The Oxford Handbook of
THE HISTORY OF
MODERN COSMOLOGY
THE OXFORD HANDBOOK OF THE HISTORY
OF MODERN COSMOLOGY
Frontispiece: The whole-sky map of the cosmic microwave background radiation (CMB) as observed by the Planck satellite of the European Space Agency (ESA). The image shows large-scale structures in the Universe only 380,000 years after the Big Bang. It encodes a huge amount of information about the cosmological parameters which describe our Universe. (Courtesy of ESA and the Planck Collaboration)
The Oxford Handbook of the History of Modern Cosmology

Edited by

Helge Kragh
Malcolm S. Longair
In 1967 and 1975 the two pioneering Russian astrophysicists Yakov Zeldovich and Igor Novikov published two monographs under the common title *Relyativistskaya Astrofizica* (Relativistic Astrophysics) which were subsequently translated into English. The second of the volumes, titled *The Structure and Evolution of the Universe*, was a comprehensive account of modern cosmology with an emphasis on the victorious big-bang model. In their introduction Zeldovich and Novikov stated that ‘Our philosophy is that the history of the Universe is infinitely more interesting than the history of the study of the Universe.’ This is a defensible view but not the philosophy underlying the present handbook on the history of modern cosmology. This volume aims at describing in detail how our present understanding of the universe has emerged through a long and complex series of investigations with roots back in the nineteenth century and even earlier.

There are indeed two ways in which one can speak of the history of the universe. We know today and have known since about 1930 that the universe has its own physical history, in the sense that it evolves over time and can be ascribed a definite age. The details of this evolution, ranging from the big bang to the present and even into the future, are the business of astronomers, astrophysicists and cosmologists. But evolution is not the same as history in the ordinary sense of the term. Darwin’s masterwork of 1859 carried the title ‘The Evolution of Species’ and not ‘The History of Species’. History of cosmology is not primarily concerned with stars, galaxies, curved space, and dark energy, but with the scientists who explored the universe and the collective results of whom have provided the picture of the universe accepted at any given time. The ‘history’ of the universe is without a doubt highly interesting but not necessarily more interesting than the history of the study of the universe. Moreover, the two kinds of history are evidently closely connected, their relationship being synergetic rather than contradictory. It is only through the historical approach that we recognize that our present picture of the universe is not inevitable but the result of a long, bumpy and contingent development which might conceivably have resulted in a very different picture.

It is worthwhile contemplating the terminology and changed meanings of some of the key concepts of cosmology. First, to the ancient Greeks the term *cosmos* (κόσμος) carried connotations such as order, regular behaviour, and beauty—it is no coincidence that the term appears also in cosmetics and cosmetology. The Greeks boldly claimed that the universe as a whole is a cosmos rather than a chaos and for this reason it must be possible to understand it rationally. Cosmology must be possible. It was a wildly, almost reckless claim but more than two thousand years later it turned out to be more than just a beautiful dream. While cosmology was for a long time the playground of speculative natural philosophy, today it is no less scientific than other branches of the physical sciences. And yet it is and will always remain a different science, principally
because the domain of cosmology is the universe at large, a unique and epistemically extraordinary concept.

The universe or cosmos is everything that has, has had, or will have physical existence, whether matter, energy, space, or time. There is nothing outside the universe and literally everything of a physical nature is inside it. Cosmology in the traditional sense refers to the study of the structure of the universe, what in the seventeenth century was often known as cosmography, a term which stresses the mapping of the universe and which could also refer to what we would consider as geography today. Although the term is rarely used today, much of modern cosmology is in the older cosmographical tradition except that the universe under study is no longer assumed to be static. Edwin Hubble, aptly described by his biographer Gale Christianson as a ‘mariner of the nebulae’, was as much a cosmographer as a cosmologist. The dynamic conception of the universe is largely a product of the twentieth century, and yet scientists of the past often dealt with cosmogony, which literally means the study of how the universe came to be what it is. The earliest cosmological views we know of, those of the Mesopotamian and Egyptian cultures, were cosmogonies rather than cosmographies. They were mythical tales of how the world and the gods came into existence, not science.

While the name cosmology does not in itself include a temporal dimension, cosmogony does. However, from a modern perspective older terms such as cosmography and cosmogony can be misleading as they often referred to the description and formation of the planetary system and not to the universe as a whole. Henri Poincaré’s important work of 1913 entitled *Hypothèses Cosmogoniques* (Cosmogonic Hypotheses) was primarily concerned with theories of the formation of the planetary system and had almost nothing to say about cosmology in the modern meaning of the term.

These and other aspects of a semantic nature are not unimportant when it comes to a historical understanding of how the science of the universe has progressed. They illustrate the simple point known from all branches of culture that words come and go. Moreover, although the same word may be used during a very long period of time, in many cases the meaning of the word changes and sometimes drastically so. Just as the ‘electron’ of 1900 was very different from what physicists called an electron a few decades later, so the name ‘galaxy’ changed its meaning during roughly the same period. For practical reasons we retain the name universe, but we should not believe that when astronomers wrote about the universe in the early twentieth century they referred to the same concept that the name encompasses today.

There is another point with regard to the name cosmology that needs to be commented upon. While in the present context the name refers to the science of the universe, in other and equally legitimate contexts it refers to a world view or ideology common to a particular society or culture. Historians and anthropologists speak of the romantic ideology or the communist ideology; and they investigate the cosmology of Australian aboriginals or that of the Hopi Indians. In a broad historical perspective the two widely different meanings of cosmology were originally connected, but this is no longer the case. We generally have no problem distinguishing scientific cosmology from what may be loosely called philosophical cosmology. This volume is about the former and not about the latter kind of cosmology. Nonetheless, even today one cannot completely separate the philosophical-religious aspects of cosmology from the scientific aspects. As illustrated in
some of the chapters, changes in scientific cosmology have often resulted in discussions of a much wider nature related to philosophical, political and religious issues.

When did the universe become modern? The question is largely equivalent to asking when the study of the universe as a whole became a science based on mathematical models combined with astronomical observations and deductions from physical theory. The seeds for a new and immensely fruitful chapter in the history of cosmology were planted in the second half of the nineteenth century when the galaxies moved to the forefront of astronomical research and the art of spectroscopy made it possible to study the heavenly bodies from the perspective of physics and chemistry. During the same period the second law of thermodynamics suggested that a cosmic arrow of time might be ascribed to the universe. And yet all this was only prolegomena to the later development. It was Einstein’s cosmological field equations of 1917 which provided a solid theoretical framework for understanding the universe at large and which today are often hailed as the true beginning of modern cosmology at least as far as theory is concerned. Based on the theory of general relativity and Hubble’s measurements of extragalactic nebulae the traditional static picture of the universe was eventually replaced by the then mind-boggling idea of a universe in expansion.

The further development from the Friedman–Lemaître expanding universe to the present standard model was far from smooth as it followed a route with many false trails and ideas later recognized to be wrong. These false trails and wrong ideas are as much part of the history of modern cosmology as are the developments which today enter textbooks as the milestones that led to our present picture of the universe. From a proper historical perspective the important developments were those which were considered important at that time and not those which current cosmologists recognize as the preferred cosmological model. Consequently a substantial part of the chapters in this volume deals with theories and observational claims which are not only outdated but wrong.

Until recently the historiography of modern cosmology and in particular the development in the post-World War II period was an underdeveloped field of history of science. The low priority was evident not only if compared to the history of other branches of the modern physical sciences but even more so in comparison to the hundreds or thousands of scholarly works devoted to ancient cosmology or the Copernican revolution. Still by 1980 and possibly even today the situation was that our knowledge of how the heliocentric system of the world came into existence was more complete and fine-grained than our knowledge of the emergence of big-bang theories in modern cosmology. And yet it would be difficult to argue that the twentieth-century picture of the evolution of the universe is less revolutionary than the pictures constructed by Ptolemy, Copernicus, and Kepler.

There were a few works which covered in a comprehensive of scholarly manner the post-war development, such as John North’s *The Measure of the Universe* and Jacques Merleau-Ponty’s *Cosmologie du XXème Siècle*. But unfortunately both of these works were completed just before the magic year of 1965 when the cosmic microwave background (CMB) radiation was discovered. During the last two or three decades the situation has improved markedly with the publication of many books and articles describing the developments from about 1965 to the present. Although today there is no shortage of
literature on the history of modern cosmology, by far most of it is either of a specialized nature or belonging to the genre of popular history of science.

The present volume aims at presenting a comprehensive overview of the development of cosmology from the late nineteenth century to the early twenty-first century and to include not only the necessary scientific details but also the broader contexts. Although comprehensive, of course it does not give a complete account of the history. One of the issues which are missing is a sociologically oriented analysis of how and when cosmology became professionalized and accepted as a proper scientific discipline with its own journals, associations, and reward system. After all, modern cosmology is more than just a subfield of astronomy and physics. Unfortunately the social aspects of modern cosmology are a subject which has received almost no serious attention from historians and sociologists of science.

The volume describes and explains the historical background to what we know about the universe today but does so in a non-linear, more historically authentic way that pays attention also to the many developments that turned out to be unfruitful or just wrong. The book is organized in thirteen roughly chronologically ordered chapters, with some focusing on theory and others more on observations, experiments and technological advances. A few of the chapters are of a more general nature, relating to larger contexts such as politics, economy, philosophy and world views. While each chapter can be read separately it also connects to the other chapters, and the book thus aims at describing the history coherently and as a whole. The content and style of the chapters differ, of course, but they all form part of the same grand story.

The authors come from diverse backgrounds, some professional astrophysicists and cosmologists who have lived through the exciting and dramatic developments of the last 50 years. Others come from a background in the history and philosophy of science where the perspective is different from and complementary to that of the practitioners at the coalface. As editors, we have made no to effort to smooth out the difference in approach and perception. We find the synergies and tensions between the approaches stimulating, revealing, and thought-provoking.

The contents

The first chapter in this ‘Handbook of the History of Modern Cosmology’, written by Helge Kragh, surveys cosmological ideas in the pre-Einsteinian era from approximately 1860 to 1910. During this period cosmology was not recognized as a distinct field of scientific study and yet there was a lively discussion of subjects which were de facto cosmological. Some of these discussions were stimulated by Olbers’ paradox and others by the controversial application of the law of entropy increase to the universe at large. Importantly, astrophysics emerged as a new and powerful transdisciplinary science which substantially changed classical astronomy and eventually also cosmology. During the same period the idea of non-Euclidean geometry aroused great mathematical interest and was even embraced by a few astronomers and physicists. But throughout the half-century cosmological theory remained characteristically speculative and qualitative, largely without a stable foundation in the form of relevant observations. Most
professional astronomers either adopted an agnostic attitude or simply ignored the universe at large as a possible domain of science.

Chapter 2 by Robert W. Smith covers the same period as Chapter 1 but from the perspective of observational astronomy and cosmology and extending the period up to about 1940. With the establishment of state-of-the-art observatories in the US in the early twentieth century astronomers began to address the nebulae, one of the major results being V.M. Slipher’s discovery of redshifts in the spectra of spiral nebulae. The so-called Great Debate in 1920 was primarily concerned with the relationship of spiral nebulae to the Milky Way Galaxy, but the result of the debate was inconclusive. It was only with E. Hubble’s discovery of Cepheids in the Andromeda Nebula that a distance indicator was established and the question resolved. With the new distance scale and more data for nebular redshifts astronomers looked for a redshift–distance relation that would eventually become the main evidence for the expanding universe. To a large extent through the work of Hubble, extragalactic astronomy became a reality in the 1930s and immediately applied to test cosmological models.

As detailed in Chapter 3 by Matteo Realdi, during the 1920s cosmological models based on general relativity theory meant essentially the matter-filled Einstein model or the empty universe of Willem de Sitter. Only in 1930 was it recognized that none of the static models were adequate and that a better choice was the expanding universe which had earlier been proposed by Alexander Friedman and Georges Lemaître. With the new revelation physicists and astronomers such as Otto Heckmann, Howard P. Robertson, and Richard C. Tolman systematized the various models satisfying the requirements of general relativity. Some of the models led to a finite age of the universe smaller than the age of the Earth and much smaller than the age of stars and galaxies. The notorious timescale difficulty continued to haunt relativistic cosmology until the mid-1950s. Nonetheless, by the late 1930s the expanding universe was generally accepted and disseminated to a broader public in the form of popular books and articles. On the other hand, Lemaître’s suggestion of a kind of big-bang model was unfavourably received.

It is tempting but unhistorical to identify cosmological models with solutions to Einstein’s field equations. As discussed in Chapter 4 by Helge Kragh, since the 1920s a series of alternative cosmological theories have been proposed, some of them radically alternative and others mere modifications of relativistic models. Apart from discussing the general issue of what ‘alternative’ means, the chapter focuses on a small number of more or less heterodox theories. Some of them, such as E.A. Milne’s kinematic cosmology, once played an important role but now belongs to history. The same is the case with the steady-state theory which, because of its historical importance, is considered separately (Chapter 5). Other ideas, such as oscillating models and the introduction of time-varying constants of nature, are still discussed if mainly outside mainstream cosmology. The chapter also describes plasma models and tired-light hypotheses, and it briefly relates to more recent sceptical views regarding cosmology’s status as a proper science. Such views, found in the recent literature, have a long historical heritage but are increasingly marginalized by the remarkable developments related in this Handbook.

Still in 1960 it was far from evident that we live in a big-bang universe. Chapter 5, also by Helge Kragh, deals with the extensive and on occasion dramatic cosmological controversy that raged throughout the period 1948–1965 between two radically different
conceptions of the universe. In the late 1940s George Gamow and collaborators revived the idea of a finite-age, exploding universe which they developed into a quantitative model of the early hot universe. At the same time Fred Hoyle, Thomas Gold, and Hermann Bondi proposed the steady-state model resting on assumptions quite different from those of relativistic cosmology. The chapter follows the controversy up to the discovery of the CMB radiation which essentially sounded the death knell of the steady-state theory. And yet, according to a minority of cosmologists the death sentence was premature. They continued to fight the big-bang orthodoxy and suggest modified steady-state models in accordance with observations. But their fight was unsuccessful and, some may say, misguided. The case of the cosmological controversy is not only of historical importance, it also provides a nice illustration of the classical problem of theory choice in science.

Chapter 6 by Malcolm Longair is devoted to observational and astrophysical cosmology in the period from 1940 to 1980 and thus continues the earlier story covered in Chapter 2. The timescale problem which had plagued most relativistic models was solved with Walter Baade’s revision of Hubble’s constant in about 1953, but later in the century a new dispute emerged regarding the value of the constant. Other disputes of mainly an observational nature concerned the value of the deceleration parameter and the possibility of a non-zero cosmological constant. Perhaps the leading observational cosmologist of the period, Allan Sandage, believed that the correct cosmological model could be nailed down with the 200-inch telescope on Mount Palomar. Other issues dealt with in the chapter, apart from the determination of cosmological parameters, are the formation of large-scale structures in the universe and the role played by radio astronomy in the resolution of the cosmological controversy. The chapter also discusses the work that led to a reliable thermal history of the universe based on observations and begins the search for the density fluctuations in the CMB radiation which are further dealt with in other chapters.

