Cosmology's Century

AN INSIDE HISTORY OF OUR MODERN UNDERSTANDING OF THE UNIVERSE

P. J. E. Peebles

WINNER OF THE NOBEL PRIZE IN PHYSICS

More praise for *Cosmology's Century*

"A century of big ideas and powerful instruments has led us to the current model of our universe, with its inflationary beginning, cosmic structure built by the gravity of dark-matter particles, and accelerated expansion caused by dark energy. *Cosmology's Century* is a firsthand account of that remarkable period by Jim Peebles, who led this grand adventure with his manifold contributions and broad influence. A must-read for any serious student of cosmology."

—Michael S. Turner, Kavli Foundation and University of Chicago

"Jim Peebles has surely contributed more to the history of our understanding of the large-scale structure and evolution of the universe than anyone else still in a position to write about it; so written about it he has, and magnificently!"

—Virginia Trimble, Former President, International Astronomical Union, Division of Galaxies and the Universe

"An inspiring history of cosmic ideas."

—Joseph Silk, author of *The Infinite Cosmos: Questions from the Frontiers of Cosmology*

"Peebles offers a broad and deep description of cosmology, presenting the history of the field as well as many of the side turns, dead ends, and wrong paths that researchers explored along the way. I really enjoyed reading this book."

—David W. Hogg, New York University

COSMOLOGY'S CENTURY

Cosmology's Century

AN INSIDE HISTORY OF OUR MODERN UNDERSTANDING OF THE UNIVERSE

P. J. E. Peebles

PRINCETON UNIVERSITY PRESS
PRINCETON & OXFORD

Published by Princeton University Press
41 William Street, Princeton, New Jersey 08540
6 Oxford Street, Woodstock, Oxfordshire OX20 1TR

press.princeton.edu

ISBN 978-0-691-19602-2
ISBN (e-book) 978-0-691-20166-5

British Library Cataloging-in-Publication Data is available

Editorial: Jessica Yao and Arthur Werneck
Production Editorial: Brigitte Pelner
Jacket/Cover Design: Chris Ferrante
Production: Jacqueline Poirier
Publicity: Matthew Taylor (US), Katie Lewis (UK)
Copyeditor: Cyd Westmoreland

Jacket Image: Planck's Cosmic Microwave Background (CMB) Map, 2013, ESA and the Planck Collaboration

This book has been composed in Miller

Printed on acid-free paper ∞

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To Alison, my best friend for six decades

CONTENTS

IT IS REMARKABLE that we can say with some confidence what the universe was like far away and in the remote past. The well-tested theory grew out of starting ideas from a century ago, in a reasonably simple way compared to other branches of natural science, and relatively few people were sifting through the clues to how to make progress. I have been one of that party for over half of the century since Einstein started us in about the right direction, and this book is my opportunity to put down what I understand to be what was happening and my impressions of why. It is generally acknowledged that in natural science, we take poor care of our history. I aim to present the story warts and all: the brilliant insights and lucky guesses, the roads not taken and the mistakes large and small, and the accumulation bits of evidence that at last began to fit together in a way that makes sense. The relative simplicity of the story makes it a good illustration of how natural science really is done.

I intend this book to combine an objective history of this subject with my own recollections, the latter indicated by use of the first person. I think the two are not seriously incompatible. We are not approaching a final theory (if there is such a thing), which means that assessments of where we are within our incomplete and approximate state of establishment of natural science call for judgments that cannot be objective. We must operate with subjective assessments of what we hope is reasonably objective evidence, while bearing in mind that some pieces of evidence are a lot more objective and informative than others.

I date the convergence of evidence to a credible theory of the large-scale nature and evolution of the universe to the years 1998 to 2003, and I end my account at this revolutionary half decade around the turn of the century. As I complete writing this book, in 2019, I enter occasional comments about what has happened since the revolution. It would be tedious to keep repeating "at the time of writing," so I leave this to be understood when I feel the context suggests it.

I begin this history with Einstein's introduction of the transformative general theory of relativity, which allows quantitative analyses of the nature of a universe without edges. What came before Einstein is important—it informed later thinking—but my comments are limited to the conceptual problems with an unbounded universe in Newtonian physics. I have taken the liberty of simplifying the story of what happened after Einstein by omitting paths that I do not think were useful (even as foils to more successful thinking) and are not likely to be missed. I apologize for and would appreciate being informed of inadvertent omissions of lines of research that arguably have socially redeeming value.

The post-revolution history of this subject is important, and it is interesting also to consider what things might be like in the distant future and what might have happened in the remote past, before the earliest stages of evolution about which we have useful evidence. But I do not discuss these considerations.

People now contributing to advances in research in cosmology should be aware of the history of their subject. I offer this book as a place to look up what happened years ago that helped set the community straight, more or less; it is complicated, of course. Cosmologists already know the technical aspects of this history, which are not all that difficult. I intend the explanations in this book to be understandable to an undergraduate who is thinking of majoring in physical science, and to a nonscientist who is fascinated by what has been learned about stars, galaxies, and the expanding universe and is willing to skip over the technicalities and pay attention to the descriptions. Details are useful to those inclined to examine them, so I have placed in footnotes those comments that I consider relevant but not essential to the broader picture described in the text. Footnotes also offer definitions of astronomers' sometimes curious conventions; they are best tracked down through the index. There are equations in the main text, because they are important to the story. Where the equations are dense, I intend the narrative to keep the big picture visible. There's nothing wrong with skimming over equations, to be sorted out later if they're found to be really needed for the reader's purpose. And I offer introductory and concluding sections in which I discuss the situation without equations.

I term this book a history, but it is written in the tradition of the physics I know and love, save only for my attempt to avoid our superficial creation stories. Discussions with professional historians have taught me that my approach certainly could be complemented by assessments in the traditions of historians and sociologists. I have written little about personalities, for example, or the evolving nature of support for research, and I do not mention much about means of communication, in earlier times by letters exchanged within the old boy network, now perhaps through blogs, of all things. Communications at conferences remain important, but conferences in this subject have become increasingly specialized as the reach of cosmology has grown, a troubling development in the eyes of many in my generation. But I must leave all this to real historians, philosophers, and sociologists of science, who I hope will understand that I transgress on their traditions because I operate in the traditions of a practicing physicist.

I offer thoughts about the nature and philosophy of the enterprise of natural science, informed by what happened in cosmology, in Section 1.1 and Chapter 10. We see in this history that scientists act as they do because they behave much like people in general, though they tend to be more compulsive about it. And with each advance in science, we see an addition to the evidence that there is an objective physical reality and that we are probing ever more deeply into its nature. There is nothing new in all this, but I think the examples to be drawn from the history of modern cosmology are particularly clear and informative, because the subject is relatively simple.

I have already presented portions of this story. What happened in George Gamow's research group in 1948 is considered in detail in Peebles (2014). The work in Bob Dicke's Gravity Research Group that was so important to the development of experimental gravity physics—and led to the recognition of the sea of thermal radiation remnant from the hot early universe—is reviewed in Peebles (2017). Recollections of research in the 1960s by those who were involved in the identification and interpretation of this fossil thermal radiation are in the book *Finding the Big Bang* by Peebles, Page, and Partridge (2009).

References to the research papers I consider important to the story are indicated by authors' names followed by the year and are listed in the References section at the end of the book. The list is dismayingly long, but it has to be: although this has been a relatively small science, its development took a lot of work. I have selected samples of the pioneering contributions and apologize to colleagues who have different opinions about this subjective matter. Page numbers following references indicate where the papers are cited in the text.

