Discovering the Expanding Universe

The discovery of the expanding universe is one of the most exciting exploits in astronomy. This book explores its history, from the beginnings of modern cosmology with Einstein in 1917, through Lemaître's discovery of the expanding universe in 1927 and his suggestion of a Big Bang origin, to Hubble's contribution of 1929 and the subsequent years when Hubble and Humason provided the essential observations for further developing modern cosmology, and finally to Einstein's conversion to the expanding universe in 1931. As a prelude, the book traces the evolution of some of the notions of modern cosmology from the late Middle Ages up to the final acceptance of the concept of galaxies in 1925.

Written in non-technical language, with a mathematical appendix, the book will appeal to scientists, students and all those interested in the history of astronomy and cosmology.

HARRY NUSSBAUMER is Professor Emeritus at the Institute of Astronomy, ETH Zurich.

LYDIA BIERI is Assistant Professor in the Department of Mathematics at Harvard University.

Discovering the Expanding Universe

HARRY NUSSBAUMER
Institute of Astronomy, ETH Zurich (Switzerland)
LYDIA BIERI
Department of Mathematics, Harvard University (USA)

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521514842

First published 2009

Printed in the United Kingdom at the University Press, Cambridge

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

ISBN 978-0-521-51484-2 hardback

Cambridge University Press
978-0-521-51484-2 - Discovering the Expanding Universe
Harry Nussbaumer and Lydia Bieri
Frontmatter
More information

Contents

Acknowledgements ix
Foreword xi

1 Introduction 1

2 Cosmological concepts at the end of the Middle Ages 5
The spherically closed universe of antiquity and the Middle Ages 6
Cusanus and his universe without centre or boundary 10
A warning by the Church 13
Copernicus and the question of an infinite universe 14
Thomas Digges: How far do stars extend? 15
A crucial step in observational techniques 16
Kepler's finite universe and Galileo's telescope 17
Descartes: An evolving universe 21
Newton's *Principia* and Bentley's sermon 24

3 Nebulae as a new astronomical phenomenon 27
Early reports on nebulae 27
Edmond Halley on several nebulae or lucid spots like clouds 29
The universes of Wright, Kant and Laplace 31

4 On the construction of the heavens 35
Herschel confirms and then rejects island universes 35
The Leviathan of William Parsons, the Earl of Rosse 40
Huggins applies spectroscopy on nebulae 42
On the spectrum of the great nebula in Andromeda 45

5 Island universes turn into astronomical facts: A universe of galaxies 46
Kapteyn's Galaxy 47
Shapley's changing view of the Milky Way 48
Slipher favours island universes 56
Curtis and his novae 57
Was there a villain in the game? 58
The Great Debate 59
Öpik finds the distance to Andromeda 60
Hubble cuts the Gordian knot 60

6 The early cosmology of Einstein and de Sitter 63
Some fundamental relations 65
The static universe of Einstein 72
The static universe of de Sitter 76
De Sitter's Trojan horse 77
No energy conservation in relativistic cosmology! 85
Schwarzschild's vision of curved space 87

7 The dynamical universe of Friedmann 88
An alternative to Einstein and de Sitter 89
Einstein's rejection 92

8 Redshifts: How to reconcile Slipher and de Sitter? 93
Redshifts 93
Distances 95
Early interpretation of redshifts 96

9 Lemaître discovers the expanding universe 99
Lemaître, a student of Eddington and research fellow in Shapley's group 99
Doubts about de Sitter's choice of coordinates 101
The discovery of the expanding universe 103
The derivation of the linear velocity–distance relationship 106
Lemaître determines the 'Hubble constant' from observations 108
Lemaître's interpretation of theory and observations 109
Lemaître's debt to Friedmann 110
Einstein judges Lemaître's interpretation as 'abominable' 111

10 Hubble's contribution of 1929 114
Hubble finds the linear velocity–distance relationship from observations 115

Hubble's interpretation of the velocity–distance relationship 117
What was Hubble's motive for his 1929 investigation? 117
The reception of Hubble's discovery 118
Hubble and the expansion of the Universe 119

11 The breakthrough for the expanding universe 121
The Friday, 10 January 1930 meeting of the Royal Astronomical Society 121
Sir Arthur Stanley Eddington 124
The spreading of the gospel 125

12 Hubble's anger about de Sitter 129
De Sitter's check on Hubble 129
Hubble's angry letter 130
Hubble and Humason enlarge the sample 134

