The Human Condition

OUR PLACE IN THE COSMOS & IN LIFE

Stefan Wurm

THE HUMAN CONDITION

OUR PLACE IN THE COSMOS & IN LIFE

— *Discovering our Universe* —
— *Understanding our Home Planet Earth* —
— *Exploring the Evolution of Life* —
— *Considering the Rise & Future of our Species* —

Stefan Wurm

A·T·I·C·E

ATICE LLC, Albany NY

Published in the United States of America.

Library of Congress Control Number: 2019955721 (print edition).

First ATICE e-Book Edition | ISBN 978-1-951894-00-9.

For Brigitte, Geraldine, & Tristan

And our four-legged friends:

Houston
Caesar
Struppi
Miss Marple
Rollo
Mini Mundus
Krumi
Microbe
Fungus
Pebble
Figaro
Aramis vulgo "Margret"
Portos vulgo "Disturbed"
D'Artagnan vulgo "Der Dicke"

Stufen

Wie jede Blüte welkt und jede Jugend
Dem Alter weicht, blüht jede Lebensstufe,
Blüht jede Weisheit auch und jede Tugend
Zu ihrer Zeit und darf nicht ewig dauern.
Es muß das Herz bei jedem Lebensrufe
Bereit zum Abschied sein und Neubeginne,
Um sich in Tapferkeit und ohne Trauern
In andre, neue Bindungen zu geben.
Und jedem Anfang wohnt ein Zauber inne,
Der uns beschützt und der uns hilft, zu leben.

Wir sollen heiter Raum um Raum durchschreiten,
An keinem wie an einer Heimat hängen,
Der Weltgeist will nicht fesseln uns und engen,
Er will uns Stuf´ um Stufe heben, weiten.
Kaum sind wir heimisch einem Lebenskreise
Und traulich eingewohnt, so droht Erschlaffen;
Nur wer bereit zu Aufbruch ist und Reise,
Mag lähmender Gewöhnung sich entraffen.

Es wird vielleicht auch noch die Todesstunde
Uns neuen Räumen jung entgegen senden,
Des Lebens Ruf an uns wird niemals enden ...
Wohlan denn, Herz, nimm Abschied und gesunde!

Aus dem Roman *Das Glasperlenspiel* von Hermann Hesse (1877 - 1962).

Steps

As every flower wilts and every youth
gives way to old age, blossoms every stage of life,
blossoms every wisdom too and every virtue,
in their time and not allowed to last forever.
It must the heart with every call of life
be prepared to bid farewell and begin anew,
to courageously and without grief
bind itself to other, new attachments.
And to every new beginning there is a magic,
protecting us and helping us, to live.

We shall serenely stride through space after space,
not hanging on to any like to a homeland,
the world-spirit seeks not to fetter and constrain us,
step by step it wants to raise us, widen.
Barely at home in one sphere of life
and cozily settled, sluggishness threatens us;
Only those ready for departure and journey,
may snatch themselves from paralyzing routine.

Still, perhaps even the hour of death
will send us youthful towards new spaces,
Life will never stop calling for us …
Well then, heart, bid farewell and heal!

From the novel *The Glass Bead Game* by Hermann Hesse (1877 - 1962).

Translated into English by the author

Contents

Prologue

Falling I

For many years I have been looking to find the time to write this book. As it is with so many of us, our careers take their toll. We always are good at convincing ourselves that we have much more important things to do that require our immediate attention. Hence, we focus on these rather than on what we really would like to do. That of course is the standard excuse for most of us and it has been no different for me. Finally, I was able to rearrange my priorities, not so much because of my decisions but rather because of one of life's events, which helped me reset my perspectives on what really matters. So why do I believe that writing this book is an important task and what do I hope to accomplish with it? To answer these questions, we only have to look around us. Never has it been more challenging to keep up with what is happening right in front of our own eyes. Understanding how our world changes at an ever-quicker pace and how that may shape our future and more importantly the future of generations to come, is all but simple. Seeking to learn more about these major transformations and their potential consequences for humanity is something in which we all have a stake. Equally important is to understand how we can navigate what clearly are unprecedented challenges on a societal and individual level. A substantial fraction of humanity lives in relative prosperity and in open societies that to a greater or lesser degree enable them to participate in shaping their future. However, a large and growing number of people in these countries, many of them democracies, feel increasingly powerless to alter the trajectories their societies are currently set on. It has become a widely shared perception that without seminal changes as to how we conduct our human affairs, the inevitable result will be the destruction of much of our living Earth, including ourselves. This sentiment is quite understandable. Change is difficult for most of us and the more so if we have virtually no control over it. But change has always been part of humankind and change we can again. Our ability to rapidly pivot and adopt is likely one of our human key characteristics and advantages as a species. But in that we are pretty much alone. While we may be able to adapt to a rapidly changing planet Earth, given our ever-more sophisticated technologies, other life on this planet does not have the same option. Much of Earth's life is already struggling to keep pace with the changes that humanity continues to force on itself and everything else on this planet. How long can Earth's many endangered species continue before it becomes too much, and they quietly vanish into history? Figuratively speaking, the living world has been shrinking since the arrival of modern humans. The large mega-fauna, think of mastodons for example, has already vanished a long time ago. Our insatiable hunger for resources and food has changed ecosystems in ways that many of the large animals are now extinct or will be going extinct soon as we continue to take away their last habitats supporting such life; or we kill them for other purposes, most of them motivated by nothing else than greed and ignorance. However, many lifeforms will likely continue to thrive. Those are the small ones, most of which we cannot see with the unaided eye. Our planet always has been the planet of bacteria and it will remain the planet of bacteria long until after all higher lifeforms should have vanished; if it ever comes to that.