The subject of Chapter 7, also by Malcolm Longair, is relativistic astrophysics and the role it has played as a tool in the cosmological arena. Although astrophysics in the relativistic domain began in the 1930s with work on supernovae and neutron stars, it was only three decades later that the new research area attracted massive attention and became of crucial importance in cosmology and in the testing of theories of gravity. The discovery of extragalactic radio sources and quasars entered the controversy over the steady-state theory and it was followed in 1967 by the discovery of pulsars. The new and strange objects provided valuable input for cosmological and gravitational theories. The same was the case with the surprising discovery in 1997 that the high-redshift gamma bursts are at cosmological distances and thus must have enormous energies. The chapter also recounts the story of celestial X-rays and gamma rays which dates back to rocket flights in the early 1960s, and it discusses the attempts to detect gravitational waves which culminated in the recent successful LIGO experiments rewarded with the 2017 Nobel Prize in physics. Through the whole of this period, general relativity changed from a specialised mathematical discipline to an everyday tool of the astrophysicist.

The discovery and further study of the CMB has been of unrivalled importance to our understanding of the early universe and to cosmology generally. Chapter 8 by Bruce
Partridge is devoted to the epic story of the CMB with a focus on the years from the mid-1960s to the present day. Although predicted by Ralph Alpher and Robert Herman as early as 1948, the CMB became a reality only with its serendipitous discovery by Arno Penzias and Robert Wilson in 1964. Initially it was questioned whether or not the CMB was cosmic, but measurements at other wavelengths soon silenced the scepticism. The chapter deals with the advances in instrument technologies that turned the CMB to a phenomenon of quantitative studies and ultimately discovered the tiny anisotropies predicted by theory. The detailed study of the radiation’s power spectrum opened up a Pandora’s box full of new insights and new problems. The CMB became an foundational tool for calculations of nucleosynthesis and led to the necessity of dark matter and dark energy as major constituents of our universe, not to mention hypotheses of an early inflationary phase. This study reveals the enormous effort needed from many physicists, engineers and cosmologists, as well as funding agencies, to determine unambiguously the central role which the CMB occupies in modern cosmology.

The era of big science space programmes coincided with the Cold War. As suggested by Silvia De Bianchi in Chapter 9, even cosmology became involved in the political, military, and technological competition between the USA and the Soviet Union. Rather than focusing on the ideological scene, she tells the story, based on archival sources, of how the two superpowers developed research programmes of both scientific and military significance to test theories of cosmology and gravitation. Radio astronomy and radar imaging techniques were in part developed for the purpose of planetary astronomy and in part for military purposes. Advanced laser technology, of obvious military importance, provided a tool for testing general relativity and determining whether or not the constant of gravitation varied slowly in time. Although the general picture during the 1960s and 1970s was characterized by competition, there were also collaborative space science programmes. Generally the chapter argues that the historiography of cosmology and gravitation in the period needs to take into account Cold War space science and its associated technologies.

Chapter 10 by Malcolm Longair continues the story from Chapter 6, now bringing the development in observational and astrophysical cosmology up to the present. Since about 1980 our empirical knowledge of the universe has progressed tremendously and what was once a dream, precision cosmology, has become a reality. The remarkable development has largely been technology-driven, the result of increased computer power, new generations of telescopes for all wavebands, semiconductor-based detectors (CCDs), and much more. The discipline has also benefitted from the influx of experimental and theoretical physicists into the cosmological arena. The chapter describes how the accuracy and reliability of the cosmological parameters have increased drastically and established the $\Lambda$CDM consensus model. The age of the universe has been determined by different techniques resulting in the same age of about 13.8 billion years. A major advance in this period involved the supernovae projects which in the late 1990s led to the conclusion that the universe is in a state of acceleration. Most of its gravitating mass is non-baryonic and most of its total energy density is due to vacuum energy. The chapter also surveys several other crucial areas of modern cosmology, including light element nucleosynthesis and the formation of galaxies and large-scale structures.
Chapter 11, written by Malcolm Longair and Chris Smeenk, deals with concepts that are key components of modern cosmology but the nature of which are not well understood. While dark non-baryonic matter is generally accepted on observational grounds, there is no sure knowledge of what this kind of matter consists of. The history of dark energy is part of the history of Einstein’s cosmological constant, or perhaps it is the other way around. It may well be that the dark energy is the cosmological constant, but several alternatives are still discussed. With the proposal of the inflationary scenario for the very early universe in the early 1980s, a major revision of the very early phases of the big bang model saw the light of day. Studies of the CMB later resulted in a more sophisticated picture of inflation and its role in structure formation. Although many cosmologists assert that inflation has become part of the present consensus model, critics argue that it is too early to claim with confidence that the universe started in an explosive inflationary phase.

Some of the themes in Chapter 11 are further discussed in Chapter 12 by Milan M. Ćirković if from a different, theoretical and to some extent speculative perspective. The history of the universe supposedly started with the big bang at $t = 0$ but theories of quantum gravity make it possible to extend the history arbitrarily into the past, so-called pre-big-bang theories. And as to the far future, physicists have produced scenarios of the evolution of the universe and its constituent objects zillions of years from now, what is sometimes known as physical eschatology. The notorious anthropic principle entered cosmology in 1974 as a selection effect for observations and an argument for other possible universes. The controversial idea of a multiverse has several roots, including string theory, the many-worlds interpretation of quantum mechanics, and self-reproducing inflationary models. Thus, cosmology in the twenty-first century is not only characterized by precise measurements and an observationally confirmed standard model but also by an increased engagement with other fields of fundamental physics, resulting in fields such as string cosmology and its bedfellows. Once again the development has stimulated discussions of a philosophical nature.

Although cosmology has long ago become a genuine physical science, it is not and cannot be totally divorced from philosophical concerns. The final Chapter 13 by Chris Smeenk takes up some of the pertinent philosophical issues relating to modern cosmology. Theology has a place too, as in the classical and still living discussion of the ultimate origin of the universe. During the controversy over the steady-state theory in the 1950s (Chapter 5) the very nature and aims of cosmology were hotly debated and not all of the questions raised then have received a final answer. Cosmology rests on extrapolations. It is empirically underdetermined in a strong sense as observational knowledge is restricted to just a small part of the universe. More recently the question of the initial state has been reconsidered from the perspective of theories of quantum gravity. Anthropic reasoning and the epistemic status of multiverse hypotheses have attracted as much philosophical as scientific interest. In short, for philosophers, aspects of modern cosmology question traditional modes of scientific inquiry related to fundamental issues such as explanation, prediction, and falsification. And there is more.
A number of the authors of the papers which include images reproduced in this handbook are deceased or uncontactable. We are most grateful to the publishers who have been helpful in giving permission for the use of the figures from journals, books, and other media for which they hold the copyright. Every effort has been made to track down the copyright owners for all the pictures, but a very few of them have proved to be beyond what we consider to be reasonable effort.

Helge Kragh and Malcolm S. Longair
Contents

1 Cosmological theories before and without Einstein
   Helge Kragh

2 Observations and the universe
   Robert W. Smith

3 Relativistic models and the expanding universe
   Matteo Realdi

4 Alternative cosmological theories
   Helge Kragh

5 Steady-state theory and the cosmological controversy
   Helge Kragh

6 Observational and astrophysical cosmology: 1940–1980
   Malcolm S. Longair

7 Relativistic astrophysics and cosmology
   Malcolm S. Longair

8 The cosmic microwave background: from discovery to precision
   cosmology
   R. Bruce Partridge

9 Space science and technological progress: testing theories
   of relativistic gravity and cosmology during the Cold War
   Silvia De Bianchi

10 Observational and astrophysical cosmology: 1980–2018
    Malcolm S. Longair

11 Inflation, dark matter, and dark energy
    Malcolm S. Longair and Chris Smeenk

12 Stranger things: multiverse, string cosmology, physical
    eschatology
    Milan M. Ćirković
13 Philosophical aspects of cosmology 497

Chris Smeenk

References 531
Subject index 595
Author index 602
List of Contributors

Milan M. Ćirković. Research Professor, Astronomical Observatory of Belgrade, Serbia. mcirkovic@aob.rs

Silvia De Bianchi. Ramón y Cajal Fellow, Department of Philosophy and Centre for the History of Science, Autonomous University, Barcelona, Spain. silvia.debianchi@uab.cat

Helge Kragh. Emeritus Professor, history of science, Niels Bohr Institute, University of Copenhagen, Denmark. helge.kragh@nbi.ku.dk

Malcolm S. Longair. Emeritus Professor, natural philosophy, Cavendish Laboratory, Cambridge, UK. msl1000@cam.ac.uk

R. Bruce Partridge. Emeritus Professor, astronomy, Haverford College, PA, USA. bpartrid@haverford.edu

Matteo Realdi. Postdoc, history of astronomy, Institute for History and Social Aspects of Science, Vrije Universiteit Amsterdam, The Netherlands. m.realdi@vu.nl

Chris Smeenk. Associate Professor, history and philosophy of physics, Department of Philosophy, University of Western Ontario, London, Canada. csmeenk2@uwo.ca

Robert W. Smith. Professor, history of science, Department of History and Classics, University of Alberta, Edmonton, Canada. rwsmith@ualberta.ca
1

Cosmological theories before and without Einstein

Helge Kragh

*During the ages there seems to be an oscillating between two different types of
cosmologies: during some periods it was believed that the structure of the universe can
be understood through religious-philosophical-theoretical speculations, during others
it is claimed that an empirical-observational approach is preferable.*

Alfvén (1983), p. 323

1.1 Introduction

It is a common misunderstanding, not least among physicists and astronomers, that
cosmology only became a science, as distinct from philosophy, in the twentieth century.
Many modern scientists will be inclined to point to Einstein’s pioneering theory of 1917
as a precondition for a truly scientific cosmology, perhaps the very beginning of the
field. Others will identify the beginning with the discovery of the cosmic microwave
background in 1965 and others again with the COBE satellite experiments starting in
1989. But attempts to understand the entire universe in rational and scientific terms
go much farther back in time. One can reasonably speak of scientific (or at least proto-
scientific) cosmology in the cases of Aristotle, Hipparchus, Ptolemy, and other thinkers of
Greek antiquity. The later achievements of Copernicus, Kepler, and Newton were more
than just contributions to theoretical planetary astronomy. They were also concerned
with the structure of the universe as a whole and therefore qualify as cosmology.¹

Although with roots that go back to the earliest civilizations, cosmology in essen-
tially the way we understand the term today only emerged in the enlightenment era.
Perhaps characteristically, it is in this period we find the first books carrying the word
“cosmology” in their titles. The first time may have been in 1731 when the influential
German philosopher Christian Wolff published his *Cosmologia Generalis*, a book that was
primarily in the older philosophical tradition.
Twenty-four years later another and more famous philosopher, Immanuel Kant from Königsberg in Germany (the present Kaliningrad), published anonymously a small book entitled Allgemeine Naturgeschichte und Theorie des Himmels (General Natural History and Theory of the Heaven; Fig. 1.1). Although Kant referred frequently to God and presented his theory as theistic, in reality his references to the Creator were largely rhetorical except for the original creation of matter. Contrary to the much-admired Newton he found no place for divine miracles in the universe. “A world-constitution, which without a miracle does not maintain itself does not have the steadiness which is the hallmark of God’s choice,” he piously wrote.² Perhaps the most important innovation in Kant’s cosmology was that he provided the universe with an evolutionary perspective. He started with a primeval, divinely created chaos of particles at rest, distributed throughout an infinite void. The initial chaos was unstable, he said (but did not prove), and out of the chaos condensations were formed by gravitational attraction. He claimed that the result must necessarily be a regular and orderly cosmos consisting of planetary systems and huge nebulae like the Milky Way. Ostensibly on the basis of Newtonian mechanics he

Fig. 1.1 Title page of Kant’s treatise on the constitution and evolution of the world.
pictured the present world as an “island universe,” the islands being nebulae floating in the infinite sea of void space. The term, which became popular since the late nineteenth century, cannot be found in Kant's work. It may have been coined by the American astronomer and evangelist Ormsby Mitchel who used it in 1846 in a journal he edited, *The Sidereal Messenger*.

Another novel aspect of Kant’s grandiose cosmological scenario was its dynamic and evolutionary nature. Of course, God had originally created the universe, but not in its present form for it had slowly developed from the primeval chaos solely governed by the laws of nature. The evolutionary perspective implied that the universe has a history and then an age, a concept that only appeared in modern cosmology in the 1930s. Indeed, not only did it have a beginning, it also might have an end, for Kant speculated that the universe would eventually return to its chaotic state and that this would possibly happen an infinity of times. The idea of an eternally recurrent universe, sometimes known as a “phoenix universe,” can be traced back to the earliest mythological world views and was particularly common in Indian culture. Although Newton considered the possibility, he rejected it as theologically dangerous. With Kant and later thinkers the fascinating scenario of a phoenix universe experienced a revival.

Interesting as Kant’s combined cosmology and cosmogony is from the point of view of history of ideas it did not live up to our standards of science and not even to the standards of his own time. It was still philosophical cosmology although now dressed in scientific language. To speak of the beginning of modern cosmology we need to jump ahead another century, to a period when more was known observationally of the stellar universe and when the laws of thermodynamics had entered alongside Newtonian gravitation as universally valid laws of nature. This chapter focuses on the conceptual and theoretical problems in cosmology ca. 1860–1910 rather than the astronomical observations relevant to the structure of the universe as a whole.

### 1.2 Olbers’ so-called paradox

Heinrich Wilhelm Olbers was trained in medicine and during most of his career he earned his living as a practising physician in Bremen, Germany (Fig. 1.2). However, he considered his true vocation to be astronomy in which area he was recognized as one of the most eminent researchers of the early eighteenth century. His reputation was based in particular on his pioneering observations of comets, asteroids, and meteoric showers. In 1823 Olbers submitted a theoretical paper to the *Astronomisches Jahrbuch* (Astronomical Yearbook) which contained his most well-known contribution to astronomy but which only appeared in print three years later. Olbers introduced his paper with lengthy quotations from Kant’s almost forgotten book of 1755 and in this way he contributed to a revival of interest in the cosmological vision of the Königsberg philosopher. Indeed, like Kant he firmly believed that the universe was infinite and throughout populated with stars distributed approximately uniformly. A devout Christian, Olbers argued that only an infinite universe agreed with God’s omnipotence. But then he was faced with
Fig. 1.2 The German astronomer H.W. Olbers (1758–1840). Wikimedia Commons.

an old puzzle concerning the darkness of the night sky which first had been noticed two centuries earlier.

The puzzle became eventually known as “Olbers’ paradox” although Olbers did not consider it to be paradoxical at all. The eponymous label may first have been used by Hermann Bondi, who in his textbook Cosmology from 1952 dealt with the subject in some detail and of course with the benefit of hindsight.\(^4\) According to Bondi, the riddle of cosmic darkness could be explained as a result of the redshift of distant stars and galaxies caused by the expansion of the universe.

The so-called paradox arises from the assumption that the infinite or just hugely large universe is uniformly filled with stars that shine in the same way. The light received from a star varies with the inverse square of the distance and so is negligible for distant stars; but the number of stars at a certain distance increases with the square of the distance with the result that the integrated starlight received on Earth will make the sky at night as bright as or brighter than on a sunny day. This is the brief version of Olbers’ famous paradox. More formally, consider the Earth to be surrounded by spherical shells of radius \(r\) and thickness \(dr\); assume that the distribution of stars, each of the same luminosity \(L\), is uniform and unchanging in time. If \(\rho\) denotes the average stellar density the number of stars in a shell is \(4\pi\rho r^2\, dr\) and the total apparent luminosity at the Earth is
For $R$ approaching infinity the expression becomes infinitely large.

The history of Olbers’ paradox has been traced back to Johannes Kepler who in a work of 1610, *Dissertatio cum Nuncio Sidereo* (Conversation with the Starry Messenger), used it as an argument for the finite and surprisingly small universe that he favoured. Edmond Halley, on the other hand, was convinced that the stellar universe was infinite and in a paper of 1720 he claimed to have defused the paradox by means of an argument of a somewhat obscure nature; it rested on the claim that the intervals between the stars in an infinite system decreased linearly with distance while the intensity of light from them decreased with the double square of the distance.

