



Carlos Martins

The Universe Today

Our Current Understanding
and How It Was Achieved



Springer

Astronomers' Universe

Series Editor

Martin Beech, Campion College, The University of Regina,
Regina, SK, Canada

The Astronomers' Universe series attracts scientifically curious readers with a passion for astronomy and its related fields. In this series, you will venture beyond the basics to gain a deeper understanding of the cosmos—all from the comfort of your chair.

Our books cover any and all topics related to the scientific study of the Universe and our place in it, exploring discoveries and theories in areas ranging from cosmology and astrophysics to planetary science and astrobiology.

This series bridges the gap between very basic popular science books and higher-level textbooks, providing rigorous, yet digestible forays for the intrepid lay reader. It goes beyond a beginner's level, introducing you to more complex concepts that will expand your knowledge of the cosmos. The books are written in a didactic and descriptive style, including basic mathematics where necessary.

More information about this series at <http://www.springer.com/series/6960>

Carlos Martins

The Universe Today

Our Current Understanding and How It
Was Achieved

 Springer

Carlos Martins
Centro de Astrofísica da Universidade do Porto
Porto, Portugal

ISSN 1614-659X ISSN 2197-6651 (electronic)
Astronomers' Universe
ISBN 978-3-030-49631-9 ISBN 978-3-030-49632-6 (eBook)
<https://doi.org/10.1007/978-3-030-49632-6>

© Springer Nature Switzerland AG 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

You do not receive an education any more than you receive a meal. You seek it, order or prepare it, and assimilate and digest it for yourself.

Frank Rhodes (1926–2020)

Over the past 10 years, I have had the pleasure of giving various different short courses (between 10 and 25 contact hours each, depending on the case) to three types of audiences. This book grew out of the contents of these lectures, with the encouragement of many people who attended them.

The first group are high school teachers, in the context of refresher courses for current teachers regularly organised by U. Porto and by ESERO's¹ partner in Portugal, Ciência Viva, and mainly taught at CAUP. Periodically taking some such courses is mandatory for high school teachers to progress in their careers, and naturally, the courses are broadly aligned with the goals of the national physics and chemistry school curriculum. (Unfortunately, Portugal has no national astronomy curriculum!) These are formal courses, including a final assessment (typically a written exam), and are validated by a national pedagogical committee before being offered. About 70% of the teachers who take them are physics and chemistry teachers, while the others are mathematics and biology/geology teachers.

The second group are U. Porto students (about two-thirds of them studying science or engineering and the rest from all other areas), taking complementary

¹<https://www.esero.pt/>.

courses offered by the CAUP Training Unit. Most of these are undergraduate students, with occasional M.Sc. or Ph.D. students. The students can take these courses at any point of their studies, on a voluntary basis: a final written exam is not compulsory, although it can be credited if the students choose to take the exam and pass (typically about one-third of the students take the exam, and almost all those who take it pass).

The third group are bright high school students (in the last 3 years before university, roughly 15–18 year olds), mostly in the context of an astrophysics summer school which I created in 2012 and have been organising every year since then at the Corno de Bico Protected Landscape Area² (in the Paredes de Coura municipality, in the Northwest of Portugal), the AstroCamp.³ Initially, this was only aimed at Portuguese students but has now grown into an international school, accepting applications from 42 eligible countries and supported by international partners like ESO in addition to several national partners.

These courses are given to relatively small groups (in Portuguese and to a maximum of 40 students in the first two cases or in English and to a maximum of 20 in the third case), allowing for detailed interaction with the students and for instant feedback from them. They are mostly given using slides, supplemented by the blackboard whenever needed. Occasionally, I have also used various parts of this material as stand-alone popular talks, mostly when visiting high schools, for example, during World Space Week (in this case, the groups are larger, typically 50–120 students). About half of this material is on modern astrophysics (physics of stars, relativity, the standard Big Bang model, physics beyond the standard model, etc.), while the other half is on the history of astronomy (and, to a lesser extent, that of physics).

After many queries from students on where they could find additional information on the various topics discussed, the time has come to further organise some of these materials into a book.

My aim is to give the reader an overview of our current view of the Universe, of how we gradually developed it, and of how outcomes of current research—both my own and by others—might still change this view. For this, one must bring together concepts in physics and astronomy, including some of the history of both of them. The historical part may seem of lesser interest to the average reader, but it is important to understand where we are and how things may develop in the future. In fact, I strongly believe that anyone taking

²<http://www.cornodebico.pt/portal/>.

³<http://www.astro.up.pt/astrocamp/>.

physics or astronomy today and considering a future professional career in this field should be exposed to the history (and philosophy) of these subjects—and naturally the same applies to any other scientific subject. It is the responsibility of universities to provide such possibilities in the degrees they offer.

The level of the book is not technical. Although I do not particularly like the word, it could be called ‘descriptive’, in the sense that the goal is always to highlight the crucial physical concepts. I often emphasise to students that if one understands the key ideas in a given context, one can always work out the maths later on (if and when it is needed), while trying to start with the maths without understanding the physics is far harder. Therefore, there will be very little explicit maths, and most of it will be concentrated into a few sections. That said, the level of the book will not be as simple as that of typical popular science books: the aim is to present things rigorously, and although I will simplify many things I did my best to avoid oversimplifying.

My guess is that the ‘average’ readers should be undergraduate students (not necessarily studying astronomy or physics, but still with a broad interest in these areas and/or taking an introductory course on it). The only prerequisite is a qualitative knowledge of basic physics concepts, at the high school physics level. The book (or some chapters thereof) could therefore be useful for various undergraduate introductory astronomy and physics classes. This level should also make the book relevant for high school teachers who may need to teach some parts of this material or simply acquire some background knowledge with which to answer questions from the more interested and curious of their students. (At least in the Portuguese school system, some astronomy topics are actually taught by biology/geology teachers, who sometimes have a limited knowledge of astronomy or physics.) For the same reason, the book should be appealing to bright high school students wanting to learn more physics and astronomy.