Some of the quotations in this book are taken from the literature, and the sources are so indicated. Where the quote is in French or German, I add my translation, sometimes condensed and aided by Google. This book offered an excellent opportunity to ask for recollections from those who have long memories of research in this subject. Quotes drawn from them for the purpose of this book are marked by the author's name and "personal communication." I have profited also from the advice of younger people, and from the wonders of the Internet. I am particularly thankful for NASA's Astrophysics Data System Bibliographic Services archive, a most useful tool for tracking down research papers from times past.

Figures that illustrate data can be influential, and the evolving nature of these figures is a part of the history. I am grateful to colleagues who gave me figures they made and own; their names are mentioned in the captions. The figures I made for this book, or made in times past but never published, have no references in the captions. Captions state sources of the many figures that have been taken from the literature, and the copyright holder can be traced through the reference to the publication. Copyright holders have a broad variety of prescriptions for statements of permission to reproduce, and their conditions for permission range from casual statements that reuse of figures is OK to payments required to reproduce two of the figures in this book taken from the publication of an otherwise respectable scholarly society. I take this confusion of permissions to be a consequence of the natural desire of publishers to keep some control over their content while the ease of taking figures from the literature for use in lectures can readily spill over into publications. I apologize for any permissions to reproduce I may have improperly stated or overlooked, and if notified will make amends in later printings.

The color plates that appear within chapter 9 are a sample of the actors in this history; I mean them to be reminders of the people behind those equations and measurements. I apologize to valued colleagues whose photos could have been included if space in this book and the energy to collect them had been more freely available. The text accompanying the photographs is my opportunity to comment on the stories behind the images, the analog in print of teachable moments.

My choice of units follows customs that tend to differ in different lines of research. In some parts of cosmology, the units usually are chosen so the velocity of light is unity. These equations look odd to me when the symbol c is entered, a matter of conditioning of course, but I follow tradition, which seems appropriate, since this is a history. In other places, Planck's constant $\hbar$ is unity, or Newton's constant G is unity. The old centimeter, gram, second units are being replaced by meters, kilograms, seconds. I suppose this is a sensible move, but the change is slow, and again I follow the history in staying with the former.

The index lists only a few of the pioneers of cosmology. This is a subjective choice, as is whatever else is deemed appropriate for an index. It would make no sense to place in the index the many appearances in the text of the word "redshift," so I index only the definition that appears early in the text. The word "inflation" appears a lot, too, and I enter the first significant commentary about the concept and later page numbers in which cosmological inflation is particularly relevant. But such algorithms are only of limited help with so many decisions.

This account may seem overly centered on the small town of Princeton in the small state of New Jersey. That is inevitable, in part because I have been a member of Princeton University since arriving here as a graduate student in 1958, but inevitable in even larger part because a good deal of the story happened here. My role in this story was aided by sabbatical leaves at the California Institute of Technology; the University of California, Berkeley; the Dominion Astrophysical Observatory in British Columbia; the University of Cambridge; and on two occasions, the Institute for Advanced Study in Princeton. I learned a lot at these places.

I have benefited from advice from and recollections of many colleagues: Neta Bahcall, John Barrow, Dick Bond, Steve Boughn, Michele Cappellari, Claude Carignan, Ray Carlberg, Rick Carlson, Robin Ciardullo, Don Clayton, Shaun Cole, Ramanath Cowsik, Marc Davis, Richard Dawid, Jaco de Swart, Jo Dunkley, John Ellis, Wyn Evans, Sandra Faber, Kent Ford, Ken Freeman, Carlos Frenk, Masataka Fukugita, Jim Gunn, David Hogg, Piet Hut, David

Kaiser, Steve Kent, Bob Kirshner, Al Kogut, Rocky Kolb, Andrey Kravtsov, Rich Kron, Malcolm Longair, Gary Mamon, John Mather, Adrian Melott, Liliane Moens, Richard Mushotzky, Kieth Olive, Jerry Ostriker, Lyman Page, Bruce Partridge, Will Percival, Saul Perlmutter, Mark Phillips, Joel Primack, Martin Rees, Adam Riess, Brian Schmidt, Jerry Sellwood, Joe Silk, David Spergel, Ed Spiegel, Paul Steinhardt, Matthais Steinmetz, Michael Strauss, Alex Szalay, Alar Toomre, Rien van de Weygaert, Hugo van Woerden, Steve Weinberg, Rainer Weiss, Cyd Westmoreland, Simon White, Ned Wright, Jessica Yao, and Matias Zaldarriaga. I surely have forgotten to mention some; my sincere apologies.

COSMOLOGY'S CENTURY

CHAPTER ONE

Introduction

THE STORY OF how cosmology grew is fairly simple, compared to what people have been doing in other branches of science, but still complicated enough that sorting it out requires a better plan than the common practice in science. Papers reporting research in cosmology and other parts of physics usually begin with an outline of what came before. Abandoned ideas and roads not taken are seldom mentioned, and there is the natural human tendency to follow patterns of attributions found in introductions in other recent papers. This builds evolving creation stories that efficiently set the current context for the research to be described. We tell these creation stories in the classroom for a quick introduction to what we are really interested in: the nature of the science. But the stories tend to be at best only vaguely related to what actually happened. Their gross incompleteness may not be a problem for ongoing research, except of course when good ideas have been overlooked or abandoned and lost. But the creation stories leave a woefully incomplete and inaccurate impression of how science is done.

To do better, we have to look further back in time, and we certainly have to consider the ideas that seemed interesting but were falsified or otherwise found not to be so interesting after all. A closer account of how cosmology grew presented in chronological order would be awkward, because different parts of what became the established theory were making progress at different rates following different methods and motivations until they started to come together. This account accordingly presents histories of six lines of research that were developing more or less separately. They are reviewed in Chapters 2 to 7. The advantage is a modest degree of continuity within each chapter. The disadvantage is the need to refer back and forth in time to what was happening in different lines of research. The arrangement is explained in more detail in Section 1.2 in this chapter, in the form of an outline and guide to the story to come. But first let us consider our traditions of research in the natural sciences, with particular attention to the operating conditions in cosmology.

1.1 The Science and Philosophy of Cosmology

The starting assumption for cosmology, as in all branches of natural science, is that nature operates by kinds of logic and rules that we can discover by careful examination of what is observed, informed by past experience of what has worked. The results are impressive; I urge any who might disagree to consider the rich fundamental physics employed in the construction and operation of their cellphones. But despite the many demonstrations of its power, physics, along with all the rest of natural science, is incomplete. Maybe discoveries to come will make the physical basis for science complete, revealing the final rules by which nature operates. Or maybe it's successive approximations all the way down.

The standard and accepted methods of science must be adapted to what can be done, of course. In physical cosmology and extragalactic astronomy, we can look but never touch. In cosmology, we cannot run the experiment again; we must instead resort to what can be inferred from fossils of times past. We find some fossils relatively nearby, as in the rocks on Earth and the stars in our galaxy and others, all of which have their own creation stories. Our past light cone offers us views of times past, because radiation detected here has been approaching us at the speed of light: the greater the distance of an object, the earlier in the evolution of the universe it is observed. Our light cone integrated through human history captures an exceedingly thin slice of what has been happening, but it reveals the way things were over a long range of time in a large universe that offers a lot to see and to seek to interpret.

The research path to where we are now in cosmology is marked by debates on open questions, as is usual in natural science. But the issues in cosmology have been defended and criticized with considerably more vigor than might have been expected from the modest weight of the evidence at the time. This was in part because observations that might settle questions in cosmology have tended to seem just out of reach or perhaps just barely possible. And I think an important factor has been the tendency to take a personal interest in the nature of our world. Is the universe really evolving, or might it be in a steady state? If evolving, how might it all end, in a big crunch or a big freeze? And where did it all come from? Such debates are quieter now, because we at last have a theory that passes an abundance of tests, but they continue.