13 Robertson and Tolman join the game 137
Robertson starts from first principles 137
Tolman and the annihilation of matter 139
Weyl's brief return to cosmology 142

14 The Einstein–de Sitter universe 144
Einstein's conversion 144
Einstein and de Sitter agree on the structure of the Universe 149
Eddington's 'after dinner speech' 152

15 Are the Sun and Earth older than the Universe? 153
The age of the Universe deduced from the expansion rate 153
The age of the Earth 154
The age of the Sun and the stars 154
Baade and Sandage drastically reduce the Hubble constant 156

16 In search of alternative tracks 157
Zwicky and the gravitational drag 157
Milne's static Euclidean space 158
The Steady State alternative 161

17 The seed for the Big Bang 164
Expansion out of Einstein's static state 165
Expansion out of a primeval atom: The ancestor of the Big Bang 166
A case for the cosmological constant 168
Is there a beginning to the Universe? 169

Is the Big Bang the origin, or a transition in a cyclic universe? 170
Dark energy: Lemaître equates the cosmological constant Λ with vacuum energy 171

18 Summary and Postscript 173
A brief recollection 173
The birth of modern cosmology 174
How different is today's cosmology 177
A glass of port 186

Mathematical Appendix 188
1 Chapter 6: The early cosmology of Einstein and de Sitter 188
Some fundamental relations 188
The static universe of Einstein 193
The static universe of de Sitter 195
De Sitter's Trojan horse 196
2 Chapter 7: The dynamical universe of Friedmann 197
3 Chapter 9: Lemaître's discovery of the expanding universe 198
Doubts about de Sitter's choice of coordinates 198
The discovery of the expanding universe 199
4 Chapter 13: Robertson and Tolman join the game 201
Robertson starts from first principles 201
Tolman and the annihilation of matter 203
5 Chapter 14: The Einstein–de Sitter universe 204
6 Today's presentation of fundamental cosmological relations 205
Abbreviations 209
References 210
Index 220

Acknowledgements

Several astronomers read early drafts. The comments of Kevin Briggs, Franco Joos and Hans Martin Schmid helped to define the tracks better, which we then followed. We also thank Simon Lilly for pertinent suggestions. We are particularly grateful for extended comments and suggestions of Hilmar Duerbeck, Allan Sandage and Norbert Straumann. We highly appreciate the foreword by Dr Allan Sandage, adding authentic breadth with his very personal impressions of events and personalities, and on Hubble's reluctance to accept the concept of an expanding universe.

We endeavoured to consult the original sources as much as possible, so we often depended on the help and goodwill of librarians. We thank the library of ETH Zürich, the Zentralbibliothek Zürich, the Leiden Observatory Archives, Barbara Wolff of the Albert Einstein Archives of Jerusalem, Liliane Moens of the Archives Lemaître, Jennifer Goldman and Shelley (Charlotte) Erwin of The Huntington Library. We also thank Dr H. C. Carron, as well as the Master and Fellows of Emmanuel College, Cambridge, for permission to reproduce two illustrations from *A prognostication everlasting* by Thomas Digges, and Jim Kaler for providing the spectrum of the planetary nebula NGC 6818.

Foreword

In 1929, Edwin Hubble published a paper that correlated redshifts of galaxies with distances he had estimated from his calibration of their absolute magnitudes previously made in 1926. Writers of both popular accounts and technical textbooks have often described this as the discovery of the expanding universe. It is not so. This meticulously researched book on the history of the discovery traces the complete story of that discovery. The history started a decade before Hubble's, brilliant to be sure, initial correlation, even as it was followed by the further major advance two years later by Hubble and Humason using new observational data (Hubble and Humason 1931). These had, in fact, led to the convincing conclusion that there is indeed a relation between redshift and distance, but not to the reason for a redshift–distance effect, nor to an expanding universe per se. This definitive book, so thoroughly researched for the wider history that started more than a decade earlier than 1929, uses many heretofore-unused original sources, and many others not often cited in other histories.

The misconception is broadly held in the popular press that Hubble discovered the expansion, primarily from the redshift observations of Slipher, which in 1929 Hubble had correlated with distances. Such accounts, if left bare, neglect the central theoretical underpinnings based on Einstein's General Relativity. De Sitter first introduced the notion of cosmological motion in 1917 with his discovery, now considered a mathematical curiosity, of a time-dependent component of a metric satisfying the Einstein equations that contains a factor with distance. The de Sitter prediction was followed in 1922 by Friedmann's more realistic time-dependent metric that is the basis of the modern standard model.