Our relationship with the natural world that has fostered our species evolution has radically changed since when humans were still only one element among many others in the much larger canvas of Earth's life, several ten thousand years ago. We have altered and changed the world in ways our distant ancestors could not have imagined. With that, our relationship with the natural world has changed from being part of it to seemingly owning it. In a sense, we have stepped out of this picture, the canvas of life, and started to repaint it. Humans always had to kill animals to survive but their relationship to the animals they needed to kill in order to continue was a very different one. We can still see and read this relationship in cave paintings our ancestors left us some forty thousand years ago. It was a relationship based on respect and the taking of animal life was seemingly perceived as something that needed to be atoned for, in whichever manner our ancestors may have done that; the cave paintings maybe being its visual expression. Today we slaughter animals on an industrial scale in a brutally efficient way. Annually, we consume now more than seventy billion land animals and if we throw in marine animals the count goes up to more than one trillion animals killed by the human species each year. This counts only the animals we just kill for food and does not include the countless animals that end up in other ways as casualties of how our species conducts its affairs. Most people have no idea anymore where the animal proteins they consume come from or through how much suffering meat ends up on their dinner plates. At the same time, the large majority of us cherishes their pets in ways that make them part of our families, almost no less than our other family members, sometimes even more so.

Why do we feel about our animal friends the way we do? Why do we torture and slaughter some animals while we treasure others? Why do so many have so little respect for life on Earth unless it is emotionally close to us, human or animal? Why do we seemingly defile the house we live in, which is the only house we have? There are no easy answers. However, an increased understanding of the havoc humankind has dished out to most of other life on Earth may be a good starting point to find some. We kill intentionally, like in the mass slaughter of enslaved animals but we may even kill more life unintentionally. We do that through nothing else but our sheer greed to make everything ours and exploit the resources of our home planet Earth to exhaustion. In stride, we accept the suffering and extinction of animal and plant life on Earth, which ultimately may lead to our own destruction. Why is that so? Why have we come to a point where greed in many ways has become the norm and killing off whole ecosystems in the interests of profits, regardless of however irreplaceable these biological treasures may be for life on Earth, still is an acceptable business practice? Nothing can justify such behavior but alas, it still goes on. However, there is a common denominator and that is the human perception that our species is seemingly the center and culmination of creation. A creation that supposedly was singularly undertaken for our benefit as some religions have it. As long as we do make no conscious effort to appreciate our proper place in this supposed creation, we will search for answers in vain. To ourselves, we humans are a riddle packed in an enigma. All too well we understand what we are doing and that much of it is not just simply wrong and devastating for other life but also endangers our future and the future of generations to come; but we still go on doing it.

At no time in human history has there been more information available to more people than in our time. One would think that we all have a general idea of where we live and who we are; as a species that is. However, nothing could be further from the truth than that. It just amazes what one can find by sifting through what frequently qualifies as the latest and greatest understanding of our universe, our planet Earth, and life on it, including our human species. There is a place for science fiction, and yours truly admires some of the masters of this genre, but we have come to a point where fiction has increasingly become confused with facts. The number of half-truths that continue to be propagated in various ways including all the alien fantasies and creationist myths still flourishing is astounding. Add to ignorance a good dollop of greed and you start to get a sense of what in our days increasingly informs private and public discourses and with that inevitably the future of our societies. Maybe this is stretching it a bit far, it likely is. However, it is just despicable to see how at practically every turn much of what we do seems to be driven by little else than sheer greed fortified with a good measure of ignorance. In many instances, they are mutually interdependent, as the one could possibly not do without the other. However, all too often, there are precise calculations behind the exploitation of nature at the behest of the few. Those wielding the sharp pencils with which they calculate their profits are not ignorant. They purposefully use the ignorance of others to execute their greedy schemes and while they call themselves business leaders, more often men than women, they frequently are only plain crooks clothed in business suits.

