A DIFFERENT APPROACH TO COSMOLOGY

From a static universe through the big bang towards reality

F. HOYLE

G. BURBIDGE

Center for Astrophysics and Space Sciences, University of California, San Diego

AND

J. V. NARLIKAR

Inter-University Centre for Astronomy and Astrophysics, Pune

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain

First published 2000

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times 11/13pt. 3B2 [KW]

A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication data

Hoyle, Fred, Sir.
A different approach to cosmology: from a static universe through
the big bang towards reality / F. Hoyle, G. Burbidge and
J.V. Narlikar.
p. cm.
Includes bibliographical references.
ISBN 0 521 66223 0
1. Cosmology. I. Burbidge, Geoffrey, R. II. Narlikar, Jayant
Vishnu, 1938– . III. Title.
QB981.H754 2000
523.1–dc21 99-15821 CIP

ISBN 0 521 66223 0 hardback

Contents

2

Early relativistic cosmology

As was just mentioned, like Newton, Einstein also thought that the universe is static on the large scale. Like Newton's attempted model of the universe, Einstein's universe was also imagined to be a homogeneous and isotropic distribution of matter. These criteria have been followed by most model makers in cosmology. We briefly discuss the implications of these assumptions.

The adjectives *homogeneous* and *isotropic* imply the following characteristics. Suppose that, viewed on the large scale, the universe looks the same from all vantage points. There is *no* preferred observing position in the universe; all positions are alike. This is the property of *homogeneity*. Furthermore, as we observe the universe from any such vantage point, should we notice any differences in the structure of the universe as we look in different directions? If we do not notice any directional differences, then we say that the universe is *isotropic*. In other words, if you are taken blindfolded from one spot to another in a homogeneous and isotropic universe, after removing your blindfold you cannot tell where you are or in what direction you are looking.

Even with these simplifying assumptions about the large-scale structure of the universe, the *quantitative* details were still lacking in Einstein's model. To determine these details, Einstein needed his theory of gravitation – the general theory of relativity.

The geometry of spacetime is different from Euclid's geometry in the neighborhood of a massive object like the Sun. It is the main feature of general relativity that any distribution of matter (and energy) should affect the geometry of spacetime around it. For example, the geometry of spacetime is different from Euclid's in the neighborhood of a massive object like the Sun. Einstein therefore expected that the distribution of matter (in the form of stars, galaxies, etc.) should determine the geometry of the large-scale structure of the universe. But here he encountered a major difficulty.

The equations of general relativity, as obtained by Einstein in 1915, permitted models of the universe that were homogeneous and isotropic but *not static*. This difficulty is in fact no different from that which had troubled Newton two centuries earlier. How can matter remain stationary in spite of its self-gravity?

To appreciate this difficulty within the framework of general relativity we need to begin with Einstein's equations

$$R^i_k - \frac{1}{2} g^i_k R = -\frac{8\pi G}{c^4} T^i_k \tag{2.1}$$

as obtained by him in his 1915 formulation[1]. The left-hand side of these equations contains tensors describing the geometry of spacetime whereas the right-hand side has the energy–momentum tensor for the physical contents of the universe[a]. These equations tell us in quantitative terms how the physical contents of the universe determine its geometrical structure.

The assumptions of homogeneity and isotropy tell us that space has constant curvature which Einstein assumed to be positive. We will discuss the motivation for this supposition later. Such a space is finite but unbounded, being the hypersurface of a sphere in four dimensions, expressed, say, by the cartesian coordinate relation

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = S^2. \tag{2.2}$$

To use coordinates intrinsic to the surface define

$$x_4 = S\cos\chi, \quad x_1 = S\sin\chi\cos\theta, \quad x_2 = S\sin\chi\sin\theta\cos\phi, \tag{2.3}$$

$$x_3 = S\sin\chi\sin\theta\sin\phi.$$

The spatial line element *on* the surface S is then given by

$$d\sigma^2 = (dx_1)^2 + (dx_2)^2 + (dx_3)^2 + (dx_4)^2 \tag{2.4}$$

$$= S^2\left[d\chi^2 + \sin^2\chi\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right].$$