Nevertheless, despite its strangeness, the 'de Sitter effect' spurred many observation astronomers and others to search for the effect among disparate data, for example, by using kinematic data of galactic globular clusters before the discovery of the wider universe of galaxies in the mid 1920s. A proper accounting of the complete history, and in particular of the crucial role of Lemaître in setting out the basis of the theory as we now know it, is the central theme of this important book.

The work stands so solidly on its merits that a foreword by one who is not an author is not needed, and on many grounds would be inappropriate. Why then this foreword? The short answer is that, by the circumstances of history, I became involved with the problem in 1950 as the observing assistant to Hubble at Palomar. I came to understand his purely observational approach to the problem, devoid of much of a theory, and in fact to see his reluctance to believe that the expansion is real. This, of course, is one of the ironies of the history. Because I could describe the Pasadena mood as I learned it, first by working with Hubble and later in re-determining his correction factors to apparent magnitudes to understand the basis of his reluctance, I accepted the opportunity offered by the authors to write this foreword.

Every subject of inquiry can be described and studied on different levels, whether in science, opinion, philosophy, the arts, or all other inventions of the human mind. Each level becomes part of a hierarchy ordered by complexity. It is written, 'It is an essential feature of science that one can analyze a particular level [of the hierarchy] without knowing anything of the lower and the higher levels. It will generally be true that an understanding of the lower and higher levels will enhance an understanding of a particular level, but for some purposes it is not necessary for an understanding of the system at that level.' (Murphy and Ellis 1996.)

The hierarchical approach to the history of cosmology is especially useful here as the subject developed in the middle years of the last century. Initially I knew one level of the hierarchy from the observational side working as Hubble's assistant from 1949 to 1953 in Pasadena. The theoretical levels of the hierarchy had been largely developed elsewhere, and were not needed at that time by the Mount Wilson observers for their level of the work.

As this book describes, the theoretical levels of the modern hierarchy had begun with General Relativity in 1917, twelve years before Hubble's keystone paper of 1929. Predictions amenable to observations followed from de Sitter's discovery in 1917 of his mysterious 'distance dependent time' in the equation of the metric of the differential geometry of spacetime. Astronomers who knew of the predictions, such as Lundmark and Silberstein in 1924 and 1925, had begun to search for the de Sitter effect in the kinematics of many types of astronomical objects before galaxies were understood to define the larger scale structure of the Universe.

The search centred on the problem from classical kinematic astronomy of the 'solar motion' relative to the white nebulae. This search had been started in 1916 and continued to 1925, even before the white nebulae had been proved to be the island universes of the galaxies. Leaders in this were O. H. Truman in the United States, C. Wirtz in Europe (often called the European Hubble without a telescope), and Vesto Slipher in Arizona, who had measured most of the redshifts that were used in the solar motion solutions.

Even as Slipher was after the solar motion, astute theoreticians such as Eddington had made the connection with the de Sitter effect, throwing the solar motion problem into the realm of cosmology. In 1923, Eddington (1923) had already reprinted Slipher's data in his theoretical book on relativity.

Hubble also was after the solar motion, as his 1929 announcement paper makes clear. Yet at the end of the paper he does mention the de Sitter effect, but not the much more important theoretical underpinning papers by Friedmann in 1922, Lemaître in 1927 and Robertson in 1928, which he could have known in 1929. These, like the mysterious de Sitter metric, were also based on General Relativity. Hubble's 1929 announcement was only to point out that a relation exists between redshifts and distance, but not to claim an understanding of the effect in terms of an expansion. This book provides a definitive discussion on these points.

Furthermore, from Hubble's written records throughout the 1930s, it is clear that the later advances from Pasadena did not depend on the theory developed by Lemaître and Robertson, and subsequently advanced by McVittie, Heckman, Milne and McCrea, and others. These culminated in the Mattig revolution where the relevant equations for the observational approach were put in closed form valid for all redshifts, not in series expansions valid only for small redshift values. Only the most elementary theory was needed to search for the second-order term in the observations, now called the deceleration parameter, q_0 (see Humason *et al.* 1956 and Roberston 1955).