Our insatiable greed to enrich ourselves in the short-term not only ensures the continued destruction of Earth's ecosystems and their associated biological diversities that are fundamental for our own long-term physical survival. More than that, this greed endangers our mental and spiritual health, of us as individuals but also of us as a species.[1] While our spiritual health suffers every bit as much from this affliction as it degrades us as individuals, it also makes us all poorer, individually and as societies. Look around yourself and you will find that human life in many ways has become an accounting exercise. For most of humanity, it is a brutally one-sided game of monopoly where they have little to no voice with all the odds tilted in favor of monetary interests.

Frequently, we blame the various economic systems we practice for what ails our societies. Capitalism is a frequent scapegoat but there is not just one kind of capitalism. Some capitalist models have a strong social component while others care much less about how the broader populace fares as long as the economy thrives. Undoubtedly, capitalism has brought humanity enormous benefits. However, many parts of our world, including some developed countries, still struggle with embracing a humane and socially responsible form of capitalism. Even in the twenty-first century, some of the abuses we associate with early forms of capitalism persist, notably forced child labor. The way Adam Smith (1723 – 1790) envisioned capitalism was quite different from any of the more naked forms of capitalism we still can find today. In his view, individual economic self-interest was only one necessary ingredient for a functioning capitalistic economy. Humans conducting their businesses in an ethical manner guided by moral sensibilities based on shared values was the other vital ingredient.

All too often, that second ingredient does not seem to matter anymore today, if it ever did. However, let us not forget, the capitalist economic model is the best we have for now as under every other system that has been tried out on a similar scale people have fared much worse. One can only wonder how the focus on money and enriching oneself that seemingly permeates much of our societies today is compatible with the religious beliefs that many of the practitioners of capitalism profess. Regardless of richer or poorer, for all of us our time on Earth is limited. What legacy we leave behind is up to us. If it is riches accumulated at the expense of destroying the natural world, what kind of legacy will this be? Given how we continue to destroy the foundations of life on Earth, we should not expect future generations to judge us kindly. The wastefulness of our economic activities has become increasingly visible to most of us. What is receiving far less attention is what may be a much greater loss to humanity, the enormous loss of human capital. We seemingly accept the latter as an unavoidable by-product of how our societies conduct their economic affairs. Just think about it, the multitudes of our fellow humans that never really get a chance to make their full contributions to humanities progress just because they were born in a different place than the few luckier ones born into wealth and security.

It is not only the human species induced change itself, imposed by us on all other life on Earth, which has led to the precarious situation we are in today. Increasingly, it is the rate at which this change happens that makes it ever more difficult for other lifeforms to adapt; and that rate of change is only accelerating. Until a few centuries ago, the pace of change the human species imposed on its environment and on itself was still quite moderate when compared to today. For much of humanities past we may even see this pace of change as outright slow if we only consider its impact on everything else but us. One can argue that before the industrial revolution, we may have exterminated a few species and enslaved a few others for our benefit but otherwise our spreading human activities may have not been much more than a nuisance for the rest of life on Earth. The impact humanity had back then was much more limited as simply our knowledge and technologies were as well.

In the past, our ancestors also comprehended change more easily and fully within their own generation, and if not so, the generations immediately following would certainly digest it completely. Even if the deeper nature of those changes would not have been fully understood over one or two generations and their impact on humans and everything else may have been negative, it did not matter that much. Humans back then just could not wield the kind of power over nature and every living creature in it that we have today. Human populations were still comparatively small and large parts of our planet were still unsettled. Away from the most populous and early industrializing areas, life could still retreat from humans if needed. Much of our planet back then was still wild. For most of human existence, we put the pressure of change much more on ourselves than on our fellow creatures; except for those we hunted, domesticated, and cultivated. However, humanity seems to be rapidly approaching a point of no return. Most of us are keenly aware of how our utilization of planet Earth endangers much of Earth's life. We are also getting more apprehensive that in the end this may even be our own undoing. Until quite recently, the biological, social, and technological components of our species evolution all progressed at speeds that allowed humans to successfully adapt and spread over much of the planet. However, that has changed, and these three parts of our human evolution are now progressing at increasingly diverging speeds. Our continuously growing technological prowess forces change on us at a much faster pace than our evolution as a social species. Depending on the choices we make, our human engineered biological evolution may also soon outpace our evolution as a social species. If we should not be able to close those widening gaps, this may well put the future of our species at risk.