Research in cosmology in the twentieth century usually was done in small groups, often an individual working alone or maybe with a colleague or a student or two. In the twenty-first century, ongoing research in cosmology grew richer and called for larger groups to develop special-purpose equipment for data acquisition, which in turn called for groups of comparable size to reduce the data and interpret it. Big Science has become important to this subject: We have to get used to gathering data in vast amounts, analyzing these data, and employing massive numerical simulations that help bridge the gap between theory and observation. But Big Science best takes aim at well-motivated and sharply defined questions. The main considerations in this book are about how small groups working on seemingly independent lines of research found their results coming together in a cosmology that looked good enough to call for the demanding tests afforded by Big Science. I date this revolutionary convergence to a credible theory to the half decade from 1998 to 2003.

Research certainly continued to be active and productive after the revolution; the difference is that the community had agreed on a paradigm, in Kuhn's (1962) terms. (This is what the majority was thinking, of course; not all agreed.) An example of the adherence to the normal science of cosmology is the study of how the galaxies formed and evolved, which builds theories of galaxy formation on the standard and accepted theory of the evolution of the universe. Normal scientific research of this sort may uncover anomalies that point to a still better underlying theory. This is a point of particular interest in cosmology, because the theory is at the same time well and persuasively tested and particularly incomplete.

Our present normal science of cosmology includes an excellent case for the presence of dark matter that interacts weakly if at all with ordinary matter. There are tight constraints on the properties of dark matter, but no clear evidence exists of detection of this substance other than the inference from the effects of its gravitational attraction. Some argue that dark matter will remain only hypothetical until there is more evidence of it than that: maybe detection in the laboratory, maybe indications of what it is doing to galaxies apart from holding them together. Others argue that the case for dark matter already is so tight that it is abundantly clear that the dark matter really exists. The same applies to Einstein's cosmological constant, Λ. It has gained a new name: dark energy. But that is a poor disguise for a fudge factor that we accept because it serves to unify theory and observations so well. There are other fudge factors, hypotheses to allow the theory to save the phenomena, in the present standard science of cosmology and in all the other branches of natural science. Research in the sciences continues to improve tests of our theories that, whether intended or not, may lead to better theories that inspire new tests. And they might on occasion replace fudge factors with unified theories in paradigms that bring parts of this enterprise closer together. It happens.

The physical cosmology that is the subject of this history is an empirical science, that is, it is based on and tested by what can be observed or measured by detectors, such as microscopes and telescopes and people. But we must pay attention to the role of theory, and intuition, and what Richard Dawid (2013 and 2017) terms "nonempirical theory assessment." The prime example in this history is that during most of the past century of research in cosmology, the community majority implicitly accepted Einstein's general theory of relativity. Few pointed out that this is an enormous extrapolation from the few meager tests of general relativity that we had in the 1960s. By the 1990s, as research in cosmology was starting to converge on a well-tested theory, there were demanding checks of the predictions of general relativity on scales ranging from the laboratory to the solar system, probing out to length scales of about 10^{13} cm. But the application to cosmology on the scale of the Hubble length, about 10^{28} cm, extrapolates from the precision tests by some fifteen orders of magnitude in length scale. This was not often mentioned, in my experience, and when mentioned, it tended to make some scientists a little uneasy, at least temporarily. In the first decades of the twenty-first century, the parts of general relativity that are relevant to the standard cosmology have passed an abundance of demanding tests. In short, the theory Einstein built on laboratory experiments was seriously tested only by the orbit of the planet Mercury. (The test of the prediction of the gravitational deflection of light by the mass of the sun, led by the people pictured in Plate III, was heavily cried up but in retrospect, their evidence seems marginal.) We find that this theory successfully extrapolates to applications on the immense scales of the observable universe. It is a remarkable result.

General relativity is an elegant extension of electromagnetism in flat spacetime; it has been said that it is a theory waiting to be found (though that is easier to say in hindsight). The faith in its extrapolation exemplifies the powerful influence and very real successes of nonempirical theory assessment. Of course, influential nonempirical assessments can mislead: Consider that in the 1930s through the 1990s, few objected to the assertions by respected experts that Einstein's cosmological constant, Λ, surely may be discarded. The evidence now is that Λ, under its new name—dark energy—is an essential part of our well-tested cosmology.

The practice of nonempirical assessments is sometimes termed "post-empiricism," but I have not found this term in Dawid's writing. Dawid (in a personal communication, 2018) states instead that

> non-empirical assessment as I understand it crucially depends on the ongoing collection of empirical data elsewhere in the research field and on the continued search for empirical confirmation of the theory under scrutiny. In a "post-empirical" phase where no substantially new data comes in any more, non-empirical assessment would get increasingly questionable and eventually would come to a halt as well.

This is consistent with what I understand to be normal practice in the physical sciences. That is, I have in mind the kind of nonempirical assessments we have been practicing all along without thinking much about it.

I take account of three other kinds of assessments: personal; community, though some may disagree; and pragmatic. The first two speak for themselves. I take examples of the third from cosmology. The usual practice has been to analyze data and observations in terms of general relativity. This surely has been due in part to the beauty of the theory, and in part to respect for Albert Einstein's magnificent intuition. But it was important also that the use of a common theory allowed comparisons of conclusions from independent analyses of the same or different data on a common fundamental ground. I do not imagine much thought has been given to this point, but I believe the implicitly pragmatic approach in cosmology (and I suppose in other branches of natural science) has helped reduce the chaos of multiple theories.

The pragmatic approach to science, if carried too far, could waste time and resources by directing research along a path as it grows increasingly clear that something is wrong. And even if the popular and pragmatically chosen path proves to be leading us in a useful direction, it can be important to have well-defended alternatives to standard ideas to motivate careful evaluations of approved ideas and observations. It may reveal corrections large or small that point toward a more profitable path. For example, a stimulating proposal in the mid-twentieth century was that textbook physics may have to be adjusted to include continual spontaneous creation of matter. The brave souls who argued for this steady-state cosmology were not always gently treated, but from what I saw, they gave as good as they got in debates over the relative merits of the general relativity and steady-state world views, arguments that were more intense than warranted by the evidence for or against either side. The idea of continual creation in the universe as it is now is no longer seriously considered in cosmology, but it had a healthy effect. New ideas can inspire defense and attacks that stimulate research, while a pragmatic defense of the old ways may help keep research from degenerating into confusion.

An important example of an implicitly pragmatic assessment is the general acceptance of Einstein's proposal that the universe is homogeneous in the average over local irregularities. Prior to the 1960s, there was scant evidence of this. Maps of distributions of the galaxies across the sky suggested instead that the galaxies are moving away from one another into space that is asymptotically empty or close to it, as in a fractal galaxy distribution. But whether by accident or design, this quite pertinent thought was put aside for the most part, and the main debate kept more sharply focused on the concepts of evolution or else a steady state of a nearly homogeneous universe. The first serious evidence for homogeneity came a half century after Einstein, from research for other purposes in the 1960s, as will be discussed in Chapter 2. Whether by good luck or good taste, the community was not much distracted by the elegant but wrong idea of a fractal universe.

It is not always easy to see why some issues receive much more attention than others; I suppose such things are to be considered eventualities. We do have reasonably clear standards for rejecting an apparently interesting idea. For example, the steady-state cosmology introduced in 1948 is elegant, but its predictions clearly violate the later accumulation of empirical tests. I do not know of a clear prescription for a move in the other direction, namely, the promotion of a working model to a standard theory. We might use the term "community opinion" to describe such decisions.