Olbers was aware of Halley’s much earlier paper in *Philosophical Transactions* but found its explanation to be obscure and unsatisfactory. As an alternative he assumed that the intensity of light suffered a slight absorption during its journey through interstellar space. Referring to the tails of comets, the zodiacal light, and other astronomical phenomena, he concluded that “Without doubt the universe is not absolutely transparent.” According to Olbers, “Only the slightest degree of non-transparency suffices to refute the conclusion—so contrary to our experience—that if the fixed stars stretch away to unlimited distance, the entire sky must blaze with light.” To Olbers and many of his contemporaries the proposed explanation was yet another proof of God’s infinite wisdom. In the words of Olbers: “The Almighty with benevolent wisdom has created a universe of great yet not quite perfect transparency and has thereby restricted the range of vision to a limited part of infinite space.”

From a priority point of view Olbers’ paradox should perhaps be called Chéseaux’s paradox, a reference to the Swiss astronomer Jean-Philippe Loys de Chéseaux who in an essay of 1744 analyzed more lucidly and in greater detail than Olbers the paradox of the night sky. Following Halley in conceiving the stars to be situated in concentric spherical shells, he stated that the quantity of light emitted from each shell was proportional to the sum of the squares of the apparent diameters of the stars in the shell. He did not consider the possibility that stars may fall on the same line drawn from the Earth and thus shield for some of the incoming light. Chéseaux concluded that the paradox would arise not only if the number of stars was infinite, but also if it was exceedingly large but still finite, namely larger than a value corresponding to $76 \times 10^{13}$ concentric shells.

Rather than opting for a finite, relatively small universe he suggested that the intensity of starlight decreased at a greater rate than given by the squared-distance law and that the physical cause was absorption in an interstellar ethereal fluid. Chéseaux argued that the problem would disappear if the intensity of light were to diminish by 3 percent when passing through a layer of thickness equal to the diameter of the solar system. For the nearest stars he estimated a distance 120,000 times greater, implying that we observe only the foreground stars and not the more distant ones. Olbers essentially duplicated the work of the Swiss astronomer of which he was apparently unaware. Olbers and Chéseaux had
in common that neither of them questioned the assumption of more or less uniformly distributed stars. Another assumption they implicitly agreed upon was that the stars are not subject to systematic motion in time.

Priority aside, the Chéseaux–Olbers hypothesis of a light-absorbing medium was widely accepted in the nineteenth century. It was not without problems, though. In 1848 the famous British astronomer John Herschel pointed out that radiant energy from the stars, when absorbed in the hypothetical interstellar medium, would heat up the medium until it reached a state of thermal equilibrium and the medium itself became radiant. Herschel realized that the infinite universe could be reconciled with the dark night sky even without assuming light absorption, namely if the stars were arranged in a suitable, non-uniform way. Nothing is easier, he wrote, “than to imagine modes of systematic arrangement of the stars in space . . . which shall strike away the only foundation on which [the problem] can be made to rest, while yet fully vindicating the absolute infinity of space.”6 The idea of a specially arranged hierarchic stellar universe was another solution to Olbers’ paradox, but one for which there was as little observational evidence as there was for the absorption hypothesis.

And there were more solutions, including one that did not assume interstellar absorption, a hierarchic universe, or a finite number of stars. In 1858 the German astronomer Johann Mädler, at the time director of the Dorpat Observatory in present Estonia, came up with a novel solution in his book Der Fixsternhimmel (The Heaven of the Fixed Stars). He argued that Olbers’ paradox would not arise if we only receive light from stars within a certain horizon. This might happen not as a result of absorption but as a result of the universe having a finite age:

The world is created, and hence is not eternal. Thus no motion in the universe can have lasted for infinite time; in particular, this applies to a beam of light. In the finite amount of time it could travel before it reached our eye, a light beam could pass though only a finite space no matter how large the speed of light. If we knew the moment of creation, we would be able to calculate its boundary.7

Mädler’s proposal failed to attract attention among his colleagues in astronomy except that it was restated by the Leiden astronomer Frederik Kaiser in a popular book of 1860.

More importantly, in 1901 the great physicist William Thomson (who by then was Baron Kelvin of Largs) independently examined how to “test an old and celebrated hypothesis that if we could see far enough into space the whole sky would be seen occupied with discs of stars . . . and that the reason why the whole of the night-sky and day-sky is not as bright as the sun’s disc, is that light suffers absorption in travelling through space.”8 Thomson considered a stellar Milky Way universe uniformly filled with stars up to the enormous distance of $3.3 \times 10^{14}$ light years and took into consideration that each star has a finite lifetime. Based on what he claimed was the “irrefragable dynamics” underlying his own and Hermann von Helmholtz’s contraction theory of solar energy emission, Thomson assured that a star could shine for at most 100 million years.9 This implied that the time it would take for light to travel from one of the stars farthest away would be about 3.3 million times the lifetime of a star. “Hence,” he wrote, “if all the
stars through our vast sphere commenced shining at the same time...at no one instant would light be reaching the earth from more than an excessively small proportion of the stars.” This, a much more sophisticated version of Mädler’s argument, was essentially Thomson’s solution to Olbers’ paradox. But Thomson did not refer to either Olbers or Chéseaux and was undoubtedly unaware of Mädler’s suggestion.

Olbers’ paradox was not taken very seriously during the second half of the nineteenth century, in part because it was generally assumed that interstellar space was filled with a rarefied medium that absorbed part of the starlight. The medium was sometimes identified with the physicists’ world ether, the existence of which was generally assumed. Only during the early part of the new century did observational evidence indicate that space was much more transparent than previously assumed and the problem therefore real. In 1917 Harlow Shapley formulated the paradox as support of a finite stellar universe: “Either the extent of the star-populated space is finite or ‘the heavens would be a blazing glory of light’...Then, since the heavens are not a blazing glory, and since space absorption is of little moment throughout the distance concerned in our galactic system, it follows that the defined stellar system is finite.”

Models of the Milky Way proposed at the time by J.C. Kapteyn, H. von Seeliger, and K. Schwarzschild all assumed that the absorption of starlight by interstellar matter was negligibly small.

Only in the late 1920s did astronomers begin to reconsider the reality of absorption, which was proved observationally by the Swiss-American astronomer Robert Trumpler at the Lick Observatory. As Trumpler demonstrated in 1930, the interstellar absorption effect amounted to an average change in apparent magnitude of 0.67 per kiloparsec. However, in his meticulous investigation of open star clusters Trumpler did not refer specifically to Olbers’ paradox and he did not relate his work to the cosmological problem. Incidentally, a value roughly similar to the one obtained by Trumpler followed from observational estimates reported much earlier by Friedrich Struve, the first director of the Pulkovo Observatory, in a paper of 1847. Convinced of the absorption of starlight by interstellar matter Struve praised Olbers’ earlier solution to the problem of the dark night sky.

To summarize, there were in the period basically five proposals of how to avoid Olbers’ paradox. Apart from interstellar absorption of light one could assume a hierarchic structure of the stellar universe or that there existed only a finite number of luminous bodies in space; or one could assume the stellar universe to be of finite age. The last possibility, of which more below, was to assume that space itself was finite (but unbounded) and filled uniformly with stars.

### 1.3 The rise of astrophysics

Although one can meet the term “physical astronomy” in the scientific literature of the enlightenment era, astrophysics is a child of the nineteenth century. In 1866 Friedrich Zöllner was appointed professor of astrophysics at the University of Leipzig, the first academic position of its kind. Thirty years later astrophysics was flourishing and about to gain status as a scientific discipline, such as indicated by *The Astrophysical Journal*
founded in 1895 by George E. Hale and James E. Keeler. The new interdisciplinary branch of science significantly changed the course not only of astronomy but also of physics generally. Moreover, astrophysics turned out to have very important consequences for cosmology, although these were only fully recognized much later. The emergence of the field was closely related to the introduction of spectroscopy, which from its very beginning was applied to the study of the stars.

For a couple of decades astrophysics and astrospectroscopy were nearly synonymous terms. And then we should not forget that spectroscopy was as much related to chemistry as to physics. Astrochemistry was another product of the spectroscopic revolution and no less important than astrophysics. In 1887 the prominent British amateur astronomer J. Norman Lockyer—also known as the founder of the journal *Nature*—published a voluminous monograph significantly titled *The Chemistry of the Sun*.

Advancements in the design of spectrosopes and other optical instruments proved for the first time that analysis of light received from the stars could give valuable information about the stars’ physical and chemical composition. That such information was possible at all, was a novelty. The very notion of a “star” changed from a geometrical to a physical concept, a change which implied a profound transformation of the astronomical sciences.

Still to Friedrich Bessel and his contemporaries the business of astronomy was restricted to precise measurements of the positions and orbits of celestial bodies whether planets, comets or stars. In a letter of 1832 to the great naturalist Alexander von Humboldt, Bessel wrote “Everything else that one may learn about the [heavenly] objects, for example their appearance and the constitution of their surfaces, is not unworthy of attention, but is not the proper concern of astronomy.” Only very few scientists imagined that experimental physics and chemistry would be of astronomical importance, such as Hans Christian Ørsted, the discoverer of electromagnetism, prophesied in a lecture of 1807. “Some day chemistry will have just as much influence on astronomy as mechanics so far,” he said. Thanks to advances in astrophysics, what Bessel called the “proper concern of astronomy” became much broader, incorporating important parts of the physical and chemical sciences.

Disregarding still earlier developments, astrospectroscopy began in 1814 with the Bavarian instrument maker Joseph Fraunhofer’s systematic investigation of the solar spectrum. Fraunhofer found in the spectrum a large number of dark lines which he carefully measured but without being able to explain them. What was the cause of the mysterious lines? It took more than three decades until the Fraunhofer lines attracted wide attention and it was gradually realized that the bright emission lines of chemical elements coincided with the dark lines. The real breakthrough in spectrum analysis occurred in 1859 as a result of a unique collaboration between the physicist Gustav Robert Kirchhoff and the chemist Robert Wilhelm Bunsen, both at the University of Heidelberg. In experiments using Bunsen’s new gas burner the two scientists demonstrated that the emission spectra could be used to identify small amounts of known chemical elements and also to predict new ones. Within a year the chemical power of the spectroscope was dramatically demonstrated with the discovery of two new metallic elements, cesium and rubidium. Kirchhoff immediately pointed out that the new technique had important astronomical consequences. In a paper of 1860 he
demonstrated that sodium was a constituent of the solar atmosphere and soon thereafter he also identified the presence of iron, magnesium, copper, zinc, nickel, and barium.15

Not only did the work of Bunsen and Kirchhoff establish stellar astrophysics and astrochemistry as promising fields of research, it also led Kirchhoff to a theoretical study of the thermodynamics of radiant heat and what he called blackbody radiation. This fundamental work became the starting point of a process that in 1900 culminated with Max Planck’s celebrated quantum hypothesis. By the 1920s quantum theory and the new atomic physics based on it had become a sine qua non for progress in astrophysics. Quantitative astrophysics based on quantum theory took its start in the early 1920s with important contributions of Meghnad Saha, Ralph Fowler, E. Arthur Milne, and others. Saha’s pioneering work on the temperature and pressure of stellar atmospheres, building on a combination of quantum theory, statistical mechanics and chemical equilibrium theory, was particularly important. As Eddington generously wrote, “Saha’s theory has dominated all recent progress in the observation and interpretation of stellar spectra.”16

But even earlier the new quantum and atomic physics had on occasions been applied to problems of astronomy and astrophysics.17 Thus, in Niels Bohr’s epoch-making theory of atomic structure of 1913 he famously explained some mysterious spectral lines from the star ζ Puppis which astronomers attributed to a new hydrogen series but according to Bohr were due to ionized helium in the form He⁺. It is less known that Bohr also predicted what came to be known as “Rydberg atoms” and on this basis explained why astronomers had observed many more Balmer lines than physicists had found in the laboratory. Bohr’s prediction of large and highly excited atoms in areas of space at very low density was verified in 1965 when such atoms were detected by means of radio-astronomical methods in interstellar gas clouds.18

Fig. 1.3 Early spectroscope used by W. Huggins for astronomical studies.
The versatile spectroscope was used for a variety of purposes, some chemical, others physical, and others again astronomical (Fig. 1.3). In 1842 the Austrian physicist and mathematician Johann Christian Doppler read a paper to the Royal Bohemian Society of Sciences in which he announced the principle named after him. As indicated by the title of his paper, “Über das farbige Licht der Doppelsterne” (On the colored light of double stars), he originally thought of the principle in an astronomical context. However, he wrongly thought that the color of a star would undergo a perceptible change because of its motion relative to the Earth. He even suggested that for large stellar velocities the light would be shifted to such an extent that the star would become invisible, a line of reasoning that he used to explain the color of double stars. What became known as the Doppler effect was first verified for sound waves three years later, and in 1848 the French physicist Armand Hippolyte Louis Fizeau, who was at the time unaware of Doppler’s treatise, interpreted it as a shift in wavelength for spectral lines. The principle thus originated in astronomy and after been proved in acoustics it returned to astronomy. History apart, according to the Doppler or Doppler–Fizeau principle, if a light source moves relative to the observer with radial velocity \( v \) there will be a change in wavelength given by

\[
z \equiv \frac{\Delta \lambda}{\lambda} = \frac{\lambda' - \lambda}{\lambda} = \frac{v}{c},
\]

(1.2)

where \( c \) is the velocity of light and \( \lambda' \) the measured wavelength, redshifted relative to the emitted wavelength \( \lambda \). As Einstein proved in his seminal paper of 1905 in which he introduced the special theory of relativity, if the recession velocity \( v \) is very large the formula becomes

\[
z = \frac{\sqrt{(1 + v/c)}}{(1 - v/c)} - 1,
\]

(1.3)

In the limit \( v \ll c \) Einstein’s expression reduces to Doppler’s.

The first successful attempt to observe a stellar redshift was made by the wealthy English gentleman astronomer William Huggins in 1868, in stiff competition with another pioneer of astrospectroscopy, the Italian Jesuit astronomer Angelo Secchi. Comparing the H\(_{\beta}\) line in the spectrum of the bright star Sirius with that produced in the laboratory by a discharge tube filled with hydrogen, Huggins found a shift in wavelength of about one angstrom. The displacement was at the limit of observability and Huggins had to convince himself that it was real and not due to some artefact caused by the instrument. On the assumption that the shift was a Doppler effect he cautiously concluded that it implied a recession velocity of Sirius of around \(+40\) km\(s^{-1}\). To many of his contemporaries the reported value was inconceivably high and contradicting common sense, not to mention the consensus view of a static stellar universe. Nonetheless, Huggins’ measurements were accepted as the first proof of stellar Doppler shifts. Ironically, it later turned out that the result was wrong both in amount and sign (the currently accepted value is about \(-8\) km\(s^{-1}\)).
The validity of the optical Doppler effect was for a period controversial and firmly demonstrated only in the 1880s by examinations of the rotation of the Sun. It took another twenty years until the Doppler effect for light was detected in the laboratory. Attempts to do so were made by the Russian astronomer Aristarch Belopolski at the end of the century but without success. The first unequivocal laboratory demonstration of the effect was made by the German physicist and future Nobel Prize laureate Johannes Stark, who in 1905 proved it for so-called canal rays or positive rays (which are rays of positive ions). By that time measurements of Doppler shifts had become an important tool in stellar astronomy but their significance for cosmology was not yet recognized. Nor did Vesto Melvin Slipher’s discovery of the first nebular Doppler shift in 1912 immediately lead to such recognition, but two decades later the insight of the Austrian physicist had become an indispensable tool for the new cosmology theoretically based on the general theory of relativity.