The ranges mathematically permitted of θ, ϕ and χ are given by

$$0 \le \chi \le \pi, \quad 0 \le \theta \le \pi, \quad 0 \le \phi < 2\pi. \tag{2.5}$$

But there are two geometrical alternatives open to us. The first is that χ takes the entire range $0 \le \chi \le \pi$, and this gives us what is commonly known

---

a We will assume that the reader is familiar with general relativity. For one who is not, the above discussion can be found in a number of elementary texts on the subject. In particular we follow the treatment given by Narlikar[2].

as *spherical space*. If, however, we identify antipodal points of this sphere, its connectivity is changed and the space is then known as *elliptical space*.

Another way to express $d\sigma^2$ is through coordinates r, θ, ϕ, with $r = \sin\chi\,(0 \leq r \leq 1)$. In elliptical space r runs through this range once; in spherical space it does so twice. In either case the spatial line element (2.4) takes the form

$$d\sigma^2 = S^2\left[\frac{dr^2}{1-r^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right]. \tag{2.6}$$

The constant S is called the 'radius' of the universe. The spacetime line element for the Einstein universe is therefore given by

$$\begin{aligned} ds^2 &= c^2dt^2 - d\sigma^2 \qquad (2.7)\\ &= c^2dt^2 - S^2\left[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2)\right]\\ &= c^2dt^2 - S^2\left[\frac{dr^2}{1-r^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right]. \end{aligned}$$

Note that we have derived the line element entirely from the various assumptions of symmetry. The field equations have not yet been used. We will now see what happens when we substitute the above line element into the left-hand side of Einstein's equations. We get, with S independent of time for a static universe,

$$R^0_0 - \frac{1}{2}R = -\frac{3}{S^2} \tag{2.8}$$

$$R^1_1 - \frac{1}{2}R = R^2_2 - \frac{1}{2}R = R^3_3 - \frac{1}{2}R = -\frac{1}{S^2}. \tag{2.9}$$

To complete the field equations, Einstein used the energy tensor for dust at rest and of uniform density ρ_0 in the above frame of reference, which gives

$$\begin{aligned} T^0_0 &= \rho_0 c^2 \qquad (2.10)\\ T^1_1 &= T^2_2 = T^3_3 = 0. \end{aligned}$$

This leads to two independent equations:

$$-\frac{3}{S^2} = -\frac{8\pi G}{c^2}\rho_0, \quad -\frac{1}{S^2} = 0. \tag{2.11}$$

Clearly no sensible solution is possible from these equations, thus suggesting that no static homogeneous isotropic model of the universe with $\rho_0 > 0$ is possible under the Einstein equations.

It was this inability to generate such a model that led Einstein to modify his equations to

$$R^i_k - \frac{1}{2} g^i_k R + \lambda g^i_k = -\frac{8\pi G}{c^4} T^i_k. \tag{2.12}$$

If we introduce this additional constant λ into the picture, our equations in (2.11) are modified to

$$\lambda - \frac{3}{S^2} = -\frac{8\pi G}{c^2} \rho_0 \tag{2.13}$$

and

$$\lambda - \frac{1}{S^2} = 0. \tag{2.14}$$

We now do have a sensible solution. We get

$$S = \sqrt{\frac{1}{\lambda}} = \frac{c}{2\sqrt{\pi G \rho_0}}. \tag{2.15}$$

Einstein considered this solution as justifying his conjecture that with sufficiently high density it should be possible to 'close' the universe. In (2.15) we have the radius S of the universe as given by the matter density ρ_0, with the result that the larger the value of ρ_0, the smaller is the value of S. However, if λ is a given universal constant like G, both ρ_0 and S are determined in terms of λ (as well as G and c). How big is λ?