Following the Mount Wilson developments in the 1930s, the cosmological program for the Palomar 200-inch had been largely set by 1936. At Palomar, it was to be more of the same that had been done at Mount Wilson. Central was to be the improved formulation of the redshift–distance relation extending the redshift range beyond the capabilities of the Mount Wilson telescopes, and a re-calibration of Hubble's 1930s distance scale using magnitude scales set up photoelectrically rather than by photographic methods. The work could again proceed largely without appeal to theory. To be sure, Tolman and later Robertson were in Pasadena at the California Institute of Technology, but they were studying the problem at distinctly different levels of the hierarchy.

Hubble, often interviewed by the press in the late 1930s and again near the completion of the Palomar reflector, invariably replied to questions on the meaning of the redshift that he was not interested in theoretical hypotheticals, but rather on the data to be obtained at high redshifts to test the reality of a true expansion, which he had often argued against.

I began working as Hubble's assistant while still a graduate student in astronomy at CalTech. At the same time, I was observing with Walter Baade at the Mount Wilson 60- and 100-inch reflectors for a thesis problem on globular clusters related to stellar evolution. My work with Hubble started in the summer of 1949 on a program of galaxy counts, to follow up on his early 1934 count program to find the curvature of space based on Eddington's 1932 proposal to measure how spatial volumes increase with distance (Eddington 1933, Chapter 2).

However, Hubble suffered a major heart attack in late summer of 1949. He could not continue his Palomar observational cosmology project at the telescope. The program was considered central enough for the overall Palomar program that the cosmological segment of the work must be continued despite Hubble's temporary (it would prove to be permanent) incapacity to observe. I was given the observing responsibility at Palomar for the Cepheid program and for the photometric program to obtain the apparent magnitudes of the cluster galaxies in Humason's cluster redshift program for the redshift–distance effect.

I had almost daily contact with Hubble from 1951, during which I learned much of his methods at the observational level of the hierarchy. With his death in September 1953, that part of the Palomar cosmological program concerning the distance scale, the extension of the redshift–distance program into the realm of the second parameter, and the galaxy classification program fell to me. With that event, I became a Pasadena observational cosmologist.

But as the work progressed into the 1960s, it became evident that the redshifts measured by Humason were large enough, such that theory had to be brought into discussion if the expansion were real. I also came to understand why Hubble, mistakenly, had doubts on the reality of the expansion, and why, to appreciate his reluctance, it was necessary to employ the full machinery of the Mattig revolution (Mattig 1958, 1959), with which we began to interpret the new observations and perform the tests in the early 1960s (see Sandage 1961, Sandage and Perelmuter 1991, Lubin and Sandage 2001).

Hubble's scepticism on the reality of the expansion was based on his analysis of his observations of both the redshift–distance relation, and his 1934 galaxy counts. From 1936, he had argued that if he applied his corrections for the effects of the redshift on the measured apparent magnitudes in his two observational programs (the Hubble diagram of redshift vs. apparent magnitude, and the galaxy number count–magnitude correlation) and the expansion was real, he would reach a contradiction in both programs. In contrast, if he applied only a correction that denied the expansion, the contradiction would disappear. The details are more complicated than need be set out here, but are explicit in Hubble's 1936 Oxford Rhodes Lectures (Hubble 1937). Until his death in 1953 he kept the door open as to the expansion not being real. This is seen directly in his 1953 Darwin Lecture (Hubble 1953), given only a few months before his death. In the Hubble diagram shown in that published account, there explicitly is no correction for a real recession.

In an auxiliary program, using both the Mount Wilson and the Palomar telescopes, we discovered in the 1960s that his correction terms to magnitudes for the effects of redshift were faulty, based as they were on too hot an effective spectral energy distribution used to calculate the corrections (Oke and Sandage 1968). The same had been found earlier, based on older photographic photometry (Greenstein 1938), but had been ignored at the time. Use of the correct correction terms, plus Robertson's (1938) connective equation between luminosity, proper distance and redshift, eliminated the discrepancy.

The irony, of course, is that although the discovery of the expansion is often attributed to Hubble with his 1929 paper, he never believed in its reality. It was left to the advances after the mid 1950s to establish its reality on many fronts: the Tolman surface brightness test, the agreement of the timescales of stellar evolution with the expansion age using the revised extragalactic distance scale, the time dilation in supernovae light curves with increasing redshifts, the increase of the cosmic background radiation temperature with look-back time at increasing redshift, and the existence of the background radiation itself, redshifted so that the Planck curve has precisely the correct intensity normalisation for its temperature (i.e. its zero chemical potential), satisfying again the Tolman $(1+z)^4$ surface brightness test.

