that this is the appropriate theory of gravitation to use in treating the motions of the nebulae.

Accepting Hubble's data, and accepting the relativistic theory of gravitation, we must then conclude that the trouble with time scale has arisen from the representation of nebular motions by an inappropriately chosen model. This seems, moreover, a very natural conclusion to come to since there are many reasons which would make a homogeneous distribution of nebulae seem improbable. These include (a) the high degree of observed inhomogeneity in nebular distribution within short ranges, say far enough out to include the Virgo cluster, (b) the obvious interpretation of Hubble's surveys over greater ranges as due to a non-homogeneous distribution of nebulae receding from a region where there was an exceedingly high concentration of matter, say $(3-5) \times 10^9$ years in the past, (c) my finding that a homogeneous expanding distribution of matter would not be stable against small disturbances away from homogeneity,⁹ and (d) the general improbability that the actual distribution of nebulae should turn out to be homogeneous in view of an infinite range of possible non-homogeneous distributions.

The theoretical study of non-homogeneous distributions of matter, expanding in accordance with the relativistic theory of gravitation, proves to be very

complicated even when we simplify the problem by taking the distribution of matter as spherically symmetrical around some particular origin. Dr. Guy Omer, a former student of mine, took such a study as the subject for his thesis presented last year, and has obtained results of considerable interest.

In carrying out the treatment the following simplifications are introduced. The distribution of matter is taken as spherically symmetrical, which is certainly allowable at the present stage of approximation. The observer is taken as located at the center of symmetry, which greatly simplifies the calculations, and is a natural assumption until we get sufficient observational data to locate the observer—i.e., ourselves—off-center by the right amount. The pressure of the fluid filling the model is taken as everywhere permanently zero, which is a good assumption for our surroundings at the present time, but is not a good assumption for very early stages of expansion. And finally, the behavior of the fluid is taken as describable in co-moving coordinates, which is presumably a good assumption at the present time, but is again not a good assumption for the early stages of expansion when we should expect a sorting out process for particles of different radial velocities. In addition to these simplifying assumptions, it is also assumed that the cosmological constant Λ , the value for which is now unknown, can be assigned any small value found desirable. Making these assumptions, Omer finds a wide range of models which would agree with Hubble's observational data, with varying values for the present local density of matter and for the time scale of expansion.

At present he is carrying out computations for a model with values for its parameters chosen, so as to give a close representation of Hubble's figures for red shifts and nebular counts as functions of nebular magnitude, so as to give—what the astronomers presently like— 10^{-29} g/cm³ for the smoothed out density of the matter now in our neighborhood, and so as to give 3.64×10^9 years for the time scale of the model at the origin.*

Some comments must be made on the significance of this satisfactory figure for the time scale. It is to be noted that the figure refers to a past singularity in the behavior of the model at the origin which occurs before the assumptions used in constructing the model could be regarded as even approximately valid. Hence, it is of interest to note that Omer has just calculated 3.36×10^9 years as the time which we have to go back in the past to get a density as high as 10^{-26} g/cm³ at the origin, i.e., at our location. Calculations for points not at the origin are more difficult but show that the singularity occurs at later times as we go away from the origin. It should also be remarked that longer or shorter time scales can be obtained by taking different values for the cosmological constant Λ . The possibility of choosing different values

⁹ R. C. Tolman, *Proc. Nat. Acad.* **20**, 169 (1934).

* [Omer's work has subsequently been completed, and appears in the January, 1949 *Astrophys. J.*—Ed.]

of Λ to get different types of gravitational behavior has, of course, often been employed in the past, and is a possibility which will remain open until we have some independent way of determining that quantity. It has, however, an arbitrary character that is not pleasing. It is my feeling that the importance of Omer's work lies not in providing a necessarily correct description of the recession of the nebulae, but in showing that the abandonment of the assumption of homogeneity introduces sufficient flexibility so that we do not need to expect trouble as to time scale when we apply the relativistic theory of gravitation to treat the motions of the nebulae.

I hope that you have been able to read this far. If so, I will end by saying that I look for great things from the 200-inch which will have a big effect on theory. I think that our special interest should now lie, not in the approximate linearity of red shift with distance and the approximate uniformity of nebular distribution which have been found, but in the deviations therefrom which we shall find. Perhaps we shall even be able to see out to places in the universe where contraction rather than expansion is taking place. I hope so. Cheerio.

Richard E. Tolman

REVIEWS OF MODERN PHYSICS

VOLUME 21, NUMBER 3

JULY, 1949

Postulate *versus* Observation in the Special Theory of Relativity

H. P. ROBERTSON

California Institute of Technology, Pasadena, California

INTRODUCTION

IN 1905 Einstein¹ published his theory of electrodynamics of moving bodies, which has long since been incorporated in the accepted body of physical science under the less descriptive name of the special theory of relativity. The kinematical background for this theory, an operational interpretation of the Lorentz transformation, was obtained deductively by Einstein from a general postulate concerning the relativity of motion and a more specific postulate concerning the velocity of light. At the time this work was done an inductive approach could not have led unambiguously to the theory proposed, for the principal relevant observations then available, notably the "ether-drift" experiment of Michelson and Morley² (1886), could be accounted for in other, although less appealing, ways. Because of the revolutionary character of the postulates and consequences of this theory, there is discernible in the subsequent decades a certain reluctance wholeheartedly to accept its necessity, a reluctance shared at times even by scientists whose own work paved the way to, or confirmed the predictions of, the theory.

It may therefore be appropriate on this occasion to review the present status of the theory, with special reference to the question of the degree to which postulate can now be replaced by observation in deriving the kinematics on which the theory is based. This re-examination, from a unified point of view closely allied to Einstein's original program, will emphasize the

decisive nature of the two great optical experiments of Kennedy and Thorndike³ (1932) and of Ives and Stilwell⁴ (1938) which have been performed in the interim, experiments which were designed and carried out for the explicit purpose of testing aspects of the Lorentz transformations which are insensitive to the Michelson-Morley experiment. We shall find, in confirmation of conclusions drawn by Kennedy and by Ives, that these three second-order experiments do in fact enable us to replace the greater part of Einstein's postulates with findings drawn inductively from the observations.

KINEMATICAL PRELIMINARIES

We *postulate* that there exists a reference frame Σ —Einstein's "rest-system"—in which light is propagated rectilinearly and isotropically in free space with constant speed c . In elucidation of this postulate, we have here presupposed that any observer P at rest with respect to this frame may be supplied with two independent kinds of instruments, called *rods* and *clocks*, with which he can measure space and time intervals, respectively. By independent we here mean that the fundamental measurement of one kind of interval is not to be reduced to that of the other with the aid of the postulated constancy of the velocity of light, as would, for example, be the case if the "clock" consisted of a beam of light reflected back and forth between two mirrors on the

¹ A. Einstein, Ann. d. Phys. **17**, 891 (1905).

² A. A. Michelson and E. H. Morley, Am. J. Sci. **34**, 333 (1887).

³ R. J. Kennedy and E. M. Thorndike, Phys. Rev. **42**, 400 (1932).

⁴ H. E. Ives and G. R. Stilwell, J. Opt. Soc. Am. **28**, 215 (1938); **31**, 369 (1941).