The book aims to put together materials from several different short courses but broadly includes one part on the history of astronomy (and to a lesser extent of physics) and another one on modern astrophysics and cosmology. The early chapters are not purely ‘historical’, as they could have been written by a historian of astronomy or physics. For the record, I am a working scientist, not a historian, and there are many aspects of the history of science which historians find fascinating that are of very little interest to me—naturally, the opposite is also very probably true. The point of delving into the history of the subject is that of introducing, in a historical context, concepts that are important for our view of the Universe today. By discussing how they emerged and subsequently evolved, one can more easily understand their modern relevance.

As a convenient way to summarise some of the concepts discussed in the book, I provide a short multiple-choice quiz, together with some suggested points to think about. The reader can think of the latter as invitations to write short essays on these topics, as a means to further consolidate knowledge. The answers to the multiple-choice quiz can be found online. Some of these questions have been used previously in the exams I set for students taking the course. Finally, some suggestions for further reading are also provided in an overall bibliography.

Porto, Portugal
March 2020

Carlos Martins

Acknowledgements

As mentioned in the preface, this book grew out of material from various courses I have given in the last decade or so. I am grateful to the many students in these courses, whose questions and comments gradually allowed me to develop the material in relevant and interesting directions and pitch it at an appropriate level for the various intended audiences.

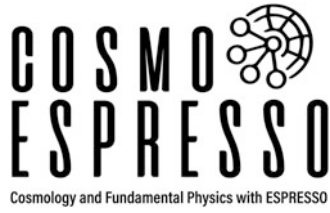
I am grateful to Paulo Maurício, with whom I co-taught many editions of the History of the Universe course, for many interesting discussions about the historical aspects, and to my students, especially Ana Catarina Leite and José Ricardo Correia, for discussions on the more modern aspects.

Many thanks also to Francesco Chiti, Liliana Sousa, Marlene Körner, Paloma Thevenet, Pedro Amaral, and Siri Berge, all former AstroCamp students, who have read several chapters and provided useful comments and suggestions.

Last but not least, I am grateful to Angela Lahee and the Springer team for their patience during the completion of this project and the many helpful suggestions during the various stages of its development.

This work was partially financed by FEDER—Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020—Operacional Programme for Competitiveness and Internationalisation (POCI), with grant number POCI-01-0145-FEDER-028987, and by Portuguese funds through

FCT—Fundação para a Ciência e a Tecnologia in the framework of the project PTDC/FIS-AST/28987/2017 (CosmoESPRESSO).



Contents

1	Introduction	1
1.1	The Scientific Method	1
1.2	The Dawn of Astronomy	9
1.2.1	Ancient Egypt	12
1.2.2	Mesopotamia	15
1.3	Clocks and Calendars	21
1.4	Science in the Modern World	26
2	Classical Astronomy	33
2.1	The Greek World View	33
2.2	Cosmology	36
2.3	The Size of the Earth	46
2.4	Astronomy	48
2.5	Epicycles	56
2.6	Aristotelian Physics	62
2.7	From Greece to the Islamic World	66
3	From the Renaissance to Modern Science	73
3.1	The Medieval World View	73
3.2	Medieval Physics	79
3.3	Mikolaj Kopernigk	83
3.4	Tyge Brahe	92
3.5	Johannes Kepler	98
3.6	Galileo Galilei	105
3.6.1	Galilean Physics	107

3.6.2	Galilean Astronomy	110
3.6.3	Physics Versus Philosophy	114
4	Classical Physics	121
4.1	Pre-Newtonian Gravity	121
4.2	Newton's Laws	123
4.3	Uranus and Neptune	129
4.4	Mercury and Vulcan	144
4.5	The Physics of Gravity and Satellites	149
4.6	Asteroids and Comets	160
5	The Physics of Relativity	167
5.1	The Decline of Classical Physics	167
5.2	Special Relativity	176
5.3	The Equivalence Principle and General Relativity	183
5.4	Gravity Is Geometry	188
5.5	Consequences and Tests	190
6	The Physics of Stars and Galaxies	197
6.1	The Life of a Star	198
6.2	Stellar Nucleosynthesis	206
6.3	White Dwarfs, Neutron Stars, and Black Holes	209
6.4	Galaxies and the Expanding Universe	218
7	The Physics of the Universe	227
7.1	The Hot Big Bang Model	227
7.2	Key Observational Tests	234
7.3	The Dark Side of the Universe	240
7.4	Scalar Fields and Extra Dimensions	248
8	New Frontiers	259
8.1	Topological Defects and Cosmic Palaeontology	259
8.2	Are the Laws of Physics Universal?	268
8.3	The Dawn of Real-Time Cosmology	279
8.4	The Quantum World	282
8.5	The Search for Life in the Universe	286

	Contents	xiii
A	Final Exam	297
A.1	Multiple Choice Questions	297
A.2	Essay Topics	309
B	Further Reading	313
	Index	317

Acronyms

BCE	Before the Common Era
CAUP	Center for Astrophysics of the University of Porto
DNA	Deoxyribonucleic acid
EEP	Einstein equivalence principle
ELT	Extremely Large Telescope
EPR	Einstein–Podolsky–Rosen
ESA	European Space Agency
ESERO	European Space Education Resource Office
ESO	European Southern Observatory
GPS	Global Positioning System
GUT	Grand Unified Theory
HEO	High Earth orbit
HST	Hubble Space Telescope
ISS	International Space Station
LEO	Low Earth orbit
LHC	Large Hadron Collider
LIGO	Laser Interferometer Gravitational-Wave Observatory
MACHO	Massive compact halo object
MEO	Medium Earth orbit
NASA	National Aeronautics and Space Administration
NEO	Near Earth object
NOAA	National Oceanic and Atmospheric Administration
RAS	Royal Astronomical Society
VLT	Very Large Telescope
WIMP	Weakly interacting massive particle



Introduction

We will start with a brief discussion of the scientific method, highlighting what distinguishes science from other human endeavours and reflecting on how what we now call science originated historically. We then take a more detailed look at these origins in the specific case of astronomy, in Egypt and Babylon, also mentioning the development of tools such as calendars and clocks. Finally, we reflect on the importance of scientific literacy (and its mathematical sibling, numeracy) in the modern world, and briefly mention astrology in this context.