Astrospectroscopy also became crucially important in relation to one of the most discussed issues of nineteenth-century astronomy, the so-called nebular hypothesis with roots in Kant’s cosmogony of 1755 and Pierre–Simon Laplace’s 1796 theory of the formation of the planetary system. According to the nebular hypothesis as understood by William Herschel and others in the early part of the nineteenth century, some of the nebulae were composed of hot gaseous clouds that would undergo different stages of condensation and eventually end up as stars. The hypothesis was widely associated with the fashionable but also controversial view of nature being in a state of continual evolution. But if all nebulae could be resolved into discrete stars nothing would be left of the nebular hypothesis, and by the mid-nineteenth century this seemed to be the result of observations performed with the most powerful telescopes of the period. The apparently refuted nebular hypothesis was dramatically revived by astrospectroscopic observations made by Huggins in the 1860s.

Huggins knew from Kirchhoff’s work that emission line spectra were produced only by gaseous bodies, whereas hot solids yielded a continuous spectrum. Observations of planetary nebulae in 1864 convinced him that they were composed of glowing gas, and the following year he turned to the Orion Nebulae which he considered a crucial test of the nebular hypothesis. His spectroscope revealed no continuous spectrum but only three bright emission lines. Huggins later recalled:

The riddle of the nebulae was resolved. The answer, which had come to us in light itself, read: Not an aggregation of stars, but a luminous gas…. There remained no room for doubt that the nebulae, which our telescopes reveal to us, are the early stages of long processions of cosmical events, which correspond broadly to those required by the nebular hypothesis in one or other of its forms.

By the end of the nineteenth century, the nebular hypothesis was widely accepted if still being criticized by a minority of astronomers.

As mentioned, chemistry was no less part of the spectroscopic revolution than were physics and astronomy. During the late Victorian era it was generally assumed that terrestrial elements were all over in the universe but also that there possibly existed
chemical elements in the stars that are not found on the Earth. Moreover, some researchers such as William Crookes and Norman Lockyer in England believed that stellar spectra provided evidence for the complexity of the chemical atom. Perhaps, they speculated, our present elements are the products of evolutionary processes going on in the hot stars. “Let us start at the moment when the first elements came into existence,” Crookes said in a lecture of 1886. He continued:

Before that time matter, as we know it, was not. It is equally impossible to conceive of matter without energy, as of energy without matter; from one point of view the two are convertible terms. Before the birth of atoms all those forms of energy which become evident when matter acts upon matter could not have existed—they were locked up in the protyle as latent potentialities only. Coincident with the creation of atoms all those attributes and properties which form the means of discriminating one chemical element from another start into existence fully endowed with energy.22

Does Crookes’ speculation count as an anticipation of the big-bang scenario?

Although most spectral lines from the stars could be identified with lines known from laboratory experiments, some unidentified lines might indicate the presence of elements, or states of elements, particular to the stars and not known on Earth. On the basis of stellar, solar and cometary spectra, scientists claimed the existence of several spurious elements, among them “coronium”, “nebulium,” and “asterium.” The hypothesis of the first element was based on a line in the coronal spectrum first observed by Charles Young in 1869. Coronium was sometimes believed to be an element lighter than hydrogen and in a few cases placed in the periodic system. The mystery of the spectral line was only solved in 1939, when Walter Grotrian and Bengt Edlén independently concluded that the line was due to highly ionized iron (Fe$^{13+}$). The existence of nebulium was first proposed in 1864 and it took until 1927 before Ira Bowen showed that the pair of nebular lines could be explained as “forbidden transitions” in doubly ionized oxygen (O$^{2+}$).23

Most of the claims of celestial elements were wrong, but not all were. In 1868 Lockyer studied spectral lines from a solar prominence noticing a yellow line with a wavelength of 5876 angstroms. The new line, which he named D$_3$ (because it was close to sodium’s D doublet), did not correspond to any line from a known element (Fig. 1.4). Contrary to what is usually stated, Lockyer did not originally claim that he had detected a new element that might exist in the Sun only and nor did he suggest the name “helium” for it.24 That only came some years later. For more than two decades helium remained a ghost element that the majority of chemists refused to accept, but in 1895 its status changed abruptly when the chemist William Ramsay identified the gas in cleveite, a uranium mineral. Ramsay, and not Lockyer, discovered the element.

Helium was originally believed to be exceedingly rare, and since it was as chemically inert as the more abundant argon it was considered little more than a curiosity. In the early part of the twentieth century nobody could foresee helium’s central role in the universe, just as nobody could foresee the equally central role that the Doppler redshifts would come to play in cosmology. The cosmological significance of helium was dimly recognized in the 1930s, but became of crucial importance only in the 1960s when the
study of helium isotopes in the universe became an important part of the development that led to and consolidated the new hot big-bang theory.  

Speculations concerning the composite nature of atoms and their evolution from a common primeval substance were *en vogue* at the turn of the nineteenth century and during the first decade of the new century. As one of many examples consider the American geologist and astronomer Thomas C. Chamberlin, professor at the University of Chicago and a leading critic of the nebular worldview. In a paper of 1899 Chamberlin pointed out that gravity was not the only force of astronomical significance and that the fate of the stars, and of the entire universe, might depend on the still unknown structure of the atom. “It is not improbable that they [atoms] are complex organizations and the seats of enormous energies,” he said, elaborating:

Certainly, no competent chemist would affirm either that the atoms are really elementary or that there may not be locked up in them energies of the first order of magnitude…. Nor would he probably be prepared to affirm or deny that the extraordinary conditions which reside at the center of the sun may not set free a portion of this energy…. Why should not atoms, atomecules [sic], and whatever lies below, one after another have their energies squeezed out of them; and the outer regions be heated and lighted for an unknowable period at their expense?  

As a later example consider the British physicist John Nicholson, a rival to Bohr in the art of atom-building and an early advocate of the nuclear atom. Foreshadowing much later developments in the physics of elementary particles, Nicholson suggested that, in the future, fundamental physics would have to rely on astrophysics. “Astronomy in the wider interpretation of its scope which is now general, owes much to Physics” he pointed out. “A point appears to have been reached in its [astronomy’s] development at which it becomes capable of repaying some of this debt, and of placing Physical Science in its turn
under an obligation.” Nicholson believed that astrophysics might become “an arbiter of the destinies of ultimate physical theories” because the astrophysicist studied nature in a more primordial form than the laboratory physicist examining terrestrial matter and the laws governing them.

Fin-de-siècle speculations of a more or less cosmological nature spanned a wide range. The British physicist Arthur Schuster suggested in 1898 that there might exist in the universe a hitherto unknown form of matter with the remarkable property that it would be repelled gravitationally by ordinary matter. He coined the word “anti-matter” for the hypothetical substance. “Worlds may have formed of this stuff, with elements and compounds possessing identical properties with our own, undistinguishable in fact from them until they are brought into each other’s vicinity.” Schuster further speculated that atoms and anti-atoms might enter into chemical combinations, with the short-range attractive forces dominating over the long-range gravitational repulsion. “Large tracts of space might thus be filled unknown to us with a substance in which gravity is practically non-existent, until by some accidental cause... unstable equilibrium is established, the matter collecting on one side, the anti-matter on the other until two worlds are formed separating from each other, never to unite again.” The hypothetical stuff made up of matter and anti-matter, he went on, “we cannot call a substance because it possesses none of the attributes which characterize matter ready to be called into life by the creative spark? Was this the beginning of the world?” Schuster was well aware that his cosmological scenario was nothing but an innocent speculation. About seventy years later a different sort of antimatter, the real antimatter based on quantum mechanics, turned out to be of great interest to cosmologists (see Section 4.6).

1.4 Thermodynamics and the universe

The term “thermodynamics” only entered the scientific language through a manual of steam engines published by the Scottish engineer and physicist William Rankine in 1859. Rankine adapted it from the phrase “thermo-dynamic engine” coined by William Thomson four years earlier. The new general theory of heat that emerged in the mid nineteenth century was claimed to be universally valid, and for this reason it was bound to have important cosmological implications. The two laws on which the theory was founded were supposed to work not only for steam engines and chemical processes but also for the solar system and beyond, indeed for the universe as a whole.

From the law of energy conservation it followed that although the energy of the universe might change in form, overall it would remain constant. The second law with its claim of a common direction for all natural processes had consequences for both extremes of the cosmic timescale. If extrapolated to the far future, it indicated that the world would come to an end; and if extrapolated to the far past, it might lead to the conclusion that the world had not always existed but had a beginning in time. Neither of these predictions could be tested (nor can they today), but this only made them more popular as subjects of discussion far outside the small communities of professional scientists.29
In an important paper of 1850, Rudolf Clausius introduced the second law of thermodynamics as a natural and inevitable tendency for heat to equalize temperature, and four years later he reformulated his theory by basing it on a function that in 1865 reappeared under the new name “entropy.” For two states A and B at absolute temperature $T$ Clausius defined the difference in entropy as

$$S_A - S_B = \int_A^B \frac{\delta q}{T},$$  \hspace{1cm} (1.4)

where $\delta q$ denotes an infinitesimal change in heat. Armed with his entropy concept, Clausius now stated the second law as “the entropy of the world tends towards a maximum.” He similarly expressed the first law globally, as “the energy of the world is constant.” The cosmological connection was cultivated more fully in William Thomson’s alternative route to the second law. In a paper of 1852 Thomson derived as a consequence of the law that there would come a time when the Earth became “unfit for habitation of man as at present constituted.” While Thomson suggested a terrestrial “heat death” in 1852, Helmholtz was the first to extend the idea to the entire universe, such as he did in a famous lecture given in Königsberg in 1854. According to Helmholtz’s gloomy prediction, the universe would necessarily approach a state of equilibrium and, when this state had been reached, it would be condemned to eternal rest.

At the Liverpool meeting of the British Association for the Advancement of Science in 1854 Thomson invited his audience to trace backwards in time the actions of the laws of physics. He speculated that the source of mechanical energy in the universe might be sought in “some finite epoch [with] a state of matter derivable from no antecedent by natural laws.” However, such an origin of matter and motion, mechanically unexplainable and different from any known process, contradicted his sense of both causality and uniformitarianism. “Although we can conceive of such a state of matter,” he wrote, “yet we have no indications whatever of natural instances of it, and in the present state of science we may look for mechanical antecedents to every natural state of matter which we either know or can conceive at any past epoch however remote.”30 Here we have the second law used not to predict the far future but to speculate about a singular and primordial state in the distant past.

From about 1860 the heat death scenario was expounded by several leading physicists and significantly entered, either explicitly or implicitly, the scientific literature. Clausius, who coined the term “heat death” (Wärmetod), formulated it in terms of the continual increase of entropy. Imagine, he said, a time when the entropy of the universe had attained its maximum value. Then “no further change could evermore take place, and the universe would be in a state of unchanging death.”31 Here we have the idea of “physical eschatology,” a speculation that would return in a different disguise in late twentieth-century cosmology (see Chapter 12). Clausius further emphasized that the second law, whether formulated in terms of entropy or dissipation, contradicted any idea of a cyclic universe. Not only for Clausius, but even more so for Thomson and his circle of Christian scientists, was it an appealing feature of the second law that it countered what they considered the materialistic and un-Christian notion of a world evolving cyclically.
Expectedly, the claim of a future heat death did not go uncontested. Many scientists and even more non-scientists felt it unbearable that life in the universe should one day cease to exist, never to return. They came up with a variety of suggestions to avoid the scenario, either by devising counter-entropic processes or by questioning the premises upon which the heat death prediction rested. Some of the critics flatly denied the validity of the second law; not a few failed to understand it.

Among scientists a common objection was to question that the second law was valid for the universe as a whole. After all, how can such a huge extrapolation possibly be justified? Reputed physicists such as Pierre Duhem in France and Ernst Mach in Austria did not believe in the heat death prediction. As Duhem argued in an important book of 1905, *The Aim and Structure of Physical Theory*, it did not follow from the entropy law that there is any lower or upper limit of the entropy of the universe. It only follows that the entropy increases endlessly. The entropy could increase asymptotically for an infinite number of years, following for example a logistic curve as given by

$$S(t) = \frac{S^*}{1 + e^{-kt}}. \quad (1.5)$$

The aim of Duhem, an orthodox Catholic, was not to argue against a universe of finite age but to warn that physical theory does not justify such long-term extrapolations. “It is absurd,” he wrote, “to question this theory [thermodynamics] for information concerning events which might have happened in the extremely remote past, and absurd to demand of it predictions of events a very long way off.” More generally Duhem argued that physics and cosmology were radically different sciences and that the former could not answer questions belonging to the latter:

The methods by which the physicist develops his theories are without force when it comes to proving that a certain proposition of cosmology is true or false; the propositions of cosmology, on the one hand, and the theorems of theoretic physics, on the other hand, are judgments… [which] can neither agree with nor contradict one another… [And yet] it is very plain that a cosmological system cannot be reasonably constituted without any knowledge of physics.32

Much of the cosmological literature in the second half of the nineteenth century was of a philosophical and speculative nature. It was often associated with the rise of materialism and positivism as alternatives to the traditional way of conceiving nature, culture, and society. In this kind of literature astronomical observations played no substantial role compared to speculations of what the world ought to look like to be in harmony with the author’s philosophical and political desires. Although it was not cosmology in our present scientific sense, at the time it was an important and widely read genre of literature.

To mention but one example, a German author by the name Hermann Sonnenschmidt published in 1880 a book titled *Kosmologie* which was largely based on the popular Kant–Laplace nebular world view.33 In Sonnenschmidt’s universe, time had always existed and would remain for ever, and space was infinite, continuously filled with matter
or ether. Although he accepted that the entropy law led to the decay of individual parts of the universe, such as stars and nebulae, he denied the cosmic heat death and argued for a cyclic universe. Sonnenschmidt explicitly denied that the world had come into existence by a divine act, a view he considered to be plainly unscientific and nothing but superstition. In German-speaking Europe literature of the same genre as Sonnenschmidt’s was no less popular than the more scientifically informed books and articles written by the astronomers. The common denominator in much of the literature was the claim of an eternally cyclic universe.

According to Ludwig Boltzmann’s statistical theory of thermodynamics, dating from the 1870s, there was a non-zero probability that the entropy of a closed system would decrease rather than increase. And yet he believed that all attempts to save the universe from the heat death were futile because the probability was ridiculously small. Indeed, for a gas volume of 1 cm$^3$ he calculated that it would take $10^{109}$ years until the molecules returned to their initial state! However, in 1895 he came up with a new, speculative answer to the old problem of why we have not already suffered the heat death. His answer involved a remarkable scenario of anti-entropic pockets in a universe which as a whole is in thermal equilibrium:

If we assume the universe great enough we can… also make the probability great that, though the whole universe is in thermal equilibrium, our world is in its present state. But can we imagine, on the other side, how small a part of the whole universe this world is? Assuming the universe great enough, the probability that such a small part of it as our present world be in its present state, is no longer small. If this assumption were correct, our world would return more and more to thermal equilibrium; but because the whole universe is so great, it might be probable that at some future time some other world might deviate as far from thermal equilibrium as our world does at present.$^{34}$

In this passage Boltzmann assumed implicitly that the universe as a whole had existed eternally or for an exceedingly long time. He did not consider the possibility of a beginning. In general he suggested a many-worlds picture, although the worlds were apparently taken to be just different parts of the universe and not, as in some later many-worlds or multiverse theories, causally separated regions. In the second volume of his influential textbook on gas theory, published in 1898, Boltzmann included a section in which he repeated and amplified his ideas of many worlds, entropy fluctuations and time reversal.$^{35}$ He was of course fully aware that these cosmological considerations were highly speculative, but found them to be consistent and also useful because they opened up new perspectives on the range of fundamental physics.