In a time of unprecedented change, most of us go about life not much different than our ancestors did. Evidently, things are more comfortable for many of us today, certainly more so than any of our ancestors could have dreamed off, but we are still essentially the same kind of people. Our lives remain focused on raising families, earning an income to support our loved ones and ourselves, and hoping for a better life for our children. Our time horizon is dominated by the generational clockwork we are all familiar with as we move from childhood to becoming adults, parents, and then eventually to hugging our own grandchildren. This experience is what defines our time horizon; this is where for most of us our responsibilities really carry weight. However, it is extremely hard for us to envision a distant future for our species and consider the potential impact that our decisions and actions today will have on future generations calling this planet home in hundreds, thousands, or millions of years. On a daily basis we are being inundated by all types of information which possibly could educate us on almost anything that is going on anywhere on our planet Earth. Unfortunately, for most of us, our attention horizon measures in days, and we tend to acknowledge and forget the seminal news just the same as the trivial ones. This seems to keep us from recognizing that collectively, as never before, we are changing the framework for human life on this planet in a non-reversible manner. Some of these changes will only develop their full impact in the far distant future and may have only a minor impact on the next generations coming after us.

We all can see ourselves as members of our families and in many cases also as members of various associations and organizations in our societies. However, few of us likely will consciously picture themselves as members of the human species. We just do not have a species consciousness. Because of that, we have not learned yet to plan as a species. However, we must learn to do so in order for us to survive and thrive through the ages to come. No species has ever achieved the level of societal and technological development that humans have, up and including the capability to destroy the basis for life on Earth for good. However, the human record on Earth is a rather short one and in the history of life on Earth many species, none of them endowed with such advanced capabilities, lasted for much longer than humans have been around. Nonetheless, the human species impact on Earth and its life is likely already greater than that of most species enduring for some ten or hundred million years longer than the handful of million years we can trace our species ancestry.

For mysterious reasons, we tend to believe that humanity has been an unprecedented success story in the history of life on Earth. Such a perception reflects a biased perspective on life. Other life before us has been successful for much longer and our success has not been so good for other living beings with whom we have shared this planet for a few million years. As a species, we must take responsibility for the planet and the animal and plant life it supports. We only got to where we are today as a species by being ruthless in our dealings with other creatures and exploiting the Earth for our sole benefit. Many other species have suffered greatly because of that. Ironically, the most successful animals in a world dominated by our species have been those that we enslaved while the ones that once were our prey or those that we could not make effective use of are now in many cases extinct. We continue to foul the Earth we live on and the air that we breathe while we poison the waters that cover much of our blue planet and drive ever more animal and plant life that we could not yet exploit to extinction.

We have a lot to improve if we want to be successful in securing humanities long-term success. This will require us to become a lot humbler than we are today and realize what our true position in the world is. We will need to learn how to live and evolve as a true social species that understands that its survival depends on our stewardship of planet Earth and all the life it supports. We need to acknowledge and control the dark side of our nature we are all still beholden to and start considering ourselves as members of a single species whose responsibility it is to secure a future for the life of Earth.

Leaving tribalism behind on our path towards becoming the true stewards of all life on Earth to which our species is so intricately connected will be a long and arduous journey. As with every journey, it is important to be aware of our point of departure or else we could end up in a quite different place from where we may have intended to go. This point of departure is what defines much of our human condition and pretty much everything else that matters for humanity flows from it. Our self-perception, as a species and as individuals, must be anchored in a solid understanding of the wider cosmos, of our home planet, and of the life that has evolved on Earth. It would be natural to expect that this is what schools should teach our children but unfortunately that is quite frequently not the case anymore, if it ever was. The partisan political discourse dominating television screens today has managed to confuse fiction with fact, sometimes out of ignorance, more often out of political calculus. Why exactly are so many of our politicians or shall we say elites just so shortsighted, is it because of greed, ignorance, deceit, or bribery? Whatever the reason, the result is inevitably the same and we or rather the generations coming after us will have to pay the price.

Without knowing the basic facts that shape and inform our human condition, we can have as many debates as we like, they will be meaningless and just as fictional as quite a few of the arguments frequently thrown around in such discourses today. Nothing prevents anyone of us from looking up what we would like to know. Most of us have the resources to separate fact from fiction and can discern between what may and may not be true. Some may assert that for everyday people it is difficult to understand any of these things, as they require training in science. Nonsense, there is nothing about our understanding of the universe, of our planet Earth, and the evolution of life on it that could not be understood by anyone with an open mind using their own common sense.