1.1 The Scientific Method

Science is a way of trying not to fool yourself. [...] The first principle is that you must not fool yourself, and you are the easiest person to fool.

Richard Feynman (1918–1988)

One of the more noteworthy and alarming paradoxes of our modern civilisation is that the more our everyday lives rely on science and technology, the less the common person knows about them. If you're not sure about what I mean, I suggest the following exercise. Try to spend 24 h of your life using only the technologies you understand—by 'understand' I mean that you can explain how they work, in a simplified but otherwise accurate way, to a teenager.

For example, if you don't understand how your mobile phone works, and how it captures your voice and image and sends them to the nearest tower, and from there ultimately to the person you're talking to, then don't use a mobile phone. If you don't understand how the engine of your car (or bus, or train) works, then don't use them during that day. If you don't understand how an airplane flies, then don't fly. And we could go on to computers, digital cameras, microwave ovens, and so on.

Notice that I'm not talking about building these tools—they are all tools to enhance our own capabilities, one way or the other. I'm simply talking about understanding, at a non-technical level, what basic principles (of physics, chemistry, biology, maths) enable them to do for you what they are supposed to do. So whenever you reach out to grab one of these tools, pause for a second and ask yourself whether you understand how it works. I'm sure that you will have a very interesting 24 h.

This highlights the importance of science in technology in your daily life. And since it is clearly affecting your life, it should also make you think about what this thing is that we call science. Certainly, you will have read or heard many news stories saying that scientists have just discovered or announced something new. That could be something that is easy to grasp (possibly about a new species of mammal in the Amazon forest), but it could also be something completely outside your everyday experience, for which you have no previous experience, even if it makes you curious—maybe something about the big bang or black holes?

So how do scientists do what they do, and arrive at their conclusions on such topics? And, more importantly, why should you know about these things, and believe the things that scientists say? Clearly, you're not going to go to the Amazon rainforest to check that the newly discovered mammal is really there, or check some mathematical calculations, or a lab experiment, or a computer simulation, so either you simply trust the claim, or you have to hope that someone else will check and confirm (or refute) such claims for you.

Answering the second question is easy enough. Most of the challenges and threats that we will all be facing in the coming decades, and which are hotly debated today have a science component at their core (as a partial list, consider global warming, terrorism, pandemics, and 'alternative medicines'). Clearly, they are not purely scientific problems—they involve economic, political, ethical, and moral aspects, among others—but they do involve science at their core. And knowing the science is crucial for anyone to participate in the corresponding debate. If you don't know the underlying science, you may simply ignore or exclude yourself from the debate, and you will be easily manipulated, either by people who do know more than you and want you

to shift your opinion in a particular direction for their own purposes, or even by people that know less than you (although they may think that they know more).

Most cultures articulate their world picture, and systematically transmit it to the next generation, through mythology or religion. In the beginning of the seventeenth century, the Western world decided to do something different, by articulating our view of the world through science, and particularly through astronomy and physics. Nevertheless, science has always been confined to a small number of specialists, and despite four centuries of development of our education systems most people receive a comparatively poor exposure to science, either as children or later on as adults.

In this sense, our society hasn't been doing particularly well, and indeed the development of seemingly key ingredients has been rather slow. To take a simple example, the term 'scientist' was only coined fairly recently: it was first used in print by William Whewell in 1834, in his review of Mary Somerville's book 'On the Connexion of the Physical Sciences'. (As an aside, quite a few terms we now use in physics and chemistry—such as ion, electrode, cathode, or anode—were suggested by Whewell to Michael Faraday.) Also, science as a profession, in the sense that a restricted number of people are extensively trained to be able to carry out this activity (and eventually be paid for doing it, as well as training the next generation) is only a late nineteenth century development.

How did science emerge? We will look into the particular case of astronomy shortly, but for the moment let's note that, from a historical perspective, societies before ours had two main motivations that eventually led to what we now call science. The first one is abstract and conceptual, and in its proper historical context it is essentially theological—to behold God's plan for the Universe. The second one is practical. If you are a farmer and your survival in the coming months depends on producing enough food to get through the next winter you will be highly motivated to understand the cycle of the seasons so that you can sow and harvest at the appropriate times. You thus need to develop an accurate calendar, leading to astronomy. Similarly, the need for reliable maps, methods of navigation, and all sorts of tools which enhance the natural capabilities of the human body are starting points for various other branches of physical science.

As a specific example, think for a second about the night sky. For early societies this provided a calendar for farmers and a map for sailors, but also a home for the gods and a repository for countless mythological stories for everyone to tell and remember. Indeed, the night sky was like a primitive society's television—or, in more modern parlance, YouTube. What else could