Less discussed than the heat death, but of no less cosmological importance, the second law of thermodynamics was also taken to imply that the universe had a finite age. What has been called the “entropic creation argument” is simple, if not necessarily convincing. The first premise is that according to the fundamental second law the entropy of the world increases continually towards an equilibrium state. The second premise is the simple observation that the present world is far from this state. Order, structure, and life do exist, proving that we do not live in a high-entropy world. But, and this
is the conclusion, had the world existed in an eternity of past time then, given the first premise, we should not live in the world we experience—we should not live at all. The discussion that raged from about 1870 to 1910 concerning thermodynamics and cosmology involved a mixture of science, philosophy and religious belief, categories that in this case cannot be easily separated.

For some commentators in the Victorian era the argument implied that the world was created supernaturally in accordance with the Bible. After all, if the universe had an absolute beginning, what else could have caused it except a creative and transcendent being, in other words God? It is no coincidence that the entropic argument attracted much positive attention among theologians and Christian scientists; nor was it a coincidence that it was seriously criticized by atheists scientists and philosophers. Maxwell and Thomson were among those who associated the second law with divine creation, and yet they and most other scientists refrained from speculating about the origin of everything. The general attitude, as expressed by the British astronomer Edward Walter Maunder, was that the scientist “cannot go back to the absolute beginning of things, or forward to the absolute ends of things…. Not merely because these beginnings of things were of necessity outside his experience, but also because beginnings, as such, must lie outside the law by which he reasons.”36 While several prominent physicists participated in the thermodynamic controversy it is remarkable that the astronomers kept a low profile. Very few professional astronomers showed an interest in cosmological questions. Even fewer referred to the implications of thermodynamics, a branch of science which somehow failed to attract astronomical interest. Charles Augustus Young, a respected Princeton astronomer and author of astronomical texts, was one of the few exceptions. In a widely read textbook of 1893, *General Astronomy*, he included a section on the cosmological consequences of the second law. He had this to say:

If we carry our imagination backwards we reach at last a “beginning of things,” which has no intelligible antecedent: if forwards, an end of things in stagnation. That by some process or other this end of things will result in “new heavens and a new earth” we can hardly doubt, but science has as yet no word of explanation.37

If entropy was a problematic concept to use in arguing for a finite-age universe, after about 1900 there was the possibility of replacing it with another cosmic clock, the newly discovered radioactivity. Although radioactivity was studied as a terrestrial phenomenon, it was generally agreed—incorrectly it turned out—that the material composition of the Earth was largely representative of the cosmos. How could there still be radioactive elements if the world had existed in an eternity? Even uranium and other elements with a very long half-life would have disappeared. In 1911 the Austrian physicist Arthur Haas addressed the question, his answer being that the universe must have had a beginning in time. To avoid the conclusion one might object that uranium and thorium were perhaps decay products of some heavier, hypothetical elements, but Haas showed that the objection would only lead to new difficulties. His suggestion of linking together radioactivity and cosmology attracted but little attention at the time but it played some role in later cosmological thought. Notably, when Georges Lemaître proposed the first
cosmological big bang model ever in 1931, he based it in part on the radioactive creation argument.\textsuperscript{38}

A few fin de siècle scientists contemplated other ways in which radioactivity might be relevant to cosmological questions. Radioactivity not only produced new atoms but also liberated a large amount of energy apparently stored in the material structure of the atom. Moreover, at the time it was generally assumed that radioactivity was a common property of all matter, only appearing in different degrees of intensity. The British chemist Frederick Soddy was a pioneer of radioactive research and the discoverer, together with Ernest Rutherford, of the fundamental law of radioactive decay. In the first textbook ever on radioactivity he speculated that the strange irreversible phenomenon indicated an end of evolution, namely “when all the available energy shall have run its course to exhaustion.” And when he looked backwards in time the phenomenon might indicate a beginning:

Correspondingly, a sudden beginning of the universe—the time when present laws began to operate—is also fixed. It is necessary to suppose that the universe, as a thing in being, had its origin in some initial creative act, in which a certain amount of energy was conferred upon it sufficient to keep it in being for some period of years. It is possible, it is true, to avoid the end indefinitely, since the rate of change will diminish as the end is approached, and theoretically the end will require an infinite time to be attained. But the difficulty with the beginning cannot be so avoided.\textsuperscript{39}

We have here, possibly for the first time, the idea of a known physical mechanism responsible for the origin of matter in the universe. However, it was a speculation only and no more, and no more an anticipation of the big-bang scenario than Crookes’ earlier speculation. In fact, as Soddy (who was an atheist) revealed shortly later in his book, he was in favor of an eternally cyclic universe in which destructive decay processes were balanced by constructive creation processes. He suggested that this kind of cosmology was in better accord with the sciences of thermodynamics and radioactivity. “The universe,” he concluded, is “limited with reference neither to the future nor the past, and demanding neither an initial creative act to start it nor a final state of exhaustion as its necessary termination.” To many contemporary scientists this was a more appealing world view.

1.5 Gravitational collapse

At least on a relatively short timescale the near universe is stable. The sky we observe today is approximately the same that people observed in ancient Greece and the Babylonians before them. The question of the stability of the stellar universe became an issue with Newton’s universal law of gravitation which supposedly secured the stability and order of the universe. In a famous correspondence with the theologian Richard Bentley from the early 1690s Newton argued in qualitative terms that although a finite collection of stars would coalesce gravitationally, this would not be the fate of an infinite number of uniformly distributed stars. However, he did not give any proof and instead appealed to
God’s wisdom in placing the stars so delicately in space that a tiny perturbation would not cause the system to collapse.40

For two centuries Newton’s cosmological argument was accepted without being rigorously examined. This only happened in 1895 when the German astronomer Hugo von Seeliger, professor and director of the Munich Observatory, reconsidered the argument.41 Seeliger proved that an infinite Euclidean space with a roughly uniform mass distribution could not be brought into agreement with Newton’s law of gravitation. He showed that calculation of the gravitational force exerted on a body by integration over all the masses in the infinite universe does not lead to a unique result, as the integral diverges. Hence the conclusion, “Newton’s law, applied to the immeasurably extended universe, leads to insuperable difficulties and irresolvable contradictions if one regards the matter distributed through the universe as infinitely great.”42 Seeliger’s concern was not really to save Newton’s infinite stellar universe, for he rejected the notion of an actual infinity and tended to believe that the universe was finite. The following year Seeliger framed the gravitation paradox differently, by showing that the Newtonian universe allowed motions that start with finite speed and accelerate to infinitely great speeds in a finite time. As he pointed out, such motions are no less inadmissible than a collapsing universe. In a popular presentation of 1898 Seeliger summarized that whatever the mass distribution in the universe, there must occur infinitely great accelerations in it. His remedy for the conceptual sickness was to suggest that Newton’s law should be modified at very large distances (Fig. 1.5). A body of mass $m$ moving in the gravitational field of a central mass $M$ will, according to Newton, experience a gravitational pull given by

$$F(r) = G \frac{mM}{r^2}. \quad (1.6)$$

Seeliger now suggested that for very large distances, the body would move as if there were a repulsive force in addition to the attractive gravitational force. The suggestion amounted to introducing an attenuation factor of the form $\exp(-\Lambda r)$, which leads to the new law of force

$$F(r) = G \frac{mM}{r^2} e^{-\Lambda r}. \quad (1.7)$$

Seeliger contemplated that the constant $\Lambda$ might be important also in planetary astronomy and that its value might be inferred by assuming that it was responsible for the anomalous motion of Mercury’s perihelion. He found that the anomaly could be accounted for if $\Lambda = 3.8 \times 10^{-7}$ but also that this value would lead to unobserved disturbances in the motions of the other planets. For this reason he did not seriously pursue the idea. Nor did he pursue the radical idea that the law of gravity might conceivably vary in space. Yet it is interesting that he mentioned it, if only as a remote possibility. “Though highly probable,” he wrote, “it is far from self-evident that the forces of attraction follow the same law in all places throughout the universe.”43

In a book of 1896 the German physicist Carl Neumann repeated the conclusion of Seeliger, claiming that he had come to the same result of a modification of Newton’s
Die einfachste Formel, welche eine Absorption berücksichtigt, erhält man, wenn man den Newton'schen Gesetze $k^2 m m'$ den Factor $e^{-\lambda r}$ hinzufügt, wo $e$ die Basis des natürlichen Logarithmensystems ist. $\lambda$ wird nicht eine Konstante zu sein brauchen, aber sie soll als solche gelten. Die Anziehungskraft $A$ wird also ausgedrückt durch

$$A = k^2 m m' \frac{e^{-\lambda r}}{r^2}$$

Es ist ersichtlich, dass man $\lambda$ stets so klein wählen kann, dass innerhalb unseres Planetensystems mit beliebiger Annäherung das Newton'sche Gesetz hervorgeht. Andere-

Fig. 1.5 Seeliger's passage in Astronomische Nachrichten from 1895 in which he suggested his modified law of gravitation including a repulsive $\lambda$-constant.

law many years earlier. In the literature of the history of cosmology one can sometimes read of the “Neumann–Seeliger paradox.” However, as shown by John Norton this is a misunderstanding. The paradox belongs to Seeliger alone.

Seeliger’s constant can be considered a classical version of the cosmological constant that Einstein famously introduced in his cosmological field equations of 1917. In fact, at the very beginning of his paper of that year Einstein dealt with the same problem that Seeliger had identified, the incompatibility of Newtonian gravitation and an infinite stellar universe with a constant mass density. At the time Einstein was probably unaware of Seeliger’s old paper in Astronomische Nachrichten, but he later acknowledged the German astronomer’s modification of Newton’s law. The modified force law was essentially ad hoc and also arbitrary, since many other modifications might resolve the gravitation paradox in a similar way. The idea of modifying Newton’s inverse-square law was not, by itself, very original, as many such modifications were proposed in the nineteenth century. The exponential correction factor can be found in 1825 in Laplace’s Mécanique Céleste which can hardly have avoided Seeliger’s attention. However, what was original in Seeliger’s approach was that he used it in a cosmological context and not, as in most other proposals, just to solve problems of planetary astronomy such as Mercury’s anomalous revolution around the Sun.

William Thomson, apparently unaware of Seeliger’s work, arrived at basically the same results in two papers published 1901–1902. In the first paper he proved that in an infinite universe with a non-zero density of matter, “a majority of the bodies in the universe would each experience indefinitely great gravitational force.” Considering a homogeneous model universe of radius $3 \times 10^{16}$ km, he concluded that the number of stars, assumed to be of solar mass on average, must be neither too great nor too small. He found it “highly probable” that there were fewer than two billion stars within the sphere, and more than one hundred million. Thomson also calculated the time it took for the stellar system to collapse to a zero radius. This collapse time turned out to be independent of the initial size of the universe and to depend only on its mass density $\rho_0$:

$$t_{\text{collapse}} = \frac{1}{4} \sqrt{\frac{3\pi}{2}} G\rho_0.$$ (1.8)
For a universe containing one billion stars, the collapse time would be 17 million years, a figure of the same order as that which Thomson had found for the age of the Earth and the lifetime of the Sun. Contrary to Seeliger, Thomson did not present his investigation clearly as a problem for the infinite Newtonian universe, and he did not propose a way out of the problem.

There were other ways to escape gravitational collapse than modifying Newton’s law of gravitation. One could leave the law intact and change some of the cosmological assumptions of the Newtonian universe, such as the homogeneous distribution of matter. This is what the astronomy author Richard Proctor did in his widely read Other Worlds than Ours first published 1870, although he developed his model in the context of Olbers’ optical paradox and not the gravitational paradox. Proctor conceived of a hierarchic model universe in which the higher star systems were separated by increasingly larger distances from the lower ones. In that case, the contributions of light from stars lying in successive shells would not be equal, but successively less, and the total amount of light received from the infinite number of stars could be quite small.47

In 1908, Carl Charlier, professor of astronomy at Lund University, Sweden, developed in mathematical detail a hierarchic model in a paper titled “Wie eine unendliche Welt aufgebaut kann” (How an infinite world can be constructed).48 This was not the first time he expressed his interest in the grander aspects of cosmology. Unusually for professional astronomers at the time, Charlier transcended the divide between scientific astronomy and philosophical reflections on the more speculative aspects of cosmology. In 1896 he argued that Olbers’ and Seeliger’s paradoxes indicated that the universe was finite, a solution he preferred to Seeliger’s modification of Newton’s law of gravity. However, he also pointed out that the paradoxes rested on the assumption of uniformly distributed stars, an assumption that might be questioned. If the distribution of stars followed a law such that “the density of the stars decreases faster, as we move out in space,” the paradoxes did not need to arise.49 As mentioned, this idea, as far as it is related to Olbers’ paradox, goes back to John Herschel in 1848 and was revived by Proctor in 1870. As to the temporal extension of the universe Charlier thought that a beginning and end were ruled out by the law of mass conservation as well as common sense. “A finite time is a contradiction,” he wrote, “an infinite time may be difficult to conceive, but it is not contradictory.”

Let the Milky Way $S_1$ be composed of $N_1$ stars and let $N_2$ Milky Way galaxies form a second-order galaxy $S_2$; $N_3$ galaxies of type $S_2$ form a third-order system $S_3$, and so on. The system $S_1$ has a radius $R_1$. Charlier showed that if the mean density of a structure of size $R$ decreased as

$$\rho \sim R^{-\alpha} \text{ with } \alpha > 2,$$

then the inequality

$$\frac{R_{i+1}}{R_i} \geq \sqrt{N_{i+1}}$$

was satisfied. As a consequence Seeliger’s gravitation paradox would disappear and there would be no infinite velocities. In Charlier’s hierarchical model the mass of the universe
was infinite but its average density zero. Charlier further derived an inequality for the limiting angular diameter of a spiral nebula, namely

\[ \theta < N^{-1/6}. \]  

(1.11)

From this he found values for the angular diameter and apparent magnitude of the nearest spiral that agreed qualitatively with the observed values of Andromeda.

Charlier's ideas received support from Franz Selety, a Viennese physicist, who in 1922 developed them as a Newtonian alternative to the relativistic cosmology that Einstein had introduced five years earlier.\(^50\) Selety criticized Einstein's theory of a closed universe and argued that if matter became diluted with distance in a suitable way (faster than \(1/r^2\)), Newtonian theory allowed an infinite universe completely filled with matter. The local matter density in Selety's "hierarchical molecular world" was everywhere finite and yet the average density would tend to zero. A brief debate followed between Einstein and Selety in the pages of *Annalen der Physik*, but after his first reply in 1922 Einstein chose not to reply to Selety's papers.