There are many facets to what we call our human condition. Most would agree that there are externalities that define our human condition and then there is who we really are, so-to-say the internalities of our human condition. Naturally, there is an intimate connection between these two aspects of our human condition. Who we are as human beings has been molded as much by the world around us and by our understanding of it, as it has been shaped by our search for who we are. The approach taken here to explore the framework of our human condition is not unlike that for a stage play. The play's story line is the human exploration of the world around us that has shaped our understanding of how we are a part of this world, individually and as a species. Giving a full account of all participants, so-to-say the stage roll of this play, is not possible. But many of the critical actors and their roles in this drama that has been playing out for thousands of millions of years will be introduced. In the first act, we will retrace how we learned about our universe and our place in it. The second act will look into our exploration of our home planet Earth. Our discovery of the evolution of life will take center stage in the third act. Finally, in the fourth act we will turn to our own human endeavor with our species emerging from pre-history. Of course, anything considering our human endeavor will also have to touch more on some of the internalities of our human condition. However, this book focuses on our discovery of the externalities of our human condition and only dwells on internalities, as necessary. Any meaningful exploration of the internalities of our human condition such as mythology, belief, philosophy, or art, just to name a few, is beyond the scope of this book, as it would require another volume or quite likely several more volumes.

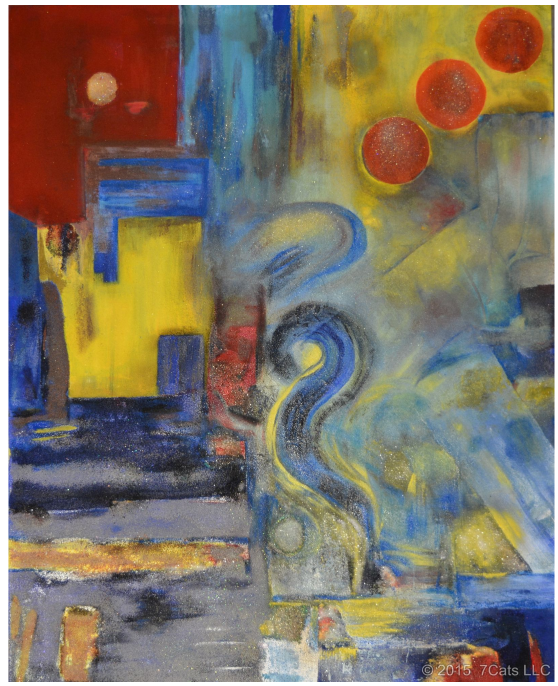
With respect to the subject matter of this book, many good books are already available, explaining our universe, our Earth, and the evolution of life on it. Some of them explore certain aspects in much more detail and depth. The challenge with writing such books is how to strike the right balance between being accessible to a broad audience and being interesting enough for a more scientifically inclined readership. Reading science graphs and formulas does not come natural to most of us and many do not speak the specific languages these sciences use. Unfortunately, this very much reduces the audience for such books. Other books with a lighter dose of science run the danger of not taking their potential readers serious enough. This book also had to strike a balance in this regard, but it does so in a different way.

I very much believe that all of us have the ability to understand the fundamental thoughts, ideas, processes, or histories that have led our human species to discover its true place in the universe and increasingly also learn more about itself. All that requires from each of us is common sense and the willingness to think through something with one's own mind. The choice made with this book was to put together the key aspects of the externalities defining our human condition in a single volume. How the understanding of these externalities has shaped and continues to shape our human condition is the story of this book. When you read this book, you will likely also sense the appreciation I do have for good science. Science is one of the greatest achievements of humankind; maybe even the greatest. However, science should not only be something for specialists. The stories of how we came to understand the framework defining our human condition is something we all should be familiar with, at least to some extent. How such understanding does change over time and how that alters our perceptions and the reality of the world around us including ourselves, truly concerns all of us. These stories need to use a language that we all can understand as they are a vital part of our shared human experience and condition. Of course, any such story telling effort has to be selective and this one is no exception.

To a good extent, what gets included is mostly dictated by keeping a coherent narrative flowing and not getting distracted in too many minute details; some of the latter is almost unavoidable and I am guilty as charged. Wherever possible, I relegated such detailed excursion to the notes section as an offer but not necessary for understanding the story. Naturally, a good part of the selection is also due to personal choice. There are many ways of how to tell this story and some may place the emphasis quite differently than I did here. It is in the nature of such accounts that no selection can adequately recognize the countless contributions by the many who have furthered human understanding in the areas considered here. Undoubtedly, some people certainly have contributed much more than others have; they are the ones we usually call out by name. However, their achievements would not have been possible without the enormous contributions of the many who are well known to the specialists in the field but remain anonymous to the broader public. The scientific journey that has brought us to this point is not a journey of individuals but of humankind itself. If the story told in the next chapters leaves you in awe of what humanity has learned about its place in the cosmos and about the story of life and how much there is still to learn, then writing this book has been well worthwhile. Frequently, such stories are crafted with the end in mind, making the outcome seem inevitable. However, that is not the case here. There is no end visible in our quest to understand the universe, our home planet, and the life of Earth, including ourselves. The exciting and perilous journey of humankind has barely begun, and the greatest challenges are still ahead of us. If we act wisely, there will be a long future for humankind with many more surprises in store. We can expect that quite a few of them will be just as or even more transforming for the understanding of our human condition than the ones retold here.