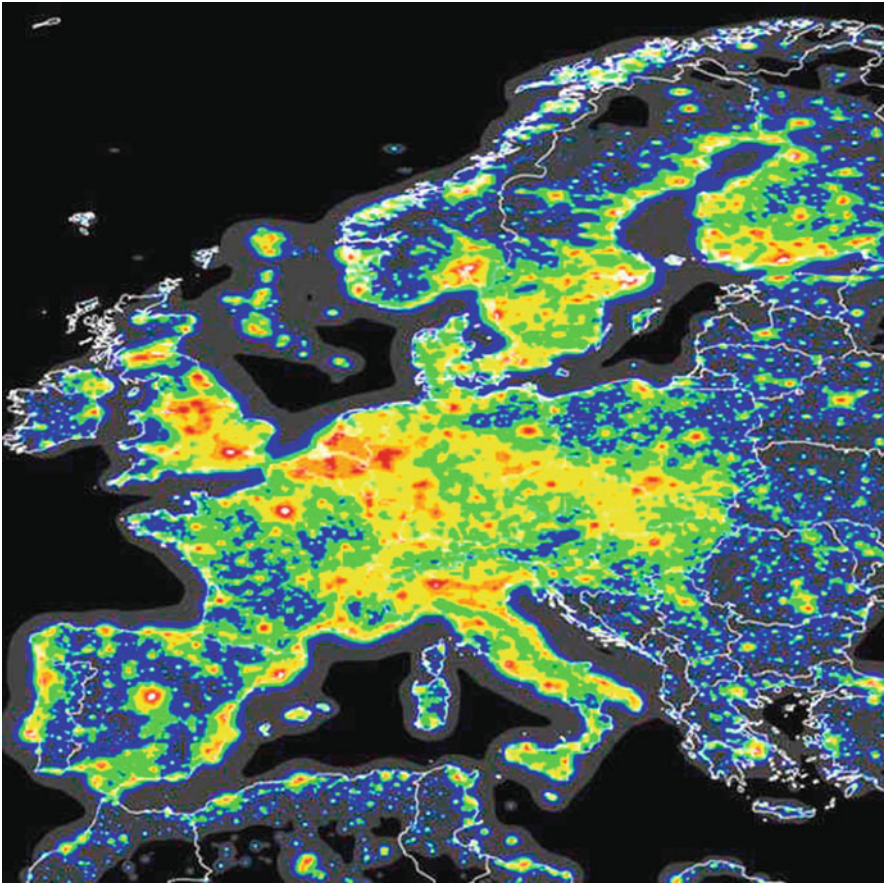


Fig. 1.1 False colour map showing the intensity of skyglow from artificial light sources. Credit: P. Cinzano, F. Falchi (University of Padova), C. D. Elvidge (NOAA National Geophysical Data Center, Boulder). Copyright Royal Astronomical Society. Reproduced from the Monthly Notices of the RAS by permission of Blackwell Science

you look at during those long nights? On the other hand, in our modern societies very few people can see a truly dark sky, free from light pollution (this is in fact impossible in Europe, see Fig. 1.1), and even comparatively simple things like seeing the Milky Way or the Andromeda galaxy with the naked eye can be a challenge. Very few people have had the opportunity to see a sky as dark as the ones Galileo saw just four centuries ago. This is particularly unfortunate since the night sky is the only part of our environment (and cultural heritage) that is common to everyone—all human beings, no matter when or where they live(d) see basically the same night sky.

It is interesting that most societies prior to our own can be neatly classified into one of the two sides of this conceptual/practical divide. The canonical examples are Classical Greece and the Roman Empire. In the former the prevailing view was that the world should be understood but not changed, and therefore any knowledge which had practical applications was deemed inferior to purely abstract knowledge. On the other hand the Romans were an eminently practical civilisation and adopted whatever worked, without ever worrying (or even thinking) about why it worked.

What is peculiar about our society is that we have somehow optimally combined the two motivations in the early seventeenth century and this led to the development of modern science. On the one hand, several developments in the previous century (by Copernicus, Tycho, and many others, which we will describe in the coming chapters) required the development of new theoretical paradigms in astronomy and physics. On the other hand, new practical tools such as the telescope and the microscope appeared at precisely this time, allowing the testing and further development of these paradigms.

How, then, does one do science? The starting point in the scientific method is a belief in the objective validity of science. This includes three different aspects that one must accept (or, perhaps more accurately, assume). Firstly, that Nature really does exist outside of and independently of us. To put it simply, you should accept that the Universe existed before you were born, and will continue to do so after you die. Secondly, that there is some set of laws of Nature which are objectively valid, without regard for our preferences or expectations. And thirdly, that we can progressively discover and understand these laws. Note that the third is conceptually different from the second: it could be the case that such laws exist but are entirely beyond our reach. These are the assumptions that every scientist is—at least implicitly—making. Science itself cannot prove the correctness of these assumptions, but as one proceeds one can gather supporting evidence for them. Historically, one could say that these assumptions were first made in a systematic way in Classical Greece around 2500 years ago, and more clearly reasserted in early seventeenth century Europe.

The scientific method is an iterative way of generating consistent knowledge about how the Universe works, gradually identifying old ideas which prove inadequate and replacing them by new ones, on the basis of observations of and experiments in the real world. One starts by observing a particular aspect of the Universe that is of interest, and formulates a starting hypothesis which is consistent with these observations. This hypothesis is then used to make further predictions, which are in turn tested by further experiments or observations. The process is then iterated until there are no noticeable

discrepancies between the hypothesis (and the underlying theory, if it already exists) and the experiments. Once this is successfully achieved, the hypothesis is validated and accepted as a new theory, or added to an existing one. By this process our knowledge about the physical world gradually grows and we acquire a deeper and more accurate understanding of the aforementioned set of laws of Nature.

An interesting question is what is the ultimate source of scientific knowledge. Two different answers are provided by rationalists (who emphasise reason and intuition) and empiricists (who emphasise experience and observation). Examples from the two camps are René Descartes (1596–1650) and David Hume (1711–1776), respectively, but this division is prevalent throughout the history of science itself, for example in Plato versus Aristotle or in Newton versus Galileo. Einstein is interesting in this regard, because he developed his special and general theories of relativity in opposite ways.

There are several different but inter-related concepts here that are worth distinguishing. A law is a scientific hypothesis for which there is an ample collection of experimental and/or observational evidence. A theory is an underlying conceptual framework which is able to explain a set of experimental and observational results and the corresponding laws, and which additionally predicts the results of new observations and experiments that can be done subsequently. A theory with a limited range of applicability, which is clearly perceived to be a first approximation, or which still lacks extensive testing, is sometimes called a model.