The non-uniform universe proposed by Charlier continued to attract interest among cosmologists who did not accept the cosmological uniformity principle. One of them, the Polish-American physicist Ludwik Silberstein, dismissed in a book of 1929 the closed Einstein model and its apparent agreement with Edwin Hubble's count of nebulae. Although Silberstein stayed on the ground of relativistic cosmology he much preferred a hierarchic world model of the type suggested by "the great Swedish astronomer."\(^51\)

One year after Charlier's 1908 paper had appeared, another Swedish scientist, the Nobel laureate Svante Arrhenius, examined the gravitation paradox. As mentioned, Arrhenius was an ardent advocate of a homogeneous and infinite universe, and consequently he felt it necessary to criticize Charlier's hierarchic model. One might have thought that he would be sympathetic to Seeliger's solution, but this was not the case. He could see no problem with the Newtonian universe and therefore found Seeliger's work to be irrelevant. (As Seeliger was quick to point out, Arrhenius had partly misunderstood his work.) "There really is no weighty reason why the world would not be sown uniformly with stars," Arrhenius concluded.\(^52\) Arrhenius' denial that there was a problem at all hardly convinced anyone except himself. When it came to the notion of a finite-age universe Arrhenius shared the view of Charlier, namely that it contradicted the fundamental conservation laws of physics and for this reason alone was ruled out. As Arrhenius phrased it in a textbook of 1903, a universe with a beginning in time "is hard to bring into agreement with the indestructability of energy and matter."\(^53\) Indeed it is, and it was not the last time that the difficulty was pointed out.

### 1.6 Non-Euclidean astronomy

Whereas curved space as a mathematical concept dates from the early nineteenth century, it took more than a century before it permeated to the physical and astronomical sciences. The eminent mathematician and polymath Karl Friedrich Gauss arrived as early as 1816 to the conclusion that the ordinary, flat or Euclidean geometry was not
true by necessity. As he wrote in a letter to Olbers, “Maybe in another life we shall
attain insights into the essence of space which are now beyond our reach. Until then
we should class geometry not with arithmetic, which stands purely a priori, but, say, with
mechanics.”\(^{54}\) Gauss realized that if the question could be settled at all it would require
astronomical measurements over very large, stellar distances. Although Gauss seems to
have been aware that information about the curvature of space could in principle be
obtained from data of stellar parallaxes, he did not pursue this kind of reasoning. In
a letter to the German-Danish astronomer Heinrich C. Schumacher of 12 July 1831
he communicated the formula for the circumference of a circle of radius \(r\) in the new
geometry, stating it to be

\[
l = \frac{\pi}{2} k \left[ \exp \left( \frac{r}{k} \right) - \exp \left( -\frac{r}{k} \right) \right].
\]  

(1.12)

The quantity \(k\) is a constant to be determined observationally and which, in the Euclidean
case, is infinite. Should space not be Euclidean, Gauss said in his letter to Schumacher,
from an observational point of view we can only say that the curvature measure \(k\) must
be incredibly large.

The true founders of non-Euclidean geometry were the Hungarian mathematician
János Bolyai and the Russian Nikolai Ivanovich Lobachevsky, both of whom published
their independent discoveries that a geometry different from and as valid as Euclid’s is
possible. Whereas Bolyai did not refer to the skies, the astronomy-trained Lobachevsky
did. Although primarily a mathematician, as a young man he had studied astronomy and
in the 1820s he served as director of the Kasan University Observatory. Already in his
1829 paper in the Kasan Messenger Lobachevsky suggested that one consequence of his
“imaginary” (or hyperbolic) geometry might be tested by astronomical means.\(^{55}\) As he
pointed out, the angle sum of a triangle would always be less than 180\(^\circ\) and the more
so the bigger the triangle becomes. From astronomical data he concluded that the angle
sum of the triangle spanning the Sun, the Earth, and Sirius deviated from the Euclidean
value of 180\(^\circ\) by at most 0.000372 arcsec, evidently much less than the observational
error. The modern value of Sirius’ parallax is 0.37 arcsec, which is less than a third of
the value adopted by Lobachevsky.

The Russian mathematician realized that while it could in principle be proved that
astronomical space is non-Euclidean, it could never be proved to be Euclidean, and
for this reason he tended to see his comparison as inconclusive. In a later work, titled
Pangeometry and translated into French in 1856, he argued that, assuming space to be
hyperbolic, there must be a minimum parallax for all stars irrespective of their distances
from the Earth. This is contrary to flat space, where the parallax tends toward zero as
the distance increases toward infinity. Consider a triangle spanned by a star and the two
positions of the Earth half a year apart in its orbit around the Sun. Let the angle at the star
be denoted \(\alpha\) and the two angles at the positions of the Earth be \(\beta\) and \(\gamma\). Lobachevsky
showed that the parallax angle \(p\) can be expressed as

\[
p = \pi - (\beta + \gamma) = \alpha - K\beta,
\]  

(1.13)
where the curvature is \( K = 0 \) in flat space. The ideas of non-Euclidean geometry pioneered by Gauss, Bolyai, and Lobachevsky circulated but slowly in the mathematical community. Only about 1870 did they truly enter the world of mathematics if not yet the world of the astronomical and physical sciences.\(^{56}\) The new ideas then resulted in a revolution in geometry. Of particular importance was the work of the Göttingen mathematician and physicist Bernhard Riemann, who in a famous address of 1854 (published only 1867) put the concept of curvature as an intrinsic property of space on a firmer basis. Importantly, his address led to the now standard distinction between the three geometries of constant curvature. The curvature \( K \), which has the dimension of an inverse area, relates to the radius of curvature \( R \) by

\[
R^2 = \frac{k}{K} \text{ with } k = 0, \pm 1. \tag{1.14}
\]

The three possibilities correspond to Euclidean space (curvature constant \( k = 0 \)), spherical space \((k = +1)\), and hyperbolic space \((k = -1)\).

Riemann referred only to astronomy in passing, pointing out that on the assumption of a constant-curvature space it follows “from astronomical measurements that it [the curvature] cannot be different from zero.”\(^{57}\) Although he accepted the idea of just one physical space, like Gauss he did not accept that the geometry of this space could be known a priori or with absolute certainty. Our experience of the physical world, he said, could be consistent with any of the three possible geometries. Moreover, he left open the possibility that on a microphysical scale the curvature of space might vary in such a way that the averaged curvature over measurable distances becomes inappreciably close to zero.

Although Riemann’s emphasis on the possibility of an unbounded yet finite space implicitly addressed an old cosmological conundrum, it failed to attract interest among astronomers. It took nearly two decades before a scientist made astronomical use of Riemann’s insight. When it happened it made no impact at all on the astronomical community. The Leipzig astronomer Johann C. Friedrich Zöllner was a pioneer astrophysicist known in particular for his contributions to astrophotometry. After 1877 he focused increasingly on what he called “transcendental physics,” the study of spiritualist phenomena based on the postulate of a fourth space dimension. As one might expect, this line of work created so much public attention that it damaged his scientific reputation.\(^{58}\) A few years before his conversion to spiritualism, Zöllner published a controversial book with the title *Natur der Cometen* (The Nature of Comets; Fig. 1.6) which included a chapter of considerable interest to the history of cosmological thought. By arguing that cosmic space might be positively curved in accordance with Riemann’s idea, Zöllner offered an original solution to Olbers’ paradox.

In his systematic discussion of the finite versus the infinite in the universe Zöllner assumed, for the sake of discussion, that there is only a finite amount of matter in the world. He then argued that in an infinite Euclidean space any finite amount of matter would eventually evaporate and dissolve to zero density. Given the actual existence of matter he concluded that either is space finite or the universe has only existed
in a limited period of time. Unwilling to accept the latter hypothesis he suggested that Riemann’s geometry might provide the key that would unravel the secrets of the universe. “It seems to me,” he wrote, “that any contradictions will disappear...if we ascribe to the constant curvature of space not the value zero but a positive value. The assumption of a positive value of the spatial curvature measure involves us in no way in contradictions with the phenomena of the experienced world if only its value is taken to be sufficiently small.” In this way he could explain Olbers’ paradox without having to assume interstellar absorption of starlight or taking recourse to a limitation of either cosmic time or space. Zöllner’s innovative cosmological speculations attracted some interest in German philosophical circles, but were ignored by almost all physicists and astronomers. They were essentially forgotten.

During the last quarter of the nineteenth century, non-Euclidean geometry became a hot topic in mathematical and philosophical circles, and it was discussed in hundreds of books and scientific papers. On the other hand, the number of astronomers who

---

**Fig. 1.6** Title page of Zöllner’s 1872 book on the nature of comets, including his proposal of a geometrically closed universe.
expressed interest in the topic can be counted on the fingers of one hand. Moreover, the interest rarely went beyond uncommitted comments and it appeared more often in popular than scientific works. The Irish astronomer Robert Stawell Ball, who from 1892 to 1913 served as professor of astronomy and geometry at Cambridge University, mentioned the possibility of a positively curved space at some occasions but without endorsing it as a reality. This cautious and ambiguous attitude was shared by his distinguished American colleague Simon Newcomb, who discussed the hypothesis in the widely read *Popular Astronomy* first published in 1878. Like Lobachevsky many years earlier, Newcomb pointed out that the hypothesis was testable although this might perhaps be more in principle than in practice. As he wrote, “Unfortunately, we cannot triangulate from star to star; our limits are the two extremes of the earth’s orbit. All we can say is that, within those narrow limits, the measures of stellar parallax give no indication that the sum of the angles of a triangle in stellar space differs from two right angles.”

A few physicists and mathematicians wondered if the puzzle of the anomalous motion of Mercury’s perihelion might be solved on the assumption that the geometry of space is non-Euclidean. Thus, in 1885 the German mathematician Wilhelm Killing derived the orbit of Mercury moving in spherical space by means of a modified law of gravitation that he found suitable for that kind of space. His investigation was of mathematical interest only and the same was the case with a similar work by Carl Neumann. The later approach of the Austrian physicist Josef Lense was different in so far that he assumed Newton’s law to be universally true and applied it to spherical space. In a paper of 1917 he derived in this way an expression for the perihelion advance but unfortunately his expression could not be translated into a numerical answer. In any case, by that time the motion of Mercury’s perihelion was no longer anomalous. The problem had been solved on the basis of Einstein’s general theory of relativity.

In the period before general relativity only two astronomers, both of them German, examined more systematically the possibility of a Newtonian stellar universe embedded in non-Euclidean space. One of them was 27-year-old Karl Schwarzschild, who in a lecture of 1900 discussed how to determine the geometry of space from astronomical observations. He summarized his results as follows:

One may, without coming into contradictions with experience, conceive the world to be contained in a hyperbolic (pseudo-spherical) space with a radius of curvature greater than 4,000,000 earth radii, or in a finite elliptic space with a radius of curvature greater than 100,000,000 earth radii, where, in the last case, one assumes an absorption of light circumnavigating the world corresponding to 40 magnitudes.

Schwarzschild saw no way to go further than this rather indefinite conclusion and decide observationally whether space really has a negative or positive curvature, or whether it really is finite or infinite. Nonetheless, from a philosophical and emotional point of view he definitely preferred a closed universe. His reason was that “then a time will come when space will have been investigated like the surface of the Earth, where macroscopic investigations are complete and only the microscopic ones need continue.”
It is noteworthy that some later cosmologists, including Eddington and Lemaître, expressed their preference for closed universe models in similar philosophical language. Eight years later Paul Harzer, professor of astronomy at the University of Kiel, developed a more detailed model of a closed stellar universe. Harzer's universe was enclosed in a finite cosmic space with a volume about 17 times that of the stellar system. This system contained the same number of stars but was compressed to a size approximately one half of what it had in other models based on flat space. The size of the entire universe was given by the time it took for a ray of light to circumnavigate it, which Harzer estimated to be 8,700 years. The Schwarzschild–Harzer suggestion of a closed space filled with stars had the conceptual advantage that it did away with the infinite empty space, but it made almost no impact on mainstream astronomy. The cosmological problem that moved to the forefront of astronomy in the 1910s was concerned with the size of our Galaxy and the question of whether the spiral nebulae were external objects or belonged to our own Galaxy. This was a problem in which the geometry of space was considered to be irrelevant. In short, the main reason why astronomers were reluctant to consider the consequences of space being non-Euclidean was that they had no need for the hypothesis.

Among the very few scientists who referred to curved space as more than just a mathematical curiosity before Einstein was also the Austrian meteorologist Wilhelm Trabert, a professor in Vienna. In a textbook of 1911 on what at the time was known as “cosmical physics” he included a chapter on cosmology in which he discussed Olbers’ paradox, the anomalous motion of Mercury’s perihelion, and other issues. In this context he contemplated that the universe might be finite because space is finite. It is conceivable, he wrote, that astronomical observations would one day demonstrate that “light, gravitation and electricity do not propagate in straight lines, but in circles.”

Schwarzschild did not consider his investigation of 1900 to be very important, but it was well known and its main conclusions reappeared in the 1911 edition of the recognized and widely read Newcomb–Engelsmanns *Populäre Astronomie*. Schwarzschild’s early work was also of some significance in the earliest phase of relativistic cosmology. When Willem de Sitter developed his model universe in 1917 based on Einstein’s field equations, he referred to the limit of stellar parallaxes which Schwarzschild had found earlier on the assumption of cosmic space being either elliptic or hyperbolic. “The limit found by Schwarzschild still corresponds to our present knowledge,” de Sitter wrote. Most likely, it was through de Sitter that Einstein first became acquainted with the astronomical works of Seeliger and Schwarzschild.

1.7 Finite or infinite?

Conceptual and speculative ideas about the universe as a whole did not occupy a major role in nineteenth-century astronomy and cosmology. In so far that such ideas were discussed the question of the size of the universe permeated much of the discussion. Briefly put, is the universe finite or infinite in extension? The “universe” might in this context refer to either space or, what was more common, to the material stellar system
Cosmological theories before and without Einstein

embedded in it. As we have seen, the question was at the basis of the discussions concerning Olbers’ paradox, the gravitation paradox, and also of the cosmic consequences derived from the second law of thermodynamics. It was not, however, a question that kept most astronomers awake at night; and they were even less concerned with the possibility of an origin of the universe.

At the end of the nineteenth century it was generally recognized that if an observationally grounded cosmology was ever to be established, a first step would be to understand the size and structure of the Milky Way. Space was usually thought to be infinite, but there was no consensus at all regarding the distribution of stars and nebulae in the universe. Many astronomers hesitated to admit it as a problem that could be solved scientifically by means of the usual combination of theory and observations. In an article of 1869 Proctor gave voice to the dilemma: “The only question for us is between an infinity of occupied space and an infinity of vacant space surrounding a finite [material] universe. Either idea is equally incomprehensible; but the former is merely beyond, the latter seems contrary to reason.”

On this issue Thomson agreed, and with basically the same argument as Proctor. He was convinced that matter was distributed endlessly through space, which was one reason why he did not accept the cosmic version of the heat death as real. “Finitude is incomprehensible, the infinite in the universe is comprehensible,” he claimed in a popular lecture of 1884. The illustration of the claim made popular sense but was far from a scientific argument: “What would you think of a universe in which you could travel one, ten or thousand miles, or even to California, and then find it come to an end? Can you suppose an end of matter or an end of space?” While only the few scientists acquainted with Riemannian geometry could conceive a finite space, to many astronomers an end of matter was perfectly intelligible. This kind of subjective argument—based on what individual scientists could comprehend or not comprehend—colored much of the debate not only in the pre-Einstein era but also in later periods.

In a book of 1890 the Irish astronomer and historian of astronomy Agnes Mary Clerke made clear that the nebulae were parts of the Milky Way rather than separate galactic systems. What, then, about the universe beyond the Milky Way? Clerke raised the question only to dismiss it as unworthy of scientific consideration: “With the infinite possibilities beyond, science has no concern.” In a later book, the massive _Problems of Astrophysics_, she likewise refrained from going beyond what she poetically called “the equatorial girdle of a sphere containing stars and nebulae.”

As she wrote:

> The whole material creation is, to our apprehension, enclosed within this sphere. We know nothing of what may lie beyond. Thought may wander into the void, but observation cannot follow. And where its faithful escort halts, positive science comes to a standstill. Fully recognizing the illimitable possibilities of omnipotence, we have no choice but to confine our researches within the bounds of the visible world.