You can read the chapters in this book in the sequence presented or you can just read them according to your preference. The sequence I choose is somewhat suggestive. It not only implies an ordering in time but also in scale. Our very own story begins with the cosmos itself even though we only appear on Earth some 13,800 million years after the birth of our universe. There is nothing more grandiose than the cosmos. Next in time and scale comes the formation of our solar system and with it the evolution of our home planet Earth. Only then could the life that came before us and the life that we are still part of evolve on Earth. The most recent aspect is of course the story of our own journey as a species, which only started a few million years ago. While we understand who we are as a species much better today than ever before, there are still many open questions as to how we got to where we are today and more importantly, where we will go from here. The objective of this book is not to seek answers to any of these questions, as for a lot of them there are no satisfying answers yet and maybe there never will be. More to the point, for many of them where we could discuss answers we would have to search more in the roam of the internalities of our human condition, which are not part of this book. Questions of human beliefs, ideas, religions, or philosophies will only be touched on here in as far as they are necessary to understand how we discovered the externalities shaping our human condition. If this book contributes to more people seriously exploring themselves and their environments about what informs our human condition and what our true responsibilities are, to each other as well as to all other life on Earth, then it has accomplished that for which I am hoping.

Gaining Perspective



Red Dots Leaving

It is nothing but amazing in how short a time and how dramatically our knowledge of the real dimensions and time scales of the universe we live in has changed. It is only over the past few generations that humans have really started to appreciate the immensity of the universe. Not more than a few hundred years ago, most of humanity still perceived itself to be at the center of the known universe with everything around it, from the planets to the stars, being part of a divine creation with humanity as its crown jewel. We could not have been more wrong. In this chapter, we will try to gain some perspective of our place in the universe and as we do so, how humanity has perceived it throughout recorded history. It is important to understand how our cosmological perspective has evolved and changed. Particularly, how dramatic the known size of the universe has expanded over the past few centuries. To begin with, we will look back into our history and revisit major milestones in our understanding of the cosmos. Looking at this process from our vantage point in the here and now should make us blush when we realize how self-centered our perspective has been for most of human history, and often still is. At the same time, we should take pride in the scientific edifice that we built over the past few centuries that continuous to push aside the murky clouds of ignorance that have obscured our views and judgments for much of human history. As we do so, we need also to remember that humans have tried to understand the mysteries of the heavens for a very long time. Often that earned them applause and recognition but sometimes they were condemned for what they found or concluded. The history of the human effort to understand the universe is a checkered one. Like seemingly all human stories, it has many exhilarating moments, bright flashes of human ingenuity and perseverance; and it has its dark sides mired by ignorance, intolerance, and cruelty. It may be hard for us to understand today that there was a time when the search for truth was not based on methodical science and verifying facts but governed by religious beliefs and superstitions. Today, these forces are not holding back modern science anymore, but they are not gone. In too many places around the world, intolerance and ignorance still rule supreme. Even in our modern societies, we are not immune against these threats.

We have come a long way since Copernicus put the Sun back at the center of our solar system, which in his time was the equivalent of the center of our universe. By taking the first steps towards a full scientific understanding of our immediate cosmological neighborhood a few hundred years ago, we have come to a point where we can today speculate about how the universe evolved without blushing. However, the more questions we can answer, at least to some extent, the more questions we seem to have. In many ways, our better understanding of the actual physical realities of our universe has only increased its mysteries. The more we learn, the more we begin to appreciate how little we still know. In a sense, we are still asking the same questions our distant ancestors asked themselves long ago.

Where we stand today in the early twenty-first century of the Western calendar, we have relinquished much of what we believed to be true throughout most of known human history, including the very concepts of time and space. However, that does not seem to matter much in our daily lives. Many of us would agree that compared to our distant ancestors, we live more secure, comfortable, richer, and longer lives. For all we know, the human species and human societies have very much evolved since our modern human ancestors first appeared some 200,000 to 300,000 years ago. On average, throughout human history, the evolution of our species and our societies had a positive direction, at least from our human perspective.[1] All humans likely share very similar basic hopes, desires, and fears; as well as weaknesses and strengths. Regardless into which societies humans are born into or when they happen to live, their overall outlook on what really matters in life tends not to be so much different from ours and has seemingly changed little over time. Of course, there always has been and continues to be a significant bandwidth as to what humans ask from life and what they aspire to. This human bandwidth is probably today quite similar compared to what it was a few thousand years ago when we can identify the first human civilizations in history. If one could transport humans from these early days of civilization into our time, at a very young age of course, they likely would grow up no different from our children. As teenagers, we could expect them to master the challenges of our modern world similar to how we see our younger generations do so today. And why should the same not be the case for humans that lived before the time of agriculture and civilizations? Why could a hunter-gatherer baby, time ported from $\sim$10,000 years ago, not grow up in our modern world such that we would not know that an ancient human was living among us? We probably would never suspect a hypothetical ancient human growing up in our modern societies to be much different from us. However, that is likely only true in a superficial way because we actually have evolved quite a bit since then and as we will discuss later, there are good arguments that in an evolutionary sense we actually may be quite different even from our more recent ancestors living only a few thousand years ago. Despite our continuously evolving biology, that what makes us quintessentially human has changed little over the past few thousands of years, if at all. It may not even have changed much since the *Cognitive Revolution* some 30,000 to 40,000 years ago. However, our civilizations have evolved way beyond what humans even a few hundred years ago could have imagined. When our civilizations fail it looks as if humans as a societal species can rise to unimaginable heights and unfortunately also degrade to unspeakable evils. Both, it seems, perfectly supported by the basic bandwidth of the human animal species.