In 1917 very little information was available about ρ_0, from which λ could be determined. The value of

$$S \approx 10^{26}\text{–}10^{27}\,\text{cm}$$

quoted in those days is therefore only of historical interest. If we take ρ_0 as $\sim 10^{-31}$ g cm^{-3} as the rough estimate of mass density in the form of galaxies, we get $S \approx 10^{29}$ cm and $\lambda \approx 10^{-58}$ cm^{-2}.

The λ-term introduces a force of repulsion between two bodies that increases in proportion to the distance between them. The above value of λ is too small to make any detectable difference from standard general relativity (that is, with $\lambda = 0$) in any of the Solar System tests. Thus the Einstein universe faced no threat from the local tests of gravity. The model, however, did not survive much longer than a decade, for reasons discussed next.

Apart from finding the solution for the large-scale structure of the universe, Einstein had further expectations from his model. First, he expected the solution to be unique, given the assumptions of homogeneity and isotropy. This would have provided a reason why the universe is the only one of its kind. Further, he believed that with the λ-term there was no possible empty space solution. In this belief he was influenced by Mach's principle which required inertia to be fully determined by matter, implying that it should be impossible to determine the spacetime geometry and test particle trajectories in the absence of bulk matter.

These expectations were not realized. In fact shortly after the publication of Einstein's paper[3], W. de Sitter published another solution of Einstein's field equations[4]. This had the line element given by

$$ds^2 = \cos^2\left(\frac{\rho}{S}\right)c^2 dT^2 - d\rho^2 - S^2 \sin^2\frac{\rho}{S}(d\theta^2 + \sin^2\theta d\phi^2), \tag{2.16}$$

with ρ used here as a radial coordinate, with the density of matter zero and $\lambda = 3/S^2$. Defining

$$R = S\ \sin\frac{\rho}{S}, \tag{2.17}$$

then the above line element becomes

$$ds^2 = \left(1 - \frac{R^2}{S^2}\right)c^2 dT^2 - \frac{dR^2}{1 - \frac{R^2}{S^2}} - R^2\left[d\theta^2 + \sin^2\theta d\phi^2\right]. \tag{2.18}$$

A further transformation of coordinates

$$R = r \exp Ht, \quad T = t - \frac{1}{2H}\ln\left(1 - \frac{H^2R^2}{c^2}\right) \tag{2.19}$$

then takes the line element to

$$ds^2 = c^2dt^2 - e^{2Ht}\left[dr^2 + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right] \tag{2.20}$$

where $H = c/S$.

Both (2.18) and (2.20) are the better-known forms of the de Sitter line element although in his 1917 paper de Sitter used only (2.16).

The de Sitter universe, being empty, offered a counter-example to Einstein's second expectation. Further, by offering an alternative model for the universe for a non-zero λ, de Sitter demonstrated that the Einstein universe is not a unique solution of the cosmological problem.

Much has been written about these theoretical implications of the de Sitter model and its role *vis-à-vis* the Einstein model. It is true that the primary issues of inertia and Mach's principle had been the motivating concepts for the genesis of the Einstein universe. Its contact with astronomical observations had been minimal except for the assumption (which later turned out to be wrong!) that the universe is static[b]. Nevertheless, de Sitter himself seems to have paid considerable attention to the then available observations in order to place limits on the cosmological parameters of his model. It is of interest to briefly give a sample of his arguments.

In units which appear strange in today's usage, de Sitter chose to express physical quantities in terms of a day for the time unit, the astronomical unit (AU) for distance and solar mass $M_\odot$ for mass. Thus $c = 173$, $G = 2.96 \times 10^{-4}$, and unit density of matter corresponds to 6×10^{-7}g cm^{-3}.