There were many voices like Clerke’s, Newcomb’s being one of them. In an essay on unsolved problems in astronomy Newcomb asked if the universe was populated
with stars all over, or if they were largely contained in the system of the Milky Way, itself floating in infinite empty space. The question, he wrote, “must always remain unanswered by us mortals... Far outside of what we call the universe might still exist other universes which we can never see.” For all practical purposes, the Milky Way “seems to form the base on which the universe is built and to bind all the stars into a system.” In an age of positivism, the general attitude was that theories and hypotheses were put forward to explain facts, and hence, when there were no facts to be explained, no theory was required. “As there are no observed facts as to what exists beyond the farthest stars, the mind of the astronomer is a complete blank on the subject,” Newcomb wrote. “Popular imagination can fill up the blank as it pleases.”

To mention yet another example of the agnostic tradition prevailing in astronomy, consider the presidential address that the physicist and astronomer George H. Darwin, a son of Charles Darwin, gave to the British Association in 1905. Does it not seem as futile to imagine that man “can discover the origin and tendency of the universe as to expect a housefly to instruct us as to the theory of the motions of the planets?” Although Darwin admitted that great progress had occurred in the sciences, he concluded that “the advance towards an explanation of the universe remains miserably slight.” His prediction of future knowledge was as pessimistic as it was wrong: “We may indeed be amazed at all that man has been able to find out, but the immeasurable magnitude of the undiscovered will throughout all time remain to humble his pride. Our children's children will still be gazing and marveling at the starry heavens, but the riddle will never be read.”

The question of the spatial and material finitude of the universe was hotly debated in the last decades of the nineteenth century, if more among philosophers, theologians and social critics than among professional astronomers. The Austrian amateur astronomer Rudolf Falb, founder of the popular-astronomy magazine *Sirius*, argued that the very concept of space had only meaning in connection with matter. From this premise followed the possibility of limited space, as the question of what is beyond the end of space was deemed illegitimate; for according to Falb there was no way to define “beyond.” His view of a necessary connection between the concepts of space and matter was a curious return to Aristotelian natural philosophy. It allowed Falb to conceive of space as limited without questioning the implicit assumption of Euclidean geometry.

While the consensus view was that temporal finitude followed spatial finitude, this was not the opinion of Falb. On the contrary, he was convinced that the universe must be as eternal as matter. A universe consisting of a limited number of stars and nebulae would eventually contract gravitationally to a giant sun, but this sun would immediately evaporate and form an expanding gas, a nebulous body of the same kind as the one that had originally given rise to the celestial bodies. Falb contended that the process would go on endlessly: “The life of the world is to be conceived as a recurrence of expansion and contraction, like the breaths of a monstrous colossus. In this way the eternity of processes becomes understandable, that is, the infinite duration of the universe.... The end of the world is at the same time the beginning of the world.”

If only in a speculative and classical sense Falb’s scenario included the idea of a finite expanding universe with expansion and contraction in eternal change. A somewhat similar idea was later proposed by Ludwig Zehnder, a Swiss-German professor of
physics in Freiburg and a student of Wilhelm Röntgen. In a book of 1897 Zehnder assumed that all celestial bodies were confined within a gigantic sphere of ether. Like several other scientists at the time, he conceived the ether as a kind of rarefied gas consisting of tiny atoms endowed with mass and other physical properties. A sidereal universe including a limited number of stars—a picture favored by Olbers’ paradox—would be subject to gravitational collapse. However, according to Zehnder this would not be the end of the world, for the cooling of the central stellar body would lead to an increase of the kinetic energy of the ether atoms and to an increased electrical repulsion between atoms of matter. The result would be an expansion of the material universe, eventually to be followed by a new contraction when the gravitational attraction came to dominate the electrical repulsion. In this way Zehnder believed he had constructed on a scientific basis an eternal and cyclic universe.73

The speculations of Falb and Zehnder referred to what can possibly be regarded as anticipations of the expanding universe, although in this case the expanding phase was just part of a longer cyclic evolution. But although one can find in their works the phrase “expanding universe” (Ausdehnung der Welt) it was of course in a sense quite different to the one that later entered the cosmological vocabulary. In this context it may be relevant to mention that the German word for expansion (Ausdehnung) may also have the different meaning of expanse. The great mathematician David Hilbert gave in 1925 a famous address on the infinite in which he stressed that although the concept made mathematical sense it had no counterpart in real nature. In this connection he spoke of die Ausdehnung der Welt, but it was not an allusion to what we call the expanding universe. Hilbert undoubtedly referred to the expanse and not to the expansion of the world.74

Whereas the notion of infinite space was largely uncontroversial it was commonly claimed that the corresponding notion of an infinity of material objects, whether atoms of stars, was absurd. The claims were mostly based on philosophical and logical arguments and in many cases they were on the rhetorical level only. Since the time of Aristotle the majority of mathematicians and philosophers had taken for granted that so-called realized or actual infinities, as distinct from potential infinities, cannot exist. There typically was a strong element of theology in these claims. For example, according to Augustin Cauchy, a brilliant mathematician and a passionate Catholic, infinity and eternity were divine attributes not to be found in nature. Many thinkers of a Christian orientation continued to hold this view. To claim that the physical universe is infinite would obscure a crucial difference between God and nature and, they suggested, open up for pantheism and like heresies.

On the other hand, it was a dogma among materialists, atheists, and positivists that the universe is infinite in extension, which was usually taken to imply spatial as well as material infinitude. When it came to infinitism, the problematic issue was usually the number of things in the universe, not the more abstract concept of spatial extension. It was relatively easy to conceive space as infinite, such as Thomson did, but much harder to conceive it as filled with matter. For in this case one was faced not only with astronomical problems, such as Olbers’, but also with the philosophical problems arising with of an actual infinity of numbers of objects or processes. According to the German mathematician Georg Cantor’s theory of transfinite numbers, dating from the 1880s, the
actual infinite existed in the same sense that finite numbers exist, namely, as a logically consistent and operational concept. But one thing is mathematical consistency; another question is whether or not an actual infinite belongs to the real world as examined by physicists and astronomers. Cantor’s theory of numbers attracted the attention of philosophers but was thought to be too abstract to apply to the real cosmos.

In the mid-1920s another brilliant German mathematician, David Hilbert, discussed the possible relevance of Cantor’s actual infinities to the real world. Although Hilbert admired Cantor’s theory, he argued that the infinite was merely an idea and that it had no role to play whatever in the empirical sciences. It was in this context that he introduced the imaginary “Hilbert’s hotel” with infinitely many rooms. Even though all rooms are occupied in the hotel there will always be free rooms for new guests—and even infinitely many of them! Only much later did Hilbert’s bizarre hotel enter cosmological thinking, first in George Gamow’s popular book *The Creation of the Universe* from 1952 and subsequently, if only peripherally, in relation to the steady-state theory of the universe.75

In the late nineteenth century ideas of space-only infinitism depended on one’s conception of space. If space was merely thought of as a container of matter and physical phenomena, as was typically the picture of the Milky Way universe, problems did not necessarily arise. The situation was different if space was thought to be filled up with ether, and especially if the ether was conceived to be endowed with energy or otherwise being substantial in nature. As mentioned, during the second part of the nineteenth century belief in the infinite universe was correlated with atheism, and finitism with theism. And yet the correlation was not generally valid. In fact, for centuries there had been an opposite tradition of linking God’s omnipotence to a divinely created, infinite universe. Descartes, Newton and Leibniz conceived the infinite universe as the only one worthy for the almighty God, and a similar view was held by Kant and Olbers. A Christian scientist as he was, Thomson nonetheless concluded that a spatially finite universe was impossible. However, at the time Thomson was an exception. Astronomers in the fin de siècle period realized that the question of the size of the universe could not be decided observationally and for this reason they rarely discussed it.

However, cosmology in the first decade of the twentieth century was far from restricted to astronomers’ analyses of the Milky Way. In the eyes of the public, Arrhenius may have been the period’s best known “cosmologist,” a reputation due to a large extent to his highly successful *Worlds in the Making* of 1908. While many astronomers conceived the material universe to be roughly identical to the Milky Way, Arrhenius based his cosmology on what was later called the cosmological principle, namely, that the universe is uniformly populated throughout with stars and nebulae (Section 4.3). And not only that, for Arrhenius’ guiding principle was “the conviction that the Universe in its essence has always been what it is now. Matter, energy, and life have only varied as to shape and position in space.”76 In other words, he built his theory on what in the later steady-state theory of the universe would be called the perfect cosmological principle (see Section 5.5). Like several other writers at the time, Arrhenius advocated an infinite, eternally cyclic universe. He argued that such a world view did not necessarily disagree with the second law of thermodynamics since radiation pressure provided a mechanism
to keep the universe going for ever. After Arrhenius’ hypothesis had been shown to be untenable by Poincaré in 1911 and by Schwarzschild two years later, little more was heard of it in scientific circles. But although most astronomers ignored his ideas, they were widely discussed by laymen and amateur astronomers.

The Norwegian physicist Kristian Birkeland, an authority in research of the aurora borealis, shared with Arrhenius his ambition of establishing cosmology on the basis of physical mechanisms. In articles published between 1911 and 1916 he extrapolated his early ideas of the origin of the aurora into a grand cosmological sketch of a universe dominated by cathode rays and other electrified particles (Fig. 1.7). According to Birkeland:

One of the most peculiar features of this cosmogony is that space beyond the heavenly bodies is assumed to be filled with flying atoms and corpuscles of all kinds in such density that the aggregate mass of the heavenly bodies...would only be a very small fraction of the aggregate mass of the flying atoms there. And we imagine that an average equilibrium exists in infinite space, between disintegration of the heavenly bodies on the one hand, and gathering and condensations of flying corpuscles on the other.77

For a brief period of time Birkeland’s theory attracted international attention, if mostly outside the astronomical community. For example, New York Times covered in detail

---

**Fig. 1.7** Kristian Birkeland (1867–1917) to the left, demonstrating discharge phenomena with his model of the Earth placed in a vacuum chamber and exposed to cathode rays. The person to the right is his assistant Karl Devik.

his “amazing picture of the future development of the universe.” The important point in Birkeland’s cosmology was the hypothesis that the stars and nebulae were potent generators of electrons and ions, with the consequence that interstellar and intergalactic space was filled with a tenuous ionized gas. In effect, he introduced the notion of what was later called space plasma. Many years after Birkeland’s death his ideas were reconsidered by the Swedish physicist Hannes Alfvén, a Nobel Prize laureate of 1970, who turned them into a theory of plasma cosmology.\textsuperscript{78}

Einstein’s theory of 1917 is rightly considered a watershed in modern cosmology, but it took a long time before the majority of astronomers recognized that a revolution had occurred. In the spring of 1924 Charlier gave a series of lectures on cosmology at the University of California which were published the following year under the title “On the Structure of the Universe.” A supporter of the island-universe theory Charlier had no problem in dealing with “the unknown world at the other side of the Milky way” which he presented in terms of his hierarchic theory of an infinite system of nebulae. His lectures were in the style of traditional cosmology built on stellar statistics, with no mention of nebular redshifts and no indication that the general theory of relativity had changed the foundation of cosmology. Although Charlier ignored the theories of Einstein and de Sitter, he was undoubtedly aware of them. His only, indirect reference to the closed Einstein universe was in relation to the heat death: “You know that there are also speculative men in our time who put the question whether space itself is finite or not, whether space is Euclidean or curved (an elliptic or hyperbolic space).” But Charlier had no like for these speculative men whom he compared to the philosopher Kant—and he did not mean it as praise. Their speculations, he said, “must be discussed on the support of facts and at present they must be considered as lying outside such a discussion.”\textsuperscript{79}

Still in the early 1930s Charlier’s classical universe was alive and taken seriously by a few astronomers. The distinguished astronomer Heber Curtis, director of the University of Michigan observatories and a specialist in the study of spiral nebulae, preferred the hierarchical system over the new expanding models of the universe based on the field equations of general relativity. “The remarkably close parallelism between the structure of the suggested Charlier universe and the observed arrangement of the accessible exterior universe, is too exact and detailed to be cast lightly aside,” he wrote in the authoritative \textit{Handbuch der Astrophysik} (Handbook of Astrophysics): “For this and other reasons, therefore, the writer prefers, solely as a matter of personal choice and belief, to adhere, until more evidence may be gathered, to a cosmogony based on the scheme of Charlier, rather than to accept one or another of those which seem as yet somewhat nebulously based upon a four-dimensional frame of reference.”\textsuperscript{80} And this was not the last time that a leading astronomer found inspiration in Charlier’s classical picture of the universe, if as a complement rather than an alternative to the expanding universe based on the theory of general relativity. As late as 1970 the French-American astronomer Gérard de Vaucouleurs suggested that “a hierarchical structure \textit{à la} Charlier must be included in any realistic cosmological model.”\textsuperscript{81}
NOTES

1. There are only few modern works which describe the historical development of
cosmology in its entirety. See Kragh (2007) and North (2008). This chapter
draws on material in the first source. Merleau-Ponty (1983), an important work on
nineteenth-century cosmology and astronomical philosophy, only exists in its French
original.
3. See Jaki (1974) for a history of the idea of the cyclic or oscillating universe. For a
4. Bondi (1952), p. 21. The history and substance of Olbers’ paradox is detailed in
9. According to the Helmholtz–Thomson theory first proposed by Helmholtz
in 1854, the source of stellar energy was gravitational. In modern astro-
physics the KH or Kelvin–Helmholtz timescale refers to the time for a stellar
body (mass \( M \) and radius \( R \)) to radiate away all of its potential energy.
The KH time is given by \( GM^2/RL \), where \( L \) is the average luminosity. See
Kragh (2016a) for details of the classical contraction theory and its cosmological
implications.
absorption, see also Seeley and Berendzen (1972a, 1972b).
11. Trumpler (1930).
14. For details on the history of astrospectroscopy, see Hearnshaw (2014), and for
spectroscopy generally, Hentschel (2002).
17. For early quantum astrophysics and references to the literature, see Tassoul and
Tassoul (2004).
19. For Doppler’s life and work, see Schuster (2005).
20. The nebular hypothesis was as much a world view as a scientific theory. For aspects
of the hypothesis and its reception, see Brush (1987) and Schaffer (1989).
Huggins’ work and retrospective accounts.
23. The riddles of the spectral lines attributed to coronium and nebulium are described in Hufbauer (1991), pp. 112–115 and Hirsh (1979).
24. Kragh (2009b) is a detailed account of the history of helium up to and including its discovery in 1895. The role of helium in the birth of astrophysics is described in Nath (2013).
25. Tayler (1995). See also Chapter 8 in the present volume.
26. Chamberlin (1899), p. 12. The term “atomecule” was invented by Chamberlin, perhaps in the meaning of a hypothetical subatomic entity. It was not used by other scientists either then or later.
28. Schuster (1898).
29. This section is largely taken from Kragh (2007). A more detailed account of the discussion concerning the cosmological consequences of the second law of thermodynamics can be found in Kragh (2008a) which includes references to the primary sources.
36. Maunder (1908).
38. “The idea of this hypothesis arose when it was noticed that natural radioactivity is a physical process which disappears gradually…If it were not for a few elements of average lifetimes comparable to $T_H$ [Hubble time], natural radioactivity would be completely extinct now.” Lemaître (1949), p. 452. On Haas, see Kragh (2008a), pp. 70–2. See also Chapter 4.
39. Soddy (1904), p. 188. In 1921 Soddy was awarded the Nobel Prize in chemistry for his work on radioactivity and isotopes.
40. See Harrison (1986).
41. The cosmological problems associated with Newton’s law of gravitation are described in Jaki (1979) and in mathematical details in Norton (1999).
42. Seeliger (1895), p. 132.
43. Seeliger (1898a), p. 547.
45. Einstein (1956), p. 65, in the 17th edition of a book originally published by Einstein in 1917, wrote of a “fundamental difficulty…which, to the best of my knowledge, was first discussed in detail by the astronomer Seeliger.”
46. Thomson (1901), and see also Norton (1999).
48. Charlier (1908) and also Charlier (1922). For his cosmological views, see Norton (1999), Jaki (1969), pp. 198–204, and Holmberg (1999), pp. 73–78.
50. On Selety’s work, see Jung (2005).
54. Quoted in Kragh (2012a), which gives details about curved-space astronomy before Einstein.
55. Daniels (1975). It is a popular and persistent myth that Gauss made high-precision geodetic measurements of a triangle extending between three mountain peaks in Germany in order to test whether physical space is flat or not. See Breitenberger (1984) for the correct story.
56. The rise of non-Euclidean geometry during the nineteenth century is illustrated by Duncan Sommerville’s bibliography of 1911 which includes about 4,000 titles published between 1830 and 1909. By far most of the publications date from after 1870 and only a few dozens of them refer to physics and astronomy. See Sommerville (1911).
57. Riemann (1873), p. 36.
58. See Kragh (2012c) for Zöllner as astronomer, cosmologist, and spiritualist. See also Jaki (1969), pp. 158–64.
59. Zöllner (1872), p. 308
61. Roseveare (1982) is a detailed account of the many attempts to solve the Mercury anomaly in the period before Einstein. See pp. 162–4 for the use of curved space.
63. Harzer (1908).
64. Trabert (1911), p. 259.
65. De Sitter (1917b), pp. 231–234. See Chapter 3 for the cosmological models of Einstein and de Sitter.
67. Thomson (1891), p. 322. As the historian of astronomy John North remarked: “It is easy to speak of the infinite, as every theologian knows, but it is difficult to speak of it meaningfully.” North (1990), p. 23.
69. Newcomb (1906), pp. 5–6. The latter quotation is from a review Newcomb wrote in The Observatory 30 (1907), p. 362.
70. Darwin (1905), p. 32.
71. The issue is discussed and provided with examples in Kragh (2008a).
73. Zehnder (1897), who in 1914 published an updated version of his cosmological theory in a book with the characteristic title Der Ewige Kreislauf des Weltsalls (The Eternal Cycle of the Universe).
74. Hilbert was well aware of Einstein’s curved-space cosmology to which he referred and he might conceivably have read also Friedman’s paper of 1922 in which the
expanding universe was introduced. But there is no evidence that Hilbert knew of Friedman’s work or had it in his mind.