A first important aspect of science is its iterative (one could say trial and error) nature. One thing that the history of science teaches us (and which is crucial for prospective scientists to be aware of) is that in many circumstances scientific progress is the direct result of the realisation that we were asking the wrong question. When this happens one can sidestep the original question by tackling a different but related one, and not uncommonly one also finds that the first question was actually irrelevant. We will see examples of this later in the book. In some sense one could say that what distinguishes science from other human endeavours is not what it allows us to know and how we deal with that knowledge, but how we confront what we still don't know.

A second important aspect of science is an asymmetry between the confirmation and the refutation of a theory or hypothesis. Refutation is always a possibility and when it happens is logically certain, but there is no logically valid way of proving (in the mathematical sense of the word) the truth of a theory from the agreement of its predictions with any finite number of observations or experiments. In other words, no hypothesis can be proved absolutely true (that would require testing it in all possible ways under all

possible circumstances), but there is always the possibility that it can be proved false, if one of its predictions is not verified by a particular experiment.

We can never be certain that a theory is correct, only that refuted theories are incorrect. So effectively, what one has are degrees of confidence in the validity of each theory, which increase with every new observation or experiment consistent with them and drop to zero if one experiment refutes them. For example, being at mid-latitudes in the northern hemisphere I am extremely confident that the Sun will rise towards the east tomorrow morning (and indeed I can easily calculate—or look up online—the time and direction, in Porto for example, for the event), but I cannot prove that this will happen. Maybe the Sun will be destroyed overnight by Vogons working on a hyperspace bypass, in which case it won't rise tomorrow morning. (If you don't know who the Vogons are and why they are building hyperspace bypasses, you are not reading enough.)

It follows from the previous paragraphs that to be scientifically useful, hypotheses must be falsifiable. Any and all scientific theories are, by their very nature, in constant danger of being proved wrong by new data or observations. This is a crucial positive aspect of scientific research: it ultimately provides the means for constantly improving our knowledge about the Universe. Therefore scientific truths are always qualified, and never absolute. New discoveries that change our view of the Universe can occur at any point, either by falsifying previously held theories or hypotheses, or by setting limits on their domains of applicability, and thus highlighting the need for more encompassing ones.

To put it in a different way, the distinguishing feature of a scientific theory is not that it can be verified but rather that it can be falsified: it must be able to make further predictions, beyond the observations that led to its development, that can be subjected to further testing. Only vulnerable hypotheses and theories that include this element of risk, in the sense that their predictions must be specific enough to be incompatible with some possible results of the further tests, can be counted as supporting evidence for the theory.

From this we also see that a further important aspect of the scientific method is measurability: the theory's predictions must be specific enough to be quantitatively measured, even if at a particular point in time the available technology is insufficient to reach the accuracy necessary to measure the predicted effects. And another important aspect is reproducibility: the predictions made by a theory should apply to all the phenomena and circumstances it claims to describe (one can't have a theory that works on Mondays but not on Saturdays).

Reading the above you may reflect on how it correlates with the public perception of what science is, and in particular on how scientists are portrayed in everyday life, in particular by the media. It is worthy of note that no scientist

(or, at least, no scientist worthy of that name) has or even claims to have the monopoly over truth. However, what scientists can do better than anyone else is to find out when hypotheses or theories are wrong and simply can't work. Indeed this is what scientists are specifically trained for, and most scientists, throughout most of their careers, spend their time excluding hypotheses and showing that many of them can't work. It is only relatively rarely that a scientist will discover something genuinely new about the way the Universe works.

That said, there are two caveats to bear in mind. The first caveat is that one must not reverse the burden of proof: anyone making new claims must also supply evidence in support of those claims. If you believe that Santa Claus does exist, the rest of the scientific community is not obliged to explicitly test and refute your claim: rather, it is your task to show what evidence you have supporting that claim. And the second caveat is that the fact that we do not know how something might work does not prevent us from finding out whether or not it works. For an example of this, have you ever had surgery which required you to be given general anesthesia? If you have, you may reflect on the fact that the number of people on the planet who understand how general anesthesia works, at the basic biochemical and neurophysiological level, is exactly zero. Local anesthesia is very simple to understand, but general anesthesia involves the brain and central nervous system, and understanding how it works at a fundamental level is a far more complex task. And yet, nobody doubts that it does work, and it has been routinely used across the world for decades, with a vast set of empirical data on the appropriate dose for the circumstances of each patient.

Finally, it is worth remembering that scientists are also human beings, and have their own preferences and biases which affect the way science is done. If a scientist has been working with a previously successful model or theory, it is often the case that data alone will not be sufficient to force him or her to completely abandon it and start again from scratch. Instead, such theories are often modified or reinterpreted in the light of the new developments. An example of this is the relation between Newtonian physics and General Relativity, to which we will come later in the book.

Max Planck famously said that a new scientific theory does not replace an older one by convincing the supporters of the latter that the new one is conceptually better, or simpler, or more accurate, but rather because the older generation that had been trained in the old one eventually dies and the new generation that replaces it is now familiar with and accepts the new theory. (Or, to put it more succinctly, science progresses with every funeral.) When they are faced with a choice between competing alternative theories, scientists often appeal (implicitly or explicitly) to further selection criteria, in addition

to the results of experiments or observations. Examples of these include beauty, symmetry, or economy of explanation. Naturally these are, to a large extent, aesthetic concepts, which cannot be quantified and even lack a generally agreed definition. But these factors do enter scientific debates.

Could science disappear? In other words, given our modern reliance on science and technology and our choice of science as the lingua franca to articulate our view of the world, is this necessarily a continuous and irreversible process, or could it be stopped and reversed by the forces of irrationality? The simple answer is that history demonstrates that this has happened several times in the past, in other societies, from Classical Greece to Medieval China. And in our modern society science is threatened in many ways.