Probably for everyone of us, there is a point in life where we ask questions such as *"Where do we come from? What is out there? Why are we here?"*, or variations thereof. There always have been two parts to these questions, the physical one and the metaphysical one. Our efforts to answers both parts have been and continue to be constrained at any point in time by what we do know and maybe also by what we can know. Humans have likely asked those questions and tried to find answers from very early on. There probably has never been a tribal society at any stage of human development that did not have the equivalent of a diviner, shaman, or seer to interpret the world surrounding it, including the heavens. A few millennia ago, the start of urbanization and state building enabled a few privileged to spend more time on pondering these questions. Since then, human societies are seeking knowledge in ever more systematic ways to answer such questions, however temporarily or incomplete. Human communities have always been keenly attentive to the heavenly bodies seemingly closest to them, the Moon and the Sun. We know this from remnants of megalithic cultures, some of which date to almost 10,000 years ago. From them we inherited enigmas such as Stonehenge and similar constructions that align to the sunset of the winter solstice, the sunrise of the summer solstice, or both.[2] It is likely safe to say that humans have been looking skywards even earlier than that and may have marveled at and adored or feared the mysteries of the heavens as soon as they could articulate such things.

In a number of cultures, the Sun, our Moon, and even planets acquired divine status. For a long time, the influences of the Sun, the Moon, and the planets were seen as responsible for what otherwise could not be explained. This included for example our temperaments, supposedly ruled by any of the planets depending on their specific constellations at one's birth. Because we know better today, we can now smile when we see someone turn towards astrology to gain a peek into her or his potential future. However, for much of human history divining the future was a very serious and often deadly business, certainly not to be trifled with or laughed at. It really mattered if you were born under Saturn, Mars, or Venus. For our ancestors, the future of kingdoms and empires was enshrined in the planetary and stellar constellations visible in the night skies above. There never seems to have been a human culture without its very own original or adopted mythical stories of how the universe originated and works. Humans always felt the need to explain where everything that we find around us comes from. These mythical accounts are fascinating in their own rights, and they have much to teach us about human imagination as well as how humans perceived themselves in relation to what we refer to as the cosmos today. Anyone seeking to understand where we come from should be familiar with them. Unfortunately, it is now rare for children to read these mythical stories. A great loss for our cultures indeed.

In the following pages we will look into the human efforts to explain what could be observed in the heavens and how those observations could be brought into alignment with the then believed and known universe. In human history, we do not see this occur in a systematic way until the ancient Greeks tried themselves at developing self-consistent explanations of the movements of the heavenly bodies and the Earth.