In 1990, general relativity usually was taken to be the appropriate basis for the study of the large-scale nature of the universe, but as argued above, it was an implicitly pragmatic assessment that the theory was serving well as a working basis for research. In 2003, after the revolution, the cosmological tests gave weight to the community opinion that the universe actually is well described by general relativity applied to the set of assumptions in what became known as the ΛCDM cosmological model. The introduction of these assumptions, including Einstein's cosmological constant Λ and the hypothetical cold dark matter, is reviewed in Section 8.2. Some disagreed, to be sure, but to most the accumulation of evidence (reviewed in Chapter 9) had become tight enough to have emboldened talk of what "really happened" far away and in the remote past, based on the ΛCDM theory. The notion of reality is complicated, so a more secure statement would be that whatever happened—and we assume something did happen—left traces that closely resemble those predicted by ΛCDM. And the traces are abundant and well enough cross-checked that the community opinion, including mine, is that this theory almost certainly is a useful though incomplete approximation to what actually happened.

1.2 An Overview

I have sorted this history of cosmology into lines of research that operated more or less independently of one another through stretches of time in the twentieth century. I consider the developments in each of the lines of research roughly in chronological order, but because different lines of research were at best only loosely coordinated, there have to be references back and forth in time as different lines of research started to interact. This outline is meant to explain how I have arranged the presentation of the research and how it all fits together, at least roughly, apart from the wrong turns taken.

I begin in Chapter 2 with considerations of Albert Einstein's (1917) proposal, from pure thought, that a philosophically sensible universe is homogeneous and isotropic: no preferred center or direction, no observable edges to the universe as we see it around us. That of course is apart from the minor irregularities of matter concentrated in people and planets and stars. Einstein's homogeneity is essential to the thought that we might be able to find a theory of the universe as a whole rather than of one or another of its parts. It was an inspired intuitive vision or maybe just a lucky guess; Einstein certainly had no observational evidence that suggested it. The history of how Einstein's thought was received and tested exemplifies the interplay in science between theory and practice, sometimes reinforcing each other; sometimes in serious tension; and, as in this case, sometimes aided by unexpected developments. Because I have not found a full discussion elsewhere, I consider in some detail the development of the evidence that supports what became known as Einstein's cosmological principle.

Einstein's general theory of relativity predicts that a close-to-homogeneous universe has to expand or contract. Expansion was indicated by astronomers' observations that starlight from galaxies of stars is shifted to the red, as if Doppler shifted, because the galaxies are moving away from us. Chapter 3 reviews the importance of the discovery that the Doppler shift, or redshift, is larger for galaxies that are farther away. This is the expected behavior if the universe is expanding in a nearly homogeneous way. The big bang cosmology discussed in Sections 3.1 and 3.2 uses the general theory of relativity to describe the evolution of a near-homogeneous expanding universe.

We should pause here to note that the name, "big bang," is inappropriate, because a bang connotes an event in spacetime. Unlike a familiar bang, this cosmology has nothing to do with a special position or time. The theory is instead a description of cosmic evolution of a universe that is homogeneous on average, and it attempts to follow cosmic evolution to the present from the earliest time of formation of fossils that can be observed and interpreted. That has come to include the epoch of light-element formation, when the temperature of the universe was some nine orders of magnitude larger than it is now. This is a spectacular extrapolation back in time, but not to a bang, and not to a singular start of things: We must assume that something different happened before the singularity. Simon Mitton (2005) concludes that Fred Hoyle coined the term "big bang" for a lecture on BBC radio in March 1949. It was meant as a pejorative; Hoyle favored the steady-state picture. Though unfortunate, the name "big bang" is commonly accepted. I have not encountered a better term, and the pragmatic assessment is that it is to be used in this book.

It was important that there were testable alternatives to the big bang picture; these alternatives inspired the search for tests. The leading idea, the steady-state model, is discussed in Section 3.3. It will be termed the "1948 steady-state model" to distinguish it from variants introduced later. In contrast to the prominence of the steady-state alternative to the big bang model through the mid-1960s, the leading alternative to Einstein's idea of homogeneity—a fractal distribution of matter—only became widely discussed after we at last had reasonably clear evidence of homogeneity (Section 2.6).

Hermann Bondi's (1952, 1960) book *Cosmology* in two editions, gives a valuable picture of thinking at the time. Which if either of the big bang or 1948 steady-state models, or perhaps some other model then still being considered, is the most reasonable and sensible, and on what grounds, empirical or nonempirical? Helge Kragh (1996) presents a historian's perspective of this mainstream research in cosmology up to the 1960s. Sections 3.4–3.7 augment these sources with my thoughts about the similarities and differences of assessments of the two cosmologies. I take it that in the 1950s and early 1960s, nonempirical issues account for the lack of popularity of the steady-state model in many quarters, despite its greater predictive power for observers.

The weaker predictive power of the big bang model may help account for the abundance of nonemipirical assessments discussed in Section 3.5.

The greatest effort devoted to the empirical study of the big bang cosmological model in the years around 1990 was the measurement of the mean mass density. Sections 3.6.3 and 3.6.4 review the considerable variety of these probes, and Section 3.6.5 offers an overview of what was learned. The motivation for this large effort was in part to see whether the mass density is large enough that its gravity will cause the expansion to stop and the universe to collapse, and the results were important for the empirical establishment of cosmology. But I think in large part the motivation became simply that this is a fascinating problem whose resolution is difficult but maybe not quite impossible.

The topic of Chapter 4 is the informative fossils left from a time when the universe was very different from now, dense and hot enough to produce the light elements and the sea of thermal radiation that nearly uniformly fills space. Since it was (and is) exceedingly difficult to imagine how the light elements and the radiation with its thermal spectrum could have originated in the universe as it is now, these fossils were a valuable addition to the evidence that our universe is evolving, not in a steady state. The book *Finding the Big Bang* (Peebles, Page, and Partrige 2009) recalls how these fossils were recognized in the mid-1960s, with recollections from those involved of how the recognition led to the research that produced the first good evidence that our universe really did evolve from a hot early state at about the rate of expansion predicted by general relativity. The tangled story of how Gamow and colleagues anticipated these fossils a decade before they were recognized is presented in the paper, "Discovery of the Hot Big Bang: What Happened in 1948" (Peebles 2014). Section 4.2 presents a shorter version of the main points. The sea of thermal radiation has become known as the cosmic microwave background, or CMB. The later developments leading to its central place in the revolution that established the ΛCDM cosmology are reviewed in Chapter 9. This theory of the expanding universe assumes the general theory of relativity applied to a close-to-homogeneous universe (Chapter 2), the presence of Einstein's cosmological constant Λ (Section 3.5), dark matter (Chapter 7), and particular choices of initial conditions (Section 5.2.6).

It was natural to explore how the very evident departures from Einstein's homogeneity—stars in galaxies in groups and clusters of galaxies—might have formed in an expanding universe. In the established cosmology, cosmic structure formed by the gravitational instability of the relativistic expanding universe. The early confusion about the physical meaning of this instability is an important part of the history. These considerations are reviewed in Chapter 5, along with assessments of early scenarios of how cosmic structure might have formed. The importance of these considerations for the convergence to the standard cosmology is a recurring topic throughout the rest of this book.

The subject of Chapter 6 is the astronomers' discoveries of apparent anomalies in the measurements of masses of galaxies and concentrations of galaxies. Other accounts of the exploration of these phenomena are in Courteau *et al.* (2014) and de Swart, Bertone, and van Dongen (2017). Fritz Zwicky was the first to recognize the phenomenon: He saw that the galaxies in the rich Coma Cluster of galaxies seem to be moving relative to one another too rapidly to be held together by the gravitational attraction of the mass seen in the stars in the galaxies in the cluster. One way to put it is that the mass required to hold this concentration of galaxies together by gravity seemed to be missing, always assuming the gravitational inverse square law of gravity (in the nonrelativistic Newtonian limit of general relativity). It was later seen that mass also seemed to be missing from the outer parts of spiral galaxies, based on the measurements discussed in Section 6.3 of circular motions of stars and gas in the discs of spiral galaxies. Much the same conclusion came from the studies described in Section 6.4 of how galaxies with prominent discs acquired their elegant spiral patterns. By the mid-1970s, it had become clear that understanding this is much easier if the seen mass is gravitationally held in near-circular motion in the disc with the help of the gravitational attraction of less-luminous matter that is more securely stabilized by more nearly random orientations of the orbits.