A scientifically and technologically advanced society comes without any guarantee that irrational thought will disappear. A particularly worrying trend is the fact that ‘fringe phenomena’ are widely spread by the media, indeed, more so than science itself. Think of astrology, creationism and intelligent design, global warming denial, anti-vaccination activists, and a whole slew of the so-called alternative medicines. It is clear that our society has a deep problem of scientific illiteracy, which is exploited by those wanting to spread scientific misinformation for ideological reasons.

The scientific community must do its part to address this problem. There are two components to it: one is making science accessible to the general public, and the other is explaining the process by which science is done. And academic institutions must provide greater incentive and facilities for their researchers to do so. Those working with younger students, and starting the process of training the next generation of citizens (and of future scientists), have a crucial role to play, too. As long as our society continues to report reality through science, and to rely on it to improve our daily lives, there’s a society-wide obligation to make science accessible to everyone. What is at stake is not just individual sanity and critical thinking skills, but ultimately social cohesion.

1.2 The Dawn of Astronomy

The first indications of a desire to understand the world around us are provided by the mythology of each society. Indeed, these myths typically include an account of how gods or other beings created the world. Such myths can therefore be said to be the deep roots of modern astronomy and cosmology. But the myths also have practical consequences, and shape the way in which each society organises itself.

Nevertheless, having a set of anthropomorphic deities capable of interfering in human affairs has a huge drawback: it necessarily leads to a capricious world, in which one cannot hope to make reliable predictions about future events, because divine intervention can occur unpredictably at any point.

Thus one key step in the development of science is to overcome the innate tendency of interpreting natural phenomena as personified and divinised. Strictly speaking, such a step was, as far as we know, only decisively taken in Classical Greece. We will discuss this in the next chapter. Here we will go further back in time, to discuss the origins of observational astronomy.

Sedentarization and the ensuing development of agriculture provided a key incentive to make careful observations of the Sun, the Moon, and also the planets and stars in the night sky, for example to track the passing of the seasons in order to determine the best times to sow and harvest. And inevitably, such observations would lead to the discovery of the regularity of some astronomical phenomena. In fact, this is even attested by prehistoric monuments such as the Almendres Cromlech, Stonehenge (see Fig. 1.2), Newgrange, and many others. Although those who built them left no written records, it is clear that the building of such monuments, which obviously required a substantial multi-generation construction effort, is witness to an existing belief in the regularity of certain astronomical phenomena, such as the solstices.

The fact that astronomy arose in Middle-Eastern civilisations is not an historical accident. The various civilisations that flourished in this part of the world differed in several ways (some of which will be discussed in what follows), but they shared at least four common factors which made them ideally placed for the development of astronomy.

The first and most obvious one is that this was (and still is) an area of the planet where the sky is often clear, so making frequent observations is not particularly difficult. Secondly, sedentarization and the concomitant specialization of different members of the society led to elites who had enough free time (either ex-officio or simply by their own personal choice) to undertake this systematic study of the heavens. Thirdly, they had written languages, which enabled them to record and conserve their observations over very long periods (as opposed to having to rely on the fallible memory of individuals) and thus gradually accumulate an extensive set of data. Fourth, they had considerable mathematical knowledge and enough tools (today we would call them algorithms) to allow them to look into the accumulated data, notice important patterns or regularities, and thereby make practical use of them to try to predict future motions of the celestial objects.



Fig. 1.2 The Almendres Cromlech (Évora, Portugal) and Stonehenge (Wiltshire, England) (Public domain images)

In what follows we will discuss how different local factors shaped the development of astronomy (and the corresponding early cosmogonies) in two adjacent but contrasting regions, Egypt and Mesopotamia. From here this knowledge gradually spread to Greece and thence to the whole Mediterranean, which one can arguably consider to be the first civilisation.

1.2.1 Ancient Egypt

In Egypt, the regular patterns of the heavenly bodies are seemingly reproduced by the regular cycle of the Nile, with its annual flooding from July to October. Indeed, Herodotus called Egypt ‘the gift of the Nile’. In this sense life in Egypt didn’t visibly change, and therefore the ancient Egyptians thought that the world was static and unchanging. In this context, time was logically thought of as cyclic, consisting of a succession of eternally repeating phases. There was therefore little sense of past and future, evolution, or history.

Egyptian mythology asserts that in the beginning there was Nun, a ‘non-being’ which had the potential for life but was not alive as such. Then Nun gave life to Atum, a ‘complete being’ and Lord of the Universe. Atum manifested himself as Ra, the Sun god: a radiant dawn that filled space with the light of life. Then Atum generated the first pair of gods, Shu (male, and representing air or light) and Tefnut (female, and representing moisture). In turn this couple begot Geb, the Earth god, and Nut, the sky goddess, and finally these had four children: Osiris, Isis, Seth, and Nephthys. The tenth god was Horus, a heavenly divinity usually represented with the features of a falcon, who was the son of Osiris and his heir to the kingdom on Earth. An important point is that the gods created the world with everything in it already organised in a regular, permanent pattern. To use more modern terms, this is a static world.

When it comes to the physical universe, ancient Egyptians described the sky as a roof placed over the world, which was supported by four columns placed at the four cardinal points. The Earth’s shape was a flat rectangle, longer from north to south than from east to west, and having (rather unsurprisingly) the Nile as its centre. Towards the South there was another river, this one in the sky, supported by mountains. It was on this river that the Sun god made his daily trip. Every evening the goddess Nut swallowed the sun; it then travelled through her body during the night, and she gave birth to it again every morning. However, this unchanging cosmic balance, with its regular and indeed predictable recurrence of the seasonal phenomena, did not occur spontaneously: its stability could only be ensured by a permanent and deliberate control. On Earth this was the function of the pharaoh: his main role was to ensure that the Sun would rise in the east and set in the west every day.