Taking the Einstein model first and noting that some of the spiral nebulae are galaxies similar to ours, de Sitter put their linear diameters as $d \approx 10^9$ and angular diameters $\simeq 5'$ (the limits of observations) to set their distance at $\leq 6 \times 10^{11}$ AU. Taking this upper limit as the value of S, the maximum extent in the elliptical space gives $\pi S/2 \leq 10^{12}$. Using the density formula (2.15) he then estimated the mass in the whole universe

$$M = \frac{2\pi^2 S}{\kappa}, \quad \kappa = \frac{8\pi G}{c^2} \tag{2.21}$$

so that $M \sim 8 \times 10^7 S$, with a density of $8 \times 10^6/S^2$ in de Sitter's units. Taking the mass of our own Galaxy as $\sim \frac{1}{3} \times 10^{10} M_\odot$ (based on Kapteyn's estimate) he found $S = 41$ which was absurdly low! Taking instead a density value of $\sim 10^{-17}$ based on the star density at the Galactic center, he found $S = 9 \times 10^{11}$ giving a total mass $M = 7 \times 10^{19} M_\odot$.

Extending arguments beyond our Galaxy, de Sitter argued that if the whole universe is filled with galaxies with their typical separation $\sim 10^{10}$, large compared to their linear dimensions, then the total mass in the form of galaxies worked out to only $\sim 2 \times 10^{16}$. He then argued that 'According to this view, only a small portion of the world-matter would be condensed into ordinary matter.'

This argument rings a bell! Today, when many cosmologists find the astronomically observed matter to be insufficient to account for the amount required by a theoretical model, they jump to the conclusion that the balance

---

b A parallel may be drawn with the genesis of special relativity which was *not* motivated by the Michelson–Morley experiment, but by considerations of accommodating electrodynamics within the framework of the principle of relativity of motion.

must be made up by 'non-ordinary' matter. de Sitter, however, went on to revise the observed estimate of the mean density of matter from 10^{-17} down to $\sim \frac{1}{3} \times 10^{-20}$ and managed to obtain a consistent picture with $S \leq 5 \times 10^{13}$. He concluded:

> We can thus consider the value ($S \leq 5 \times 10^{13}$) as an upper limit – subject, of course, to the uncertainty (which is considerable) of the hypothesis and of the numerical data thus derived.

So far as his own model was concerned, de Sitter referred to the cosmological redshift due to expansion: 'The lines in the spectra of very distant stars or nebulae must therefore be systematically displaced towards the red, giving rise to a spurious positive radial velocity.' Using the then available data on what he called 'helium stars', he used the predictions of his model to estimate $S = \frac{2}{3} \times 10^{10}$. This is perhaps the earliest application of the cosmological redshift hypothesis.

The astronomical data today would vitiate these estimates but their merit lies in the boldness with which the meagre observations were used to set limits on parameters of the theory. They also remind us that when one attempts to match theories with observations that have been obtained with considerable uncertainties (inevitable when the instruments have been stretched to their limits) the results may not always be correct.

The Friedmann–Robertson–Walker models

With the data on nebular redshifts still very uncertain in the early 1920s, this work of de Sitter was not taken seriously by astronomers. Nor was the work of Friedmann of 1922[5] and 1924[6] that extended the investigations of Einstein (matter without motion) and de Sitter (motion without matter) to world models that had both matter and motion. Friedmann discussed open as well as closed models and obtained a dynamical differential equation describing the change of the scale factor S with time.

Einstein made a brief reference to this work, and although by the mid-1920s there were sufficient data on nebular redshifts (cf. next chapter) neither he nor Friedmann himself felt the need to relate these non-stationary models to them. In fact Friedmann's papers were hardly known for nearly a decade after their publication[c].

---

c One of us, (JVN) recalls that at the 1962 Warsaw Conference on General Relativity and Gravitation, some Russian relativists were complaining that the cosmology community did not give due recognition to Friedmann when referring to the standard big-bang models. The belated recognition began to come in the late 1960s.