These observations pointed to a key idea for the establishment of cosmology: the existence of "dark matter," the new name for what was variously known as "missing," "hidden," or "subluminal" mass. The idea came almost entirely out of pursuits in astronomy, not cosmology, and for this purpose, the subluminal component need not be very exotic: low-mass stars would do, though they would have to be present in surprising abundance relative to counts of the more luminous observed stars. But in the 1970s, another key idea for cosmology was growing out of particle physicists' growing interest in the possible forms of nonbaryonic matter. Gas and plasma, people, planets, and normal stars are all forms of what is termed "baryonic matter." Most of the mass of baryonic matter is in atomic nuclei; the accompanying electrons are termed "leptons," but they are also counted in the mass of baryonic matter. The neutrinos are leptons that we now know have small but nonzero rest masses. Thus they act as nonbaryonic dark matter that contributes to the masses of galaxies, but in the standard cosmology, this contribution is much smaller than the total indicated by the astronomical evidence. We need a new kind of nonbaryonic matter.

The thought that the astronomers' subluminal matter is the particle physicists' nonbaryonic matter and the cosmologists' dark matter was and remains a conjecture at the time of writing. The only empirical evidence of the new nonbaryonic dark matter is the effect of its gravity. It has been a productive idea, however, that passes demanding checks. The particle physicists' considerations of nonbaryonic matter reviewed in Chapter 7 takes into account the condition that if this nonbaryonic matter were produced in the hot early stages of expansion of the universe, then its remnant mass density must not exceed that allowed by the relativistic big bang cosmological model (again, assuming the relativistic theory). But it is notable that cosmologists took over the notion of nonbaryonic dark matter before the particle physics community had taken much interest in the astronomers' evidence of the presence of subluminal matter.

The nonbaryonic dark matter most broadly discussed in the 1980s came in two varieties, cold and hot. The latter would be one of the known class of neutrinos with rest mass of a few tens of electron volts (Sections 5.2.7 and 7.1). The initially hot (meaning rapidly streaming) neutrinos in the early universe would have smoothed the mass distribution, and that smoothing would have tended to cause the first generation of structure to be massive systems that must have fragmented to form galaxies. The spurious indication in 1980 of a laboratory detection of a neutrino mass appropriate for the hot dark matter picture certainly enhanced interest in the indicated formation of galaxies by fragmentation. This model was considered but had to be rejected: the observations show hierarchical growth of structure, from smaller to larger mass distributions.

The prototype for the nonbaryonic matter that is an essential component of the established cosmology was introduced by particle physicists in 1977. The idea occurred to five groups who published in the space of 2 months. These papers do not exhibit much interest in the astronomers' subluminal mass phenomena, but the considerations certainly were relevant to subluminal matter. Was this a curious coincidence or an idea that somehow was "in the air?" This is considered a little further in Sections 7.2.1 and 10.4.

Sections 8.1 and 8.2 review why in the early 1980s cosmologists co-opted the astronomers' subluminal mass and the particle physicists' nonbaryonic matter in what became known as the standard cold dark matter, or sCDM, cosmological model. The letter "s" might be taken to mean that the model was designed to be simple (as it was) but it instead signified "standard," not because it was established but because it came first. It was meant to distinguish this version from the many variants to be considered in Section 8.4. A large part of the cosmology community soon adopted variants of the sCDM model as bases for exploration of how galaxies might have formed in the observed patterns of their space distribution and motions (Section 8.3), and for analyses of the effect of galaxy formation on the angular distribution of the sea of thermal radiation. This widespread adoption was arguably overenthusiastic, because it was easy to devise other models, less simple to be sure, that fit what we knew at the time. And it was complicated by the nonempirical feeling that space sections surely are flat. In general relativity that could be because the mass density is large enough to produce flat space sections, or because Einstein's cosmological constant, Λ, makes it so. The nonempirical reasons for preferring flat space sections, preferably without resorting to Λ, are discussed in Section 3.5. These reasons were influential and long-lasting enough to have played a significant role in the confusion of variants and alternatives to the sCDM idea considered in the 1990s.

The reduction of confusion in the years 1998–2003 was great enough to be termed a revolution. It was driven by the two great experimental advances discussed in Chapter 9. The first is the measurement of the relation between the redshift of the spectrum of an object and its brightness in the sky, given its luminosity: the cosmological redshift–magnitude relation. Its detection had been a goal for cosmology since the 1930s; it was at last accomplished by two independent groups at the turn of the century (Section 9.1). The second is the detailed mapping of the angular distribution of the CMB radiation. Work on this began in the mid-1960s, and coincidently also produced demanding constraints on cosmological models at the turn of the century. These results from the two sets of measurements, together with what was already known, made a tight case for the presence of Einstein's cosmological constant Λ and the nonbaryonic CDM in the relativistic hot big bang ΛCDM theory. It was a dramatic development.

It was proper to have asked whether the introduction of two very significant hypothetical components, CDM and Λ, along with all the other assumptions that go into the choice of a cosmological model, might only amount to adjusting the theory to fit the measurements. That line of debate did not become very prominent, because the ΛCDM cosmology that fit the two critical measurements brought together so many other lines of evidence in a tight network of empirical tests. This is the topic of Section 9.3.

By the year 2003, the community had at last settled on a respectably well-supported theory of the large-scale nature of the universe. Skeptics remained, as is appropriate, for this theory is an immense extension of the reach of established physics. Indeed, the 2003 theory has been modified to fit later measurements, but these changes amount to fine adjustments of parameters, not challenges to the basic framework of the theory. It is the nature of science to advance by successive approximations, and it would not be at all surprising to find that there is a still better theory than ΛCDM. But we have excellent reason to expect that a better theory will describe a universe that behaves much like ΛCDM, because ΛCDM passes an abundance of empirical tests that probe the universe in so many different ways.

I cannot think of any lesson to be drawn from this story of how cosmology has extended the boundaries of established science that cannot be drawn from other branches of natural science. This is no surprise, because cosmology operates by the methods of natural science. But I think there are lessons to be drawn with greater clarity in the relatively uncluttered historical development of this subject. My offerings are given in Chapter 10.

CHAPTER TWO

The Homogeneous Universe

MODERN COSMOLOGY GREW out of Albert Einstein's search for how his general theory of relativity might apply to the large-scale nature of the universe. Einstein's (1917) thought was that a philosophically reasonable universe is the same everywhere and in all directions, apart from minor irregularities, such as the observed concentrations of matter in planets and stars. This is a distinct departure from the tradition of research in natural science, which is to select for examination a level in a hierarchy of structure. It may be the examination of molecules; the atoms in molecules; the nuclei in atoms; the nucleons in nuclei; or the quarks and gluons in nucleons. One can examine structure on larger scales: the vast complexity of interactions of atoms and molecules in condensed matter, chemistry, and on up to biophysics; or the natures of planets around stars, stars in galaxies, or galaxies in groups and clusters and superclusters of galaxies. Einstein's thought was that this hierarchy of structures ends in something new to modern science: large-scale homogeneity. (Although not stated explicitly at first, the thought includes large-scale isotropy. That is, the universe is assumed to be invariant under rotations as well as translations.)

Einstein's homogeneity assumption allows us to consider and test the possibility of a theory of the universe as a whole, rather than a theory of a particular level in a hierarchy. If the universe is homogeneous in the large-scale average, then observations from our position may inform the theory of what the universe is like when observed from any other place. But we need evidence that this approximation is useful.