This idea of a stable and regularly repeating pattern of events, which was manifest in everyday life, naturally led to sense of security from the risk of change and decay. Thus there was no motivation for creativity or progress, which today is manifest in the fact that Egyptian art changed relatively little

over a period of 3000 years. Another manifestation of this frame of mind is that the years were not counted in a linear sequence. Instead, they were counted with reference to the reign of each particular pharaoh: the counting would be reset to one when each pharaoh took over the throne. Today we have long (indeed almost complete) royal lists, but a lack of precisely dated events because it is difficult to match this list to a standard calendar. This is manifest in the surviving works of the historian Manetho—whose work actually dates from the Ptolemaic Kingdom of Egypt.

On the practical side, the Egyptians did make a key contribution to horology, the science of calendars and time measurements. They created a civil calendar which consisted of 12 months, each containing 30 days, and, clearly noticing that such a 360 day cycle did not stay aligned with the seasons and the Nile flooding cycle, supplemented it with five additional days at the end (known as the ‘five days beyond the year’), making a total of 365. Historical evidence shows that this was in use as early as 2800 BCE. This is remarkable in being the first non-astronomical calendar.

This civil calendar had a purely empirical origin, built up by regularly observing, recording, and in the end averaging the time intervals between successive arrivals of the Nile flood. The year was divided into three seasons of 4 months each: Akhet (the Flood Season), Peret (the Growth Season) and Shemu (the Dry Season). Similarly, each month was divided into three ‘weeks’ of 10 days each, known as decans (which is actually a later Greek term)—a choice possibly related to the fact that there were ten main gods in the Egyptian pantheon.

To each decan period was associated a star or group of stars (themselves known as decans) which would rise in the east at dawn, just before the Sun itself (and after a period when they were invisible, being hidden behind the Sun). Such stars are said to be in heliacal rising. Just as the position of the Sun in the sky can be used to identify the time of day, the stars that are seen in heliacal rising on a given day (or, analogously, heliacal setting) can be used to identify the days of the year.

It is thought that, before this calendar, there was an earlier lunar-based calendar, with months beginning on the first day on which the old crescent was no longer visible in the east at dawn. This would be commensurate with the fact that their days ran from sunrise to sunrise—a convention that lasted until Hellenistic times.

Initially, the Egyptians did not realise that the astronomical year, and thus the cycle of the seasons, does not consist of exactly 365 days, but is in fact slightly longer, by almost 6 h. This difference was eventually recognised and another calendar was then introduced (probably around 2773 BCE) to track

astronomical phenomena more closely. The crucial step is thought to have been the realisation that the rising of the Nile coincided with the heliacal rising of the star known as Sopdet to the Egyptians, Sothis to the Greeks, and Sirius to us—that is, the brightest star in the night sky. This fortunate coincidence was surely seen as a meaningful omen and provided the natural beginning of the Egyptian year in the ‘Sothic’ calendar. The Sothic calendar kept pace with the seasons, while the civil calendar did not; the two move steadily apart, and only coincide at intervals of 1460 years.

The low levels of cloud cover not only enabled easy observation of the night sky but also made the Sun a convenient clock. Thus the fact that the earliest known solar clock comes from Egypt is not particularly surprising. It has been dated to about 1500 BCE. It is also known that the pharaoh Tuthmosis III (ca. 1450 BCE) referred to the hour indicated by the Sun’s shadow at a particularly important point of one of his military campaigns in Asia, which indicates that portable solar clocks were also in use.

In order to measure time at night (or in general, when the Sun was not available), the Egyptians also invented the water clock, which we now know as the ‘clepsydra’, as the Greeks later called it. Both the obvious main types are known to have been developed and used, with water flowing out of or into a graduated vessel, and clepsydrae were also used by the Greeks and Romans. Finally, the Egyptians used a third instrument to observe the transits of the relevant stars across the meridian. This was a set of two plumb lines, which they called the ‘Merkhet’. The principle behind this is the same as that of the heliacal risings: on a given night different stars (specifically, different decans) transit the meridian at different times, and for each decan the transit time varies according to the day of the year.

It is also worthy of note that our modern division of the day into 24 h ultimately stems from Egypt. That said, a subtle but crucial difference is that these hours were not of equal length. Instead, at all times of the year the periods of daylight and darkness were each separately divided into a period of 12 h. The end of the night was determined by the heliacal rising of the appropriate decan, as has already been explained. These initial divisions of the day and night into separate periods of 12 h were subsequently replaced, in Hellenistic and Roman times, by a single period of 24 ‘seasonal’ hours of the full day. Thus the actual length of 1 h varied according to the day of the year and, moreover, on each day except at the equinoxes the actual length of a day-time hour and a night-time hour were different. In a place like Egypt which is close to the Equator these daily differences would have been small, but they would of course become much greater if the same concept were to be applied at higher latitudes. In antiquity, only the Hellenistic astronomers regularly used

hours of equal length, and naturally those hours were chosen to be the same as the seasonal hours on the day of the Spring equinox. Gradually this uniform definition became the norm.

Since, following Babylonian practice, all Egyptian astronomical computations involving fractions were done using the convenient sexagesimal system (rather than our current decimal system), those Egyptian hours were then divided by the astronomers into 60 *first* small divisions (*pars minuta prima*, in Latin), or minutes, and each of these was in turn subdivided into 60 *second* small divisions (*pars minuta secunda*, in Latin); these are the origins of our terms ‘minute’ and ‘second’. Thus our modern-day convention for dividing up and subdividing the hours of the day is the result of a Hellenistic modification of an ancient Egyptian practice, combined with Babylonian mathematical conventions.