Fate has permitted a career for me in cosmology by the accident of timing. I was of such an age to fit into the transition period between the pioneer Mount Wilson observers of Hubble, Humason, Minkowski and Baade in the two decades from 1930 to 1950, and the modern theoretical cosmologists, plus the new generation of observers using telescopes in space to study what they believe are origins, in earlier times called cosmogony.

During that 50-year period from 1950 to 2000, I became acquainted with many of the players in both the observational and theoretical realms of the old school. My association and strong friendship with Milton Humason lasted well beyond his retirement from the Mount Wilson and Palomar Observatory in 1957 until his death in 1972. H. P. Robertson was my professor of theoretical physics at Caltech in 1950/51. He vetted my 1961 paper (Sandage 1961) on the program for the 200-inch Palomar telescope. It was then that he told me of his conversations with Hubble in 1928 about his theoretical paper on an expanding universe and an evaluation of the expansion rate using Hubble's distance scale of 1926. Fred Hoyle became a colleague at the Mount Wilson and Palomar Observatory in the 1950s and we wrote a paper together on the observational value of the deceleration parameter (Hoyle and Sandage 1956). We also often argued on the validity or otherwise on the Steady State model, which I had written against from the observational side. We remained friends throughout. I knew McVittie well and often discussed with him the observations being made at Palomar on the redshift–distance relation and the determination of the second-order term. His book, *General Relativity and Cosmology*, went through two editions before and after the Mattig revolution. In the first, in 1956, the equations relating redshift to distance and the spatial volumes enclosed therein were all in series expansions of the redshift. The second edition in 1965 set out all the equations in closed form valid for all values of the redshift, no matter how large, based on the 1958/59 Mattig equations.

At the time of the 1967 Prague IAU meeting I had not yet met Mattig. But there I thought I had seen, at a distance, a name badge marked 'Mattig'. My awe of his two revolutionary papers of 1958/59 was too great for me to talk directly with him, so I did not follow up my glance at the badge to make an acquaintance. I learned only much later that Mattig was not even at the Prague meeting, so it has turned out that I was in awe of a mistakenly read badge rather than the real thing. However, in the 1990s we did spend a day together in Basel with G. A. Tammann, my long-time colleague, on the distance scale problem beginning in 1963. We talked about classical cosmology which, at that time, was still centred about 'the search for two numbers' (Sandage 1970), the second of which was based on Mattig's equations. In connection with his Ph.D. examination (his Ph.D. was on sunspots), Mattig had been given the task of presenting a cosmological problem still open at the time. He succeeded in giving a closed analytical solution for the relation between redshift and bolometric luminosity. Out of that exercise came the two key papers of 1958 and 1959 that changed the course of theoretical discussions as far as they relate to observations.

I became acquainted with Heckmann in 1957 at the Vatican conference on stellar populations. We walked throughout Rome during breaks in the conference and had discussions about the history of many of the sites we visited that he knew so well. I met him again in 1961 at the Santa Barbara IAU symposium 15 on cosmology, the conference proceedings of which were edited by McVittie. Some months later I received a gift from him of his well-worn personal copy of his important book, *Theorien der Kosmologie*, published in 1942 in Berlin during the Second World War. The book was then, and is still today, most difficult to find.

I had also met Lemaître at the Vatican conference on stellar populations in 1957, where stellar evolution was only then drawing near to cosmology via the evolutionary corrections to galaxy luminosities at high look-back times. I met him again at the 1961 Santa Barbara cosmology conference. During a noon lunch break, Lemaître reintroduced himself and recalled our conversations at the Vatican conference. After some discussions about progress in the Palomar program we began a wider discussion on the beauty of the Einstein equations and the mystery of cosmology itself. Toward the end, he asked, 'Sandage, can you really envisage curved space and the beauties of Riemannian geometry, so necessary for relativity?' I replied, 'No, Father, I have tried and tried, using all the tricks known to visualise curved space, but my visualisations have so far failed.' Lemaître then sighed and said, 'I understand, but it is a pity because the visualisation is so beautiful. Perhaps it might be best for you to change fields.' He said it gently, like a father to a son.

Beginning in 1917, the road to the discovery of the expanding universe was traveled by many scientists on each side of the hierarchy between theory and observation. This remarkable book gives credit in a fair and neutral way to many who made the journey. It deserves to be studied by future historians because it is authoritative and definitive.

Allan Sandage
Observatories of the Carnegie Institution
California