2.1 *Einstein's Cosmological Principle*

Einstein's (1917) original argument for the picture of large-scale homogeneity is difficult to assess. He argued against the idea that the material content of the universe might be confined to a single concentration, an island universe in otherwise empty space. If this were so, and the escape velocity were finite, then stars would evaporate, escaping the island universe. This behavior would be contrary to his implicit assumption that the universe is in a stationary state. If the escape velocity were arbitrarily large, then statistical relaxation would produce the occasional star moving with arbitrarily large speed. This might be taken to be contrary to the observation that the velocities of nearby stars are much smaller than the velocity of light. Both points would make some sense if the universe were not evolving and the stars had had time to approach statistical equilibrium. Einstein does not seem to have paused to consider that if energy is conserved, then the stars must eventually stop shining. And if stars nevertheless shine forever, then his homogeneous universe would be full of starlight. This is Olbers' paradox, and is certainly an unacceptable situation.

The argument that may be closer to what Einstein was thinking in 1917 is stated in *The Meaning of Relativity*, the publication of his lectures at Princeton University in 1921 (Einstein 1923). He pointed out that his general relativity allows solutions in which there is a single mass concentration outside of which spacetime is empty and asymptotically flat, or as Einstein put it, quasi-Euclidean. Motions of matter in this mass concentration would have the usual properties of acceleration, such as the flattening of a gravitationally bound rotating galaxy. But in a nonrelativistic mass concentration, this rotation would be relative to empty spacetime. Thus Einstein (1923, 109) wrote: "If the universe were quasi-Euclidean, then Mach was wholly wrong in his thought that inertia, as well as gravitation, depends upon a kind of mutual action between bodies."

A similar sentiment, expressed in Einstein (1917), is that (in an English translation): "In a consistent theory of relativity there can be no inertia *relative to "space,"* but only an inertia of masses *relative to one another.*" He went on to point out that in his general relativity, a single particle of mass in otherwise flat spacetime would have inertia, contrary to his stated view of relativity.

Einstein (1923, 110) argued that it is "probable that Mach was on the right road" in the relativity of inertia, and cited three examples:

1. The inertia of a body must increase when ponderable masses are piled up in its neighborhood.
2. A body must experience an accelerating force when neighboring masses are accelerated, and, in fact, the force must be in the same direction as the acceleration.
3. A rotating hollow body must generate inside of itself a "Coriolis field," which deflects moving bodies in the sense of the rotation, and a radial centrifugal field as well.

With all respect to Einstein's genius, we must observe that the first example, if meant as a local measurement, may follow from Mach's principle, but it is not true in general relativity. This theory predicts that an observer confined to a space small enough that tidal fields may be neglected sees the same universal local physics, including the usual properties of inertia, whatever the environment. An operational meaning of the second example seems to be equivalent to the third. This is the Lense-Thirring effect: An inertial frame of reference near a rotating massive body rotates relative to distant matter as if the inertial frame were dragged by the rotation of the massive body. The effect has since been observationally checked.

The prediction in general relativity is in line with the thought that acceleration, like motion, surely is meaningful only relative to what the rest of the universe is doing. This certainly seems to be the direction of Ernst Mach's thinking (as expressed in his book, *Die Mechanik in Ihrer Entwicklung Historisch-Kritisch Dargestellt,* and on pages 283–285 in the English translation in Mach 1960). And we must consider that Einstein's reading of what he termed "Mach's principle" led him to an idea that is now clearly established: The observable universe is very close to homogeneous. Debate continues on whether Einstein was right about this for the right reason.

To make acceleration relative within general relativity, Einstein had to remove the possibility of a quasi-Euclidean universe. He did so by proposing as a sort of boundary condition that the universe is homogeneous: It has no preferred center and no edges. Space is to be pictured as nearly uniformly filled everywhere with matter and radiation.

The paper by Willem de Sitter (1917a, 3) gives some indication of Einstein's thinking:

> The most desirable and the simplest value for the $g_{\mu\nu}$ at infinity is evidently zero. Einstein has not succeeded in finding such a set of boundary values[1] and therefore makes the hypothesis that the universe is not infinite, but spherical: then no boundary conditions are needed, and the difficulty disappears. . . . The idea to make the four-dimensional world spherical in order to avoid the necessity of assigning boundary-conditions, was suggested several months ago by Prof. Ehrenfest, in a conversation with the writer. It was, however, at that time not further developed.

(I cannot follow the comments in de Sitter's footnote.) Spherical space, closed as is the surface of a sphere, has no boundary on which we must assign conditions, and it can be assumed to be close to homogeneous. We see that the bold and eventually successful idea of homogeneity grew out of some mix of philosophy and intuition, supplemented by interactions with colleagues and perhaps aided by some measure of wishful thinking. It certainly was not based on any empirical evidence.

Edward Arthur Milne recognized the power of homogeneity in formulating a cosmology, and he named the assumption "Einstein's cosmological principle." Milne (1933) showed that, independent of general relativity, this principle with

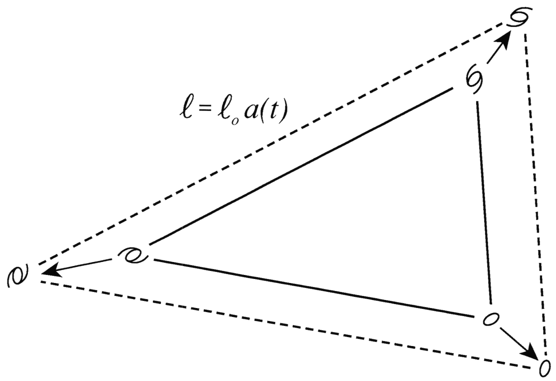

FIGURE 2.1. Homogeneous and isotropic expansion (Peebles 1980).

standard local physics accounts for a central feature of cosmology: the relation between the recession velocity v of a galaxy and its distance r,

$$v = cz = H_0 r, \tag{2.1}$$

where H_0 is the constant of proportionality. To see this, write the velocities of the galaxies as the vector relation $\vec{v} = H_0 \vec{r}$. Then an observer on galaxy a sees galaxy b moving away at velocity

$$\vec{v}_b - \vec{v}_a = H_0(\vec{r}_b - \vec{r}_a). \tag{2.2}$$

This shows that all observers see this same pattern of recession of the other galaxies, as required by homogeneity.

The expansion rate H_0 is known as Hubble's constant. The subscript is meant to indicate that H_0 is a measure of the present rate of expansion of the universe; in an evolving cosmology, the expansion rate is a function of time. Equation (2.1) is known as the redshift-distance relation, where the redshift z is defined in equation (2.1) for recession speeds well below the speed of light.

The redshift-distance relation is commonly termed Hubble's law. A vote by members of the International Astronomical Union would rename it the Hubble-Lemaître law, in recognition of Lemaître's prediction (discussed in Section 3.1). Others could have been named, too. Vesto Melvin Slipher's redshift measurements and Henrietta Leavitt's Cepheid period-luminosity relation were essential to Hubble's (1929) redshift-distance plot, and Milton Humason's redshift measurements in the 1930s were essential to establishing a clear and tight demonstration of the effect.

For another way to understand Milne's point, consider the three galaxies at the vertices of the triangle in Figure 2.1. If the galaxies are moving away from one another in a homogeneous and isotropic way, the angles of the triangle are unchanged while the length ℓ_i of each side increases by the same factor, $\ell_i \propto a(t)$. This has to be true of any triangle. That is, $a(t)$ is a universal expansion factor. With $l \propto a(t)$, the rate of change of the physical distance $l(t)$ between any two galaxies at separation l is

$$\frac{dl}{dt} = v = \frac{\dot{a}}{a}\ell(t). \tag{2.3}$$

The dot means time derivative. We see that Hubble's constant in equation (2.1) is

$$H_0 = \frac{1}{a}\frac{da}{dt}, \tag{2.4}$$

evaluated at the present epoch, at expansion time $t = t_0$.