Lemaitre[7] and Robertson[8] in the late 1920s independently developed cosmological models similar to Friedmann's. In his 1927 paper, Lemaitre derived the two equations

$$\frac{\dot{S}^2 + c^2}{S^2} = \frac{1}{3}(\lambda c^2 + \kappa\rho c^2) \tag{2.22}$$

$$\frac{2\ddot{S}}{S} + \frac{\dot{S}^2 + c^2}{S^2} = \lambda c^2 - \kappa p. \tag{2.23}$$

Whereas Friedmann had considered 'dust' models ($p = 0$), Lemaitre considered dust as well as radiation ($p_r = \rho_r c^2/3$). He also derived the energy conservation equation

$$\frac{d}{dt}\left\{S^3(\rho c^2)\right\} + 3p_r S^2 \dot{S} = 0 \tag{2.24}$$

where ρ is the combined density of matter and radiation and p_r the radiation pressure.

Lemaitre's models included Einstein's and de Sitter's models as special cases. The Friedmann models also were special cases (with zero radiation). Lemaitre also derived the cosmological redshift formula

$$1 + z = \frac{S(t_0)}{S(t_1)} \tag{2.25}$$

for radiation that left the source at epoch t_1 and arrived at the observer at epoch t_0. An approximate Hubble law also followed. Robertson[8] found a similar result in 1928.

It is interesting how prejudices have continued to govern theorists about whether the cosmological models are open or closed. At the 1931 meeting of the British Association, the Bishop of Birmingham[9] declared: 'It is fairly certain that our space is finite, though unbounded. Infinite space is simply a scandal to human thought... the alternatives are incredible.' It was against this background where only closed models were thought to matter, that Einstein and de Sitter[10] wrote a joint paper in 1932 describing a simple open model with the line element

$$ds^2 = c^2 dt^2 - S^2(t)\left[dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right] \tag{2.26}$$

and with $S(t) \propto t^{2/3}$. Known as the Einstein–de Sitter model it is perhaps the simplest of the Friedmann models. This model had $\lambda = 0$ and it belonged to a later period when Einstein had given up the idea of a cosmological constant.

This is the model that currently goes under the name 'flat $\Omega = 1$ model', and we wonder how many cosmologists of today can trace its genesis to Einstein and de Sitter.

By 1932 the observational paper of Hubble published in 1929 (next chapter) had become well known and the idea of the expanding universe had begun to take root. The solutions obtained so far were, however, somewhat *ad hoc*, being based on the simplifying assumptions decided separately by each author. The purist may have worried at the emergence of a unique time coordinate from a theory which was generally covariant. Did such 'time-dependent' solutions have a physical meaning?

A rigorous approach to cosmological models finally emerged from the independent work of Robertson[11] and Walker[12]. Starting from two well-defined assumptions, viz. the Weyl postulate and the cosmological principle of homogeneity and isotropy, they were able to obtain the most general line element as

$$ds^2 = c^2 dt^2 - S^2(t)\left[\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right], \tag{2.27}$$

with $k = 1$ for closed models and $k = 0, -1$ for open models. Here (r, θ, ϕ) are the constant comoving coordinates of a typical galaxy. The fact that such coordinates can be defined rests on the assumption that the world lines of galaxies form a bundle of non-intersecting geodesics diverging from a spacetime point in the past. Thus through each spacetime point a unique member of the bundle passes. The time coordinate is that measured by a galaxy as its proper time. This is Weyl's postulate. The cosmological principle tells us that the hypersurfaces $t =$ constant are homogeneous and isotropic. What Robertson and Walker did was to give a mathematical derivation of the line element (2.27) from these postulates. Thus the Weyl postulate and the cosmological principle single out a global coordinate system. The time coordinate t, commonly called the cosmic time, arises in this way. There is no contradiction between this global symmetry and the local covariance of general relativity.

In Chapter 12 we will return to this discussion and review it in the modern context. Most of the so-called 'standard model' in cosmology today is based on the early work of Friedmann. Whereas the cosmologists of the 1930s and the 1940s were content with modest extrapolations of the present universe, their modern counterparts are more adventurous. Their extrapolations lead them to a state of the universe that was $\sim 10^{-43}$s old with a kinetic temperature $\sim 10^{30}$K! Indeed, to what extent these ideas can be called 'physics' as opposed to 'speculations' will form part of a later discussion (see Chapter 14).