Interestingly, Egyptian mathematics used the following approximation for π :

$$\pi = 4 \left(\frac{8}{9} \right)^2 = 3 + \frac{13}{81} \sim 3.1605, \quad (1.1)$$

while Babylonian mathematics used the cruder approximation $\pi = 3$ or sometimes

$$\pi = 3 + \frac{1}{8} = 3.125, \quad (1.2)$$

and the approximation $\pi = 3$ can also be found in the Old Testament.

1.2.2 Mesopotamia

At the risk of oversimplifying, one could say that while Egypt excelled more in the arts, Mesopotamia excelled more in technology and science. Another important difference is that while the Egyptian civilisation developed in a fairly smooth way over a period of several millennia (the most dramatic difference being the final Ptolemaic period following the conquest by Alexander the Great), many different civilisations flourished in Mesopotamia and the surrounding area over the corresponding millennia.

Specifically, one should mention the Sumerians (ca. 3000–2000 BCE), Babylonians (2000–200 BCE), Hurrians and Hittites (1700–1300 BCE), and Assyrians (1400–600 BCE). Of particular interest for our present discussion is the period of the Persian empire, from 539 to 331 BCE. This is effectively

when astronomy was developed by Babylonian scholars. It was then quickly adopted by the Greeks, who eventually surpassed them at the beginning of the common era.

The earliest studies of astronomical science in Mesopotamia probably date from around the end of the fourth millennium BCE. There is considerable uncertainty in this statement, since the first written texts to have survived until the present, from the Babylonians, and subsequently from the Assyrians, only date back to the second millennium. The most famous of the Babylonian myths is known as the Poem of Creation (*Enuma elish*). This describes the sky as one half of the body of the goddess Tiamat, who was defeated by Marduk following a terrible cosmic fight. Since the heavens were considered sacred, they were thought to provide important omens, and the careful interpretation of celestial phenomena enabled priests to predict both natural and political events. It also compelled rulers to pay attention to these forecasts and react appropriately whenever necessary.

The Babylonian pantheon included Shanash (the solar disk at dawn or dusk) which was humanity's benefactor, while Nergal (the solar disk at noon) was an evil and destructive force, and Sin (the Moon) was the god of wisdom. The planet's names were Marduk (Jupiter), Ninib (Saturn), Nabu (Mercury), Ishtar (Venus), and Nergal (Mars). Surviving records of Moon risings date from about 1800 BCE, and a set of records of risings and settings of Venus over a period of 21 years (1702–1681 BCE) also survives. A particularly remarkable astronomical catalogue, known as the *Mul.apin* (see Fig. 1.3), dates from 686 BCE and survives in several copies, but is believed to have been compiled in the Assyrian period, in about 1000 BCE. It lists the names of 66 stars and constellations, as well as information on risings, settings, and culminations. The broad astronomical content and significance of the *Mul.apin* was only identified in recent times, in 1880.

The Babylonians apparently believed the Earth to be a large circular plate, surrounded by a river which no human could ever cross and beyond which there was an impassable mountain barrier. It was this mountain barrier that supported the vault of heaven, which was made of a very strong metal. Everything was resting on a cosmic sea. There was also a tunnel in the northern mountains that opened to outer space and also connected two doors, one of which was towards the east and the other towards the west. Every day, the Sun came out through the eastern door, travelled below the metallic heavens, and finally exited through the western door, spending the night in the tunnel. As you can see there are interesting similarities—but also differences—with respect to the Egyptian version.



Fig. 1.3 One of the Mul.apin tablets. Its actual height is 8.4 cm (Public domain image)

Another crucial difference between Mesopotamia and Egypt pertains to the local environment and climate: the Tigris and Euphrates were far less predictable than the Nile. While the Nile would only very rarely bring unexpected drought or flooding, the climate in Mesopotamia was far more variable. Strong winds, torrential rains, and devastating floods were frequent and would appear with little or no warning—let alone control. In this context, the regularity

of the astronomical cycles must have been particularly noticeable—as well as reassuring—in an uncertain and violent world.

Since the Sun, Moon, and planets were seen as gods (or at least visible manifestations of gods), the king and his counsellors must have watched avidly for portents that could be interpreted in terms of the intentions of these gods, so that disasters might be foreseen and possibly avoided or mitigated. The underlying assumption was that a counterpart in human events existed for every observable celestial phenomenon.

This was therefore the motivation that led priests to make careful and systematic observations of the heavenly bodies. The previously mentioned surviving evidence shows that this was already being done on a large scale in the eighteenth century BCE, but there is further evidence suggesting that particular phenomena (such as lunar eclipses) may have been regarded as ominous well before that epoch.

Gradually, the Babylonian civilisation acquired a remarkably accurate knowledge of the periodicities of the heavenly bodies: even relatively inaccurate observations can lead to accurate estimates provided one has a sufficiently long time series. Having accumulated enough data, they could use these periodicities to predict future positions of these bodies, and in particular predict the occurrence of specific phenomena such as lunar eclipses. Nevertheless, it is worth bearing in mind that this knowledge was limited to a small elite (the priesthood), and there was, as far as we know, no attempt to formulate any underlying theory of celestial motions (either geometrical or any other type).

Babylonian calculations were entirely empirical: they could predict when phenomena such as eclipses would occur, but did not (as far as we know) worry about why they occurred. Their algorithms were a collection of empirical facts without a theory—a bit like a phone book. On the other hand, the Greeks did worry. One could therefore say that astronomy was born in Babylon, but cosmology was only born later in Greece.

At this point, astronomy and astrology were born together. The predictions of what is now known as ‘judicial’ astrology applied to the royal court and the state as a whole, and not to ordinary individuals. This is to be contrasted with so-called ‘horoscopic’ astrology, which is closer to its modern remnant. The basis of horoscopic astrology is the assumption that the positions of the planets at the time of birth will determine the fate of the individual. This version only developed during the Persian domination.

In order to be able to cast such horoscopes one must have information about the positions of the planets for a given date, and this date can of course be one for which no observations have been recorded. Thus any such endeavour