The departure of a galaxy velocity from the mean value set by Hubble's law at that position is said to be the galaxy peculiar velocity. Peculiar velocities usually may be attributed to the gravitational pull of the growing clustering of mass in galaxies and concentrations of galaxies, but nongravitational forces produced by explosions may be important, too.

At nonrelativistic recession speeds, the cosmological redshift is defined as $z = v/c$, where c is the speed of light. This is a first-order Doppler shift. The distance at which Hubble's relation between distance and recession velocity extrapolates to the speed of light, $r_{\mathrm{H}} = cH_0^{-1} \sim 10^{28}$ cm, is the Hubble length. Consideration of the relativistic correction to equation (2.3) for galaxies at this great distance begins in Section 3.2.

2.2 Early Evidence of Inhomogeneity

In the 1930s, the cosmological principle passed an important empirical check: The prediction from homogeneity of the redshift-distance relation in equation (2.1) was shown to fit the tight tests discussed in Section 2.3. But homogeneity was not suggested by maps of the galaxy distribution. Charlier (1922) presented a map of the distribution across the sky of the known nebulae. Among the objects in Charlier's map are clusters of stars in our galaxy, and regions where starlight is reflected by clouds of dust, but most are extragalactic nebulae, that is, other galaxies of stars. Charlier pointed out that the map brings to mind hierarchical clustering: galaxies appear in clumps that are present in clumps of clumps, and so on, perhaps to indefinitely large scales. This was later named a "fractal universe."

A decade later, Harlow Shapley and Adelaide Ames at the Harvard College Observatory presented a catalog of the 1,249 known galaxies brighter than $m = 13$ (a measure of the brightness in the sky). Their maps of the angular positions in the two hemispheres of our galaxy are shown in Figure 2.2 (Shapley and Ames 1932). The left-hand panel shows the galaxies in the North hemisphere of our galaxy, the right-hand panel those in the South galactic hemisphere. The near absence of galaxies near the plane of our Milky Way galaxy is due to absorption of light by interstellar dust lying close to the plane

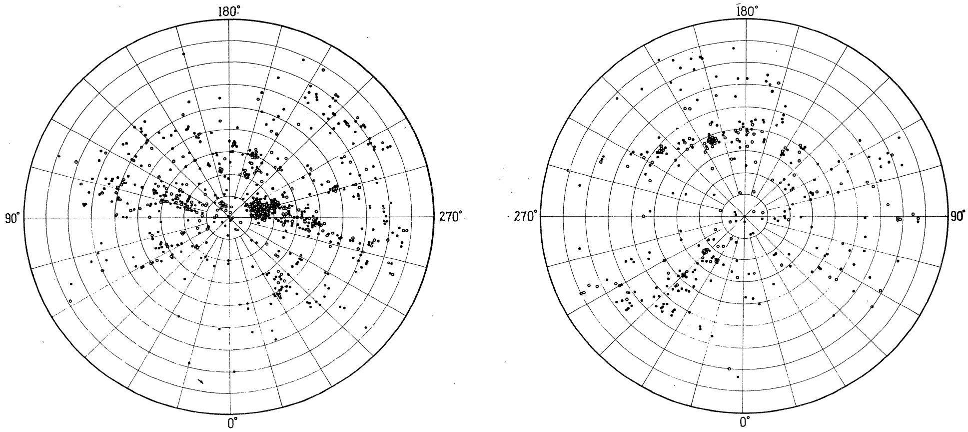

FIGURE 2.2. The Shapley and Ames (1932) map of galaxies brighter than apparent magnitude 13. Courtesy of the John G. Wolbach Library, Harvard College Library.

of our galaxy. The sky is clearer above and below the plane. The northern hemisphere on the left in the figure shows the many galaxies in the prominent concentration in and around the Virgo Cluster of galaxies. (The cluster is named for its position in the sky, near the stellar constellation Virgo.) De Vaucouleurs (1953 and 1958a) named the Virgo Cluster and the broad concentration of galaxies around it the Local Supercluster. This distinctly inhomogeneous distribution of the nearby galaxies is well established.

Willem de Sitter (1917a,b) presented discussions of Einstein's thoughts about the structure of the universe. Since de Sitter was a knowledgeable astronomer, he could have told Einstein about the nebulae, the thought that most are extragalactic, and the evidence that these extragalactic nebulae are not at all close to uniformly distributed. But I have not seen any indication that Einstein considered this observation and if so, whether it affected his thinking.

The possibilities in 1917 were that obscuration by dust is quite patchy even well away from the plane of our galaxy, or else that the observed distribution of galaxies does not at all resemble the homogeneity of the cosmological principle. Not much had changed by the 1950s except that the dust option was ruled out. The situation was recognized in the influential and informative book, *The Classical Theory of Fields* (Landau and Lifshitz 1951, the English translation of the 1948 Russian edition). It presents an admirable exposition of the special and general theories of relativity, but there is little mention of data in this book or in the others in their series on theoretical physics. A rare exception is the comment about Einstein's homogeneity assumption in Landau and Lifshitz (1951, 332):

> Although the astronomical data available at the present time give a basis for the assumption of uniformity of this density, this assumption

> can of necessity have only an approximate character, and it remains an open question whether this situation will not be changed even qualitatively as new data are obtained, and to what extent even the fundamental properties of the solutions of the equations of gravitation thus obtained agree with actuality.

As we see from Figure 2.2, this was a sensible remark, though from an empirical point of view, one might have expected another caution about the scant tests of general relativity. The situation in gravity physics was quite different from the empirical situation in the first part of their book, on the very well tested and broadly applied theory of electromagnetism.

In a report to the eleventh Solvay conference, *La structure et l'évolution de l'univers,* Oort (1958) began with the statement that "One of the most striking aspects of the universe is its inhomogeneity." As evidence, he showed the Shapley and Ames (1932) map in Figure 2.2. He could have added that Abell's (1958) catalog of the more-distant rich clusters of galaxies shows them scattered across the sky in a clumpy fashion, as in superclusters of clusters. But the distribution of clusters in Abell's map (1958, Figure 7) does look distinctly less clumpy than the distribution of the much closer galaxies in the Shapley-Ames map.

2.3 Early Evidence of Homogeneity: Isotropy

There remained the possibility that the galaxies are uniformly distributed in the average over larger volumes than Shapley and Ames had sampled. Hubble (1926 and 1934) introduced a test, the variation of the counts of faint galaxies as a function of position across the sky. Away from the areas obscured by interstellar dust close to the plane of the Milky Way, Hubble (1934) typically found about 100 galaxies per square degree (reduced to standard observing conditions) to a limiting redshift he estimated to be about $z = 0.1$. This is deep, 10 percent of the speed of light, and is about ten times the distance sampled in the Shapley-Ames map. Hubble's counts at low galactic latitudes, plotted as the lower strings of data in Figure 2.3, are smaller than at high latitudes and show a systematic variation across the sky. Both are effects of obscuration by dust in variable amounts along lines of sight near the directions of the plane of the Milky Way. The upper strings of data are counts at 40–50 degrees above the plane, plotted as filled circles in the north galactic hemisphere and open circles in the south. The counts are similar in the two hemispheres and do not show a systematic tendency to vary with position across the sky. Hubble (1934, 62) concluded that

> On the grand scale, however, the tendency to cluster averages out. The counts with large reflectors conform rather closely with the theory of

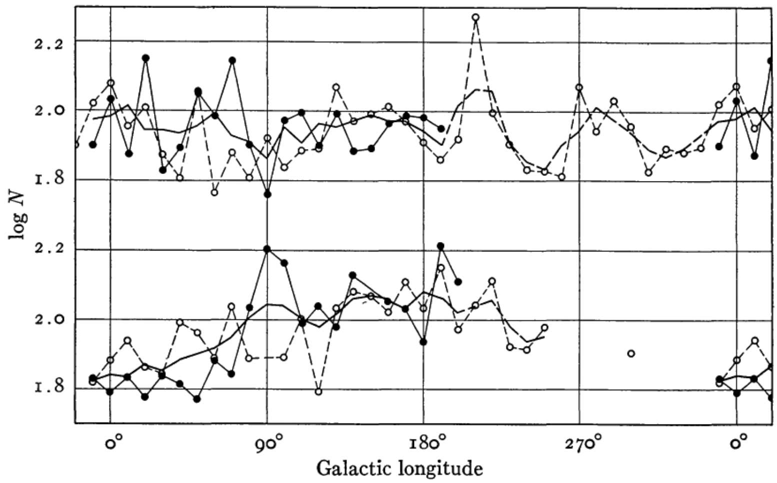

FIGURE 2.3. Hubble's (1934) counts of galaxies at high galactic latitudes in the upper curves, and at low latitudes in the lower curves. © AAS. Reproduced with permission.