75. Hilbert introduced his remarkable hotel in unpublished lectures that he gave in Göttingen in 1924–1925. For the history of Hilbert’s hotel, see Kragh (2014c). The general problem of infinities in cosmology is dealt with in North (1990), pp. 371–383.

76. Arrhenius (1908), p. xiv. For Arrhenius’ theories of radiation pressure, the aurora borealis, and the universe, see Kragh (2013a).


79. Charlier (1925b, 1925c) and Charlier (1925a), p. 182. He probably referred to the theories of Einstein and de Sitter but may also have had in mind Schwarzschild’s article of 1900.

80. Curtis (1933), p. 908. Contrary to some other critics of the expanding universe, Curtis was well acquainted with the theories of Lemaître, Eddington, Tolman and other pioneers of relativistic cosmology.

81. De Vaucouleurs (1970), p. 1211. See also Section 4.9
2 Observations and the universe

Robert W. Smith

2.1 Introduction

In the course of the middle decades of the eighteenth century, a small number of thinkers—including Immanuel Kant, Johann Heinrich Lambert, and Thomas Wright—considered the existence of what we would term galaxies beyond our galaxy, the Milky Way, as elements of speculative cosmologies in which religious concerns and ideas on extraterrestrial life were central. Kant had even suggested that some of the nebulae visible in the sky might be such distant galaxies.1

The nature of the nebulae and broad cosmological questions were of little interest for professional astronomers. As Friedrich Wilhelm Bessel, arguably the leading professional astronomer, would assert in 1832, “What astronomy must do has always been clear—it must lay down the rules for determining the motions of the heavenly bodies as they appear to us from the earth. Everything else that can be learned about the heavenly bodies, e.g., their appearance and the composition of their surfaces, is certainly not unworthy of attention; but it is not properly of astronomical interest.”2 Observational astronomy, then, as performed in professional observatories, was focused on the precise determination of the positions of astronomical objects. The telescopes and ancillary instruments in these observatories were designed for accuracy, not great light grasp. The directors of observatories, therefore, emphasized transit instruments.

William Herschel was an entirely different sort of astronomer from a typical professional. His approach to problems was more that of a natural philosopher with particular concerns in classification than a “mainstream” positional astronomer. And it was Herschel who, in the late eighteenth century, made the study of the nebulae a serious subject for research, and it was the nebulae that, it would turn out, were to be the crucial objects for observational cosmology, both in the nineteenth century and in the first decades of the twentieth. Herschel excelled too as a telescope maker, and the story of the study of the nebulae was also, we will see, very much shaped by the development of powerful telescopes.
2.2 The Herschels and nebulae

Herschel’s first career was music, but he developed a passion for astronomy that had become an obsession by the late 1770s. His life was transformed in 1781 when, in the course of a systematic hunt for double stars, he observed what he believed to be a comet. When other astronomers calculated the comet’s orbit, Herschel’s object proved to be the first major planet to be discovered in recorded history, Uranus. Uranus opened the door for Herschel to patronage from King George III of England, support that enabled him to abandon his musical career in Bath to become a full-time astronomer.

Herschel founded his astronomical career on his skills as a telescope-builder. Through his efforts, he turned the reflecting telescope from little more than a scientific toy into a serious tool of astronomical research. Although James Short had made telescopes with sizable primary mirrors, Short’s instruments were far surpassed by Herschel’s reflectors regarding size and effectiveness. With the aid of large sums from George III, William even built in 1789 a monster reflector with a 40-foot focal length and a primary mirror 48-inches in diameter. This telescope, however, was a major disappointment. Regarding light grasp and ease of operation, Herschel’s best telescope was his 20-foot reflector which had a primary mirror with a diameter of 18-inches, and which he completed in 1783.

Contrary to widespread belief, Herschel spent much of his time observing objects within our solar system. But he, along with his observing companion and long-time collaborator, his sister Caroline, devoted more attention to hunting for and examining nebulae and star-clusters than any other astronomers had ever done. Through their systematic sweeps of the skies, William and Caroline increased the total of known nebulae from the 100 or so previously identified to around 2,500.3

In 1811, Herschel explained that “A knowledge of the construction of the heavens has always been the ultimate object of my observations…” By the “construction of the heavens” he meant the arrangement of our star system (the Milky Way), as well as the arrangement and development of other star systems and nebulae. Herschel, then, was not content with discovering nebulae and star clusters and charting their positions. He also sought to classify them in the manner of a natural historian of the heavens, to employ a term he coined to describe himself, and to understand their nature.5

For Herschel, there were two crucial questions about the nebulae he sought to answer at the telescope. First, are large telescopes able to resolve nebulae into collections of stars? And, second, do the nebulae change? If the nebulae displayed changes over (astronomically) short time periods, this would argue against them being distant systems of stars. If he could answer these two questions, he could decide if the nebulae were truly clouds of luminous material, or if they were distant star systems whose milky appearance was the result of the light of the stars they contained merging together because the system was so remote that the individual stars could not be resolved.6 Herschel’s questions and issues would echo down the nineteenth century.
Herschel’s views on, and interpretations of, the nebulae shifted over time. At one stage he regarded all the nebulae as stars systems. In 1811, toward the end of his career, however, he recalled that although he had earlier

surmised nebulae to be no other than clusters of stars disguised by their very great distance...a longer experience and better acquaintance with the nature of nebulae, will not allow a general admission of such a principle, although undoubtedly a cluster of stars may assume a nebulous appearance when it is too remote for us to discern the stars of which it is composed.

Now he believed that many nebulae are single stars or comets in the process of formation. For a period Herschel viewed the Andromeda Nebula (Fig. 2.1) as a detached Milky Way. Indeed, he had calculated that its light takes about two million years to reach us. As he had reasoned in 1802:

![Fig. 2.1 Isaac Robert's photograph of the Andromeda Nebula, taken in 1888. At the time this photograph was generally regarded as showing a solar system in formation. Observational cosmology for the three decades after this photograph was taken can be understood in part as the transformation of the Andromeda Nebula into an external galaxy.](image)
A telescope with a power of penetrating into space, like my 40-feet one, has also, as it may be called, a power of penetrating into time past. To explain this, we must consider that, from the known velocity of light, it may be proved, that when we look at Sirius, the rays which enter the eye cannot have been less than 6 years and 4 1/2 months coming from that star to the observer. Hence it follows, that when we see an object of the calculated distance at which one of these very remote nebulae may still be perceived, the rays of light which convey its image to the eye, must have been more than nineteen hundred and ten thousand, that is, almost two millions of years on their way; and that, consequently, so many years ago, this object must already have had an existence in the sidereal heavens, in order to send out those rays by which we now perceive it.\(^9\)

But, by 1811, he was unsure about the Andromeda Nebula’s nature, and in that year placed it in a list “of objects of an ambiguous construction.”

The extent to which Herschel changed his position on the existence of star systems beyond the Milky Way was debated during the nineteenth century. Nevertheless, when in 1813 the poet Thomas Campbell met William and his son John on holiday in Brighton on the south coast of England, William told Campbell that

\[
\text{I have looked further into space than ever human beings did before me. I have observed stars of which the light, it can be proved, must have taken millions of years to reach the earth.}\quad 10
\]

But Herschel was also widely reckoned to have demonstrated the existence of true nebulae.

### 2.3 John Herschel and nebulae

Pierre-Simon Laplace’s nebular hypothesis—which he had first advanced at the end of his *Exposition du système du monde* of 1796 as an extended note which he elaborated on for his *Mécanique Céleste*—on the formation of the solar system relied on the existence of nebulous matter. Laplace maintained that the planets and their satellites had been born out of the shrinking outer, rotating nebulous atmosphere of the sun. Some astronomers, as we shall see later, like Laplace, looked to Herschel’s observations as proof of such material.\(^11\)

William’s fellow astronomers had admired his novel researches, but they were not inclined to go where he had led. William’s son John, however, followed his father’s example in observing and mapping the nebulae. For his observations, John exploited the 20-foot reflector, which had been William’s most successful telescope, and which he and his father had refurbished. In 1820, when he addressed the recently founded Astronomical Society of London (what would later become the Royal Astronomical Society), John Herschel claimed that one

\[
\text{of the first great steps towards an accurate knowledge of the construction of the heavens, is an acquaintance with the individual objects they present: in other words, the formation}
\]
of a complete catalogue of stars and other bodies, upon a scale infinitely more extensive than any that has yet been undertaken; and that shall comprehend the most minute objects visible in good astronomical telescopes.12

As for the nebulae, John maintained in 1826 that the

nature of nebulae, it is obvious, can never become more known to us than at present; except in two ways—either by the direct observation of changes in the form or physical condition of some one or more among them, or from the comparison of a great number, so as to establish a kind of scale or gradation from the most ambiguous, to objects of whose nature there can be no doubt.13

John extended William and Caroline’s sweeps to the southern hemisphere by transporting the 20-foot reflector to the Cape of Good Hope for a four-year stay. He also, much more so than his father, was interested in extremely detailed examinations and drawings of individual nebulae.14

John, however, was an altogether more cautious theorist than his father, and the different positions he adopted on the nebulae at different times are not easy to discern. What is clear, however, is that the four years he spent observing at the Cape of Good Hope from 1834 to 1838 were critical ones for his views on the nebulae, clusters of stars and our galactic system.

He was fascinated by the two Magellanic Clouds, and he scrutinized them with the refurbished 20-foot reflector, the first big telescope to be taken to the southern hemisphere. In presenting his results, Herschel argued that

the two Magellanic clouds, Nubecula Major and Minor, are very extraordinary objects. The greater is a congeries of stars, clusters of irregular form, globular clusters and nebulae, of various magnitudes and degrees of condensation, among which is interspersed a large portion of irresolvable nebulae, which may be, and probably is star dust, but which the powers of the twenty foot telescope show only as a general illumination of the field of view, forming a bright ground on which the other objects are scattered…

Further,

The Nubecula Major, like the Minor, consists partly of large tracts and ill-defined patches of irresolvable nebula, and of nebulosity in every stage of resolution, up to perfectly resolved stars like the Milky Way, as also of regular and irregular nebulae properly so called, of globular clusters in every stage of resolvability, and of clustering groups sufficiently insulated and condensed to come under the designation of “clusters of stars”…15

Herschel claimed that in the Large Magellanic Cloud he could see “every imaginable size of star and of nebulosity, all gathered together in the large Cloud and so at essentially the same distance from the observer…” This meant that “it was evident that no limits could be set to the variety to be found in celestial objects.”16 In this remarkable report, then,
John, as Hoskin has emphasized, broke with his father’s position that celestial objects form distinct “species” and that within each species, the variations among the objects are limited.

2.4 The nebular hypothesis, star systems, and spirals

When John Herschel returned to England from the Cape of Good Hope, his career as an astronomical observer came to an end. By this time the question of the nature of the nebulae had also become a topic of wide interest. John Pringle Nichol, the Regius Professor of Astronomy at the University of Glasgow, was also an ardent political economist and no less a figure than John Stuart Mill was one of his nominators for the post of professor of political economy at the Collège de France. Nichol also wrote on astronomy for broad audiences. His best-known book was probably *Architecture of the Heavens and Phenomena of the Solar System*, first published in 1837. It was as a result of his writings that the term “nebular hypothesis” came into widespread use, and through his *Architecture of the Heavens* the link between Laplace’s Hypothesis and William Herschel’s observations was strengthened. In this book, Nichol interpreted the different sorts of nebulae as different stages that nebulae reached on the way to contracting into groups of stars. As I have argued, following Schaffer,

Nichol and his allies fastened on the nebular hypothesis as an object of both a natural and moral science. They emphasized the stellar progress of the nebular hypothesis which they then exploited as a general model of universal progress, one that could be pressed into service to argue for political reform. Nichol, as the author of *Vestiges of the Natural History of Creation* was to do later, also wove together the nebular hypothesis, geological theorizing and speculations on the transformation of species.17

The now infamous *Vestiges of the Natural History of Creation* was published in 1844, and its anonymous author (who in fact was Robert Chambers, a Scottish publisher, writer, and political radical) provoked an enormous controversy. The book’s large number of readers even included Queen Victoria. Chambers explained the development of the Earth, the emergence of life upon it, and, in time, the appearance of human beings. Chambers contended, however, that the “nebulous matter of space, previously to the formation of stellar and planetary bodies, must have been a universal Fire Mist, an idea which we can scarcely comprehend, though the reasons for arriving at it seem irresistible.”18

As Chambers’ volume reached bookshops, the third Earl of Rosse, William Parsons, was constructing a remarkable telescope at Parsonstown in central Ireland. Observations made with Rosse’s 72-inch reflector would take the debate on nebulae in a new direction. William Herschel’s largest instrument was his 40-foot reflector, at the heart of which was a 48-inch primary mirror. Until the 1860s, telescope mirrors were generally made of speculum metal, an alloy of copper and tin. With the casting, grinding and polishing of his 48-inch speculum metal mirrors, Herschel had confronted severe problems. He