We end this theoretical discussion with a return to Newtonian cosmology. Recall that Newton and later workers failed to describe a static infinite universe in a satisfactory way. By the mid-1930s when the expanding universe had become the popular model, E.A. Milne and W.H. McCrea took up the Newtonian problem within this revised framework. Would it be possible to provide a consistent picture of the expanding universe within Newtonian physics? Milne and McCrea provided an affirmative answer to this question. In fact their Newtonian models exactly resemble the Friedmann models in their dynamical behavior, even with the λ-term, although to understand the redshifts of galaxies one has to apply the Newtonian formula for addition of velocities to the propagation of light from a receding source.

The 1929 paper of Hubble was the watershed for cosmology. Not only did Hubble's law make it clear that the bulk of the observable universe lies beyond our Galaxy, but it also held out the hope of testing the as-yet abstract mathematical models of relativistic cosmology. The following long-term programs emerged from the early work on observational cosmology.

(i) To measure the Hubble constant H_0 more accurately. This required the refinement of techniques of measuring extragalactic distances, as well as finding new ways to extend the cosmic distance ladder farther.

(ii) To measure the predicted slowing down of the rate of expansion. Within the framework of relativistic cosmology, the so-called deceleration parameter

$$q_0 = -H_0^{-2}\left(\frac{\ddot{S}}{S}\right)_{t=t_0},$$

has the range of values $0 \leq q < \frac{1}{2}$ for open models and $q_0 > \frac{1}{2}$ for the closed models ($q_0 = \frac{1}{2}$ holds for the Einstein–de Sitter model). By measuring q_0 through the redshift magnitude relation out to larger redshifts, say, to $z \approx 0.2$, it was hoped to settle the open vs. closed question.

(iii) To measure the curvature of space. Following the work of R.C. Tolman, Hubble hoped to settle the above question by counting galaxies out to increasing distances. The belief was that in a homogeneous universe the number count will reflect the volume–radius relationship and thereby enable us to decide which of the three alternative $k = 0, 1$ or -1 best fits the data.

We will follow these and other cosmological developments of the 1930s in the following chapter. By hindsight we can now say that all the three programs were doomed to inconclusive results. Paradoxically, the improved observing techniques have not helped in settling these questions; instead they have revealed the increasingly complex structure of the extragalactic universe and are only now driving home the point that the expectations of the 1930s were based on cosmological models that were too simplistic.

References

Chapter 2

1. Einstein, A. 1915, *Preuss. Akad. Wiss. Berlin, Sitzber*, 778, 799, 844.
2. Narlikar, J.V. 1993, *Introduction to Cosmology*, 2nd edn (Cambridge University Press).
3. Einstein, A. 1917, *Preuss. Akad. Wiss. Berlin, Sitzber*, 142.
4. de Sitter, W. 1917, *Proc. Akad. Weteusch Amsterdam*, **19**, 1217.
5. Friedmann, A. 1922, *Z. Phys.*, **10**, 377.
6. Friedmann, A. 1924, *Z. Phys.*, **21**, 326.
7. Lemaitre, G. 1927, *Ann. de la Societe Scientifique de Bruxelles*, **47**, 49.
8. Robertson, H.P. 1928, *Phil. Mag.*, **5**, p. 835.
9. Barnes, E.W., Bishop of Birmingham, 1931, *British Assoc. for the Adv. of Sci., Centenary Mtg.* (London, Spottiswoode, Ballantyne), p. 598.
10. Einstein, A. and de Sitter, W. 1932, *Proc. Natl. Acad. Sci.*, **18**, 213.
11. Robertson, H.P. 1935, *Ap. J.*, **82**, 248.
12. Walker, A.G. 1936, *Proc. London Math. Soc.*, **42**, 90.