> sampling for a homogeneous population. Statistically uniform distribution of nebulae appears to be a general characteristic of the observable region as a whole.

Bok's (1934, 8) considerations led him to the opposite conclusion:

> Different lines of evidence all indicate that the available material points to the existence of a widespread non-uniformity in the distribution of external galaxies, and that this tendency toward clustering is probably one of the chief characteristics of the part of the Universe within the reach of modern telescopes.

Bok was at the Harvard College Observatory, and he emphasized the clumpy distribution of galaxies in the Harvard Shapley-Ames map that came out of this observatory. He referred to Hubble (1934) but did not mention Hubble's Figure 4, which is reproduced here in Figure 2.3. Hubble took it to be an indication of approach to uniformity; Bok does not seem to have been convinced.

Hubble's interpretation seems to be the more reasonable to me, and I count it as the first indication that in the average over large enough volumes, the galaxy distribution approaches isotropy. That is easier to see now, of course. And it is easier to see that if we may take it that our position among the galaxies is not special, then the indication from this figure, though certainly preliminary, was that the galaxy distribution approaches homogeneity on large scales.

Another line of evidence opened in the 1950s with the ability to probe the universe at radio wavelengths and soon after that by X-ray and microwave detectors. Figure 2.4 shows the distribution of radio source positions across

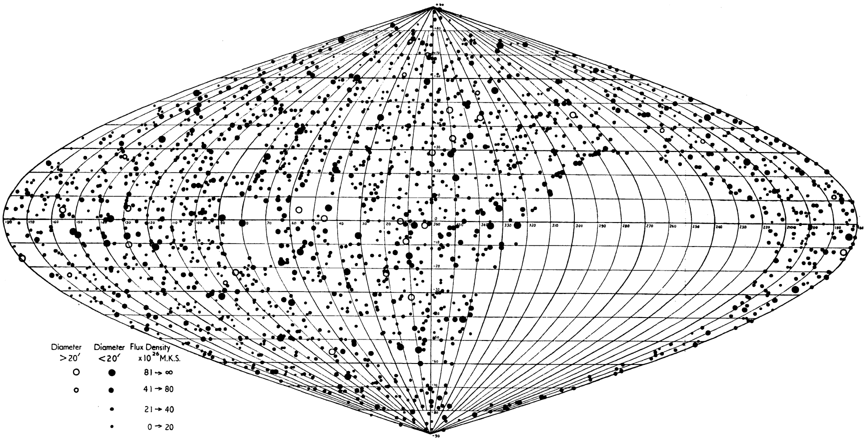

FIGURE 2.4. The Second Cambridge Catalog of Radio Sources (Shakeshaft, Ryle, Baldwin, et al. 1955).

the part of the sky surveyed in the *Second Cambridge Catalog of Radio Sources, 2C,* by Shakeshaft et al. (1955).[1] The sources were suspected then and are now known to be in galaxies. The catalog lists 1,936 sources at wavelength 3.7 meters (82 MHz). A few are close to the plane of the Milky Way and likely are in our galaxy. Others are spurious detections of sources in sidelobes, and some real sources are missing. The brightest extragalactic radio source in the sky, Cygnus A, is on the equator in this map and a quarter of the way in from the left-hand side. It is so bright in the radio that it obscures sources close to it in the map, accounting for the empty region around this object. (The large empty region to the lower right was not observed, because it always is below the horizon at the telescope.)

Optical identifications and redshift measurements of a few of these sources had suggested that many have redshifts large enough that the observations might show a detectible departure of the count of sources as a function of the radio flux density from what would be expected in the flat spacetime of special relativity. This is the cosmological test to be discussed in Section 3.4. Its application here was frustrated by spurious source detections and omissions. This systematic error has a less serious effect on the angular distribution of sources, however, and we see that the constant-area map of sources in Figure 2.4 does look about as expected in a homogeneous universe: no indication in any direction that the observations encounter an edge to the distribution of these objects.

We are in seas of X-ray and microwave radiation. The latter, later termed the "cosmic microwave background" (CMB), is the subject of Chapter 4. A 6-minute rocket flight gave the first evidence of the former, a sea of X-rays

1. This is the figure between pages 148 and 149 in Shakeshaft et al. (1955).

(Giacconi et al. 1962). The flight allowed little time for measurement of the X-ray angular distribution, but Gould's (1967) review indicated that this radiation does not vary across the sky by more than about 10 percent. Schwartz (1970) used the year-long scan of the sky by the OSO-III X-ray satellite (at 7.6 to 38 KeV, and angular resolution $\sim 10^\circ$) to bound the X-ray anisotropy to 4 percent.

Recognition of the other component, the sea of microwave radiation, was presented by Penzias and Wilson (1965). By the end of the 1960s, Wilson and Penzias (1967) and Partridge and Wilkinson (1967) had found that this radiation is isotropic to better than 0.2 percent.

The isotropy observed at optical, radio, microwave, and X-ray wavelengths seriously constrains ideas about the large-scale nature of our universe. The maps in Figures 2.3 and 2.4 require that, if the space distribution of galaxies is not close to homogeneous, then it is at least close to spherically symmetric about our position. That would seem to be a curious arrangement of matter, and it would be curious, too, that there are enormous numbers of other galaxies that would seem to have equally suitable homes for observers such as us. It is difficult to imagine we would be so special as to be close to the center of symmetry. The easier interpretation is that our universe is close to homogeneous, meaning observers on other galaxies also would see isotropic distributions of sources.

Another picture to be considered is that the X-ray and microwave radiation backgrounds are isotropic because the universe contains a homogeneous sea of radiation that has nothing to do with the galaxy distribution. This might work if spacetime is static. But if we accept the evidence from galaxy redshifts that the galaxies are moving apart, then most would have to be moving through a uniform sea of radiation. The Doppler effect would cause an observer moving through the radiation to find that the radiation is brighter than average in the direction the galaxy is moving and dimmer in the opposite direction. The radiation we observe is close to isotropic, so again we would have to be in an exceedingly special galaxy, one of the few that are moving slowly through the radiation. Most galaxies would be moving through the sea more rapidly, the most distant at near-relativistic speeds. Why should we be in this special situation?

Another situation we might consider accepts that the radiation is uniformly distributed in a curved spacetime that describes homogeneous and isotropic space sections, consistent with the cosmological principle, but that the galaxies are distributed in a clumpy fashion even on arbitrarily large scales, as in a fractal distribution. This picture might have been defended in the 1960s by supposing that the X-ray background did not come from the galaxies and that we could ignore the gravitational disturbance to spacetime caused by the mass in the regions occupied by the galaxies. Issues of this sort were discussed by Wolfe and Burbidge (1970) and Peebles (1971a). The conclusion is that it is difficult to imagine a model for clumping of matter on scales