The Universe before the Renaissance

Sumerians developed the first writing system, the cuneiform script, scratched into clay tablets and burnt for durability some 5,400 years ago. They are also credited with the invention of the sexagesimal system, using it for accounting, geometric problems such as land partitions, or weight divisions and likely for astronomical purposes too. It may be tempting to think that the Sumerian's use of sixty as the base for their number system, instead of our use of base ten, grounded somehow in astronomy because the Moon cycle is about thirty days. However, we do not know if that is so. Its origin could just as well trace back to the twelve bones in our four fingers times the number of fingers on one hand; five times twelve gives sixty. The Sumerians were also not the only ones devising a sexagesimal system. Other societies used it later but in no such case is there evidence that astronomy was behind the adoption of this numbering system. The Sumerians may have started their astronomical observations in the fifth millennium BCE and could have compiled detailed celestial maps as early as 3,000 BCE. In the second millennium BCE, its Akkadian neighbors conquered Sumer. The relationship between these two peoples is quite interesting, as the Akkadians seem to have admired Sumerian culture and incorporated it into their own; not too different from what the Romans later would do with respect to Greek culture; or what the ancestors of today's Europeans would do after that with Roman-Greek culture. The Akkadians kept for example the Sumerian language for religious and sacred purposes for a long time, just as Europeans later did with Latin. More importantly, they inherited the Sumerian's mathematical and astronomical knowledge and continued to expand it. Sumer-Akkad, united first under Akkadian and later under Sumerian rulers, eventually succumbed around 2,000 BCE to the invading Amorites, whom we would come to know as the Babylonians. The Babylonians absorbed Sumerian-Akkadian culture and astronomy seems to have been no less important to them than to the Sumerians. It is in the Babylon of the early first millennium BCE that we begin to see astronomy thrive. Babylonian religion and astronomy were closely linked, and we can assume that astronomical observations and star catalogs were not compiled for knowledge's sake but for religious purposes. Near-East cultures left us the oldest planetary observations we know, going back to as far as 4,000 years ago, although they were not put in writing until about some 2,700 years ago by the Babylonians. Given the close link between their religious beliefs and what they could observe in the skies above them it should be no surprise that they also gave us the first recorded forms of astrology. The Babylonians charted the paths of planets and stars alike but we do not know of any astronomical model that they may have developed to put their observations into a broader context. However, absence of evidence must not be construed as evidence of absence. The Babylonians may well have passed on more to the ancient Greeks than planetary or star tables. We may never know. What we do know is that the Greeks thought highly of Babylonian astronomical knowledge. When the ancient Greeks started exploring the night skies they may well have built on knowledge gained from the Chaldeans, as they called the Babylonians. The Greeks may also have benefited to some extent from ancient Egyptian astronomy, which seems to have served in Egyptian society similar purposes as it did in Sumer. However, the Egyptians did not leave the kind of detailed star catalogs in the historical records that we know from the Babylonians.

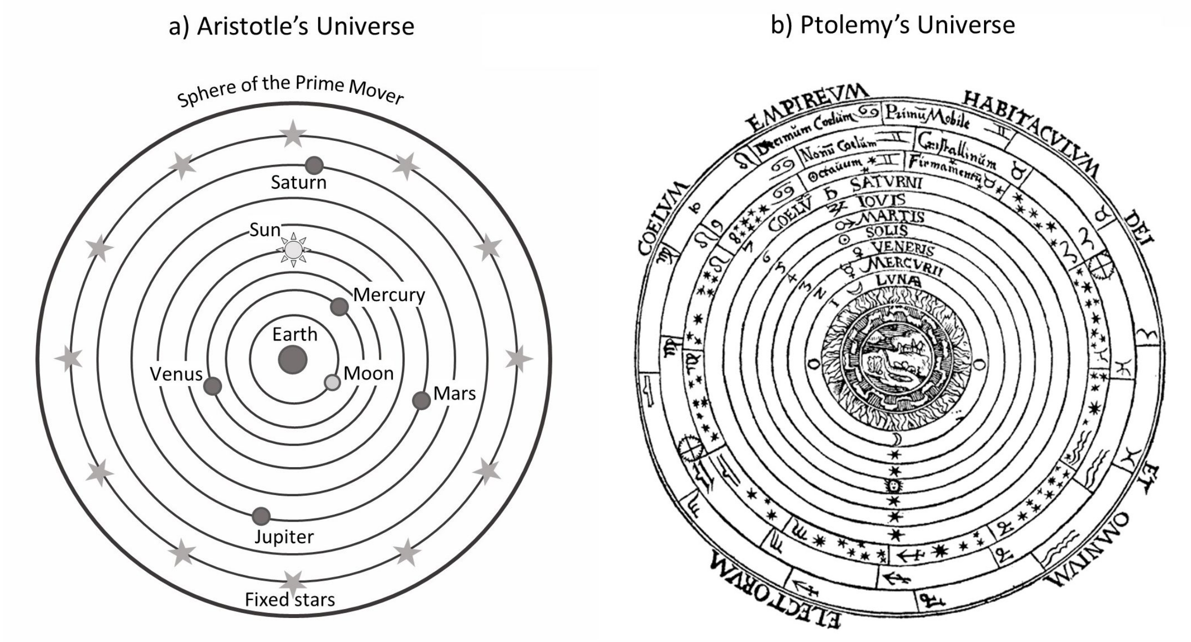


Figure 1.1: The universe in antiquity: (a) A simplified view of the universe according to Aristotle; (b) Ptolemy's astronomical system based on Aristotle's universe as depicted in *Cosmographia* published by Petrus Apianus in 1524.

In his *Histories*, published around 430 to 425 BCE, the ancient Greek historian Herodotus (c. 484 – c. 425 BCE) credits Thales of Milet (c. 624 – c. 546 BCE) with having been the first to predict a solar eclipse. Thales, a pre-Socratic philosopher from the city of Milet in Ionia, was an eminent mathematician and astronomer.[3,4] Herodotus was not the only one to give a description of the event, several other notable authors of antiquity also did. However, among all of them there is no contemporary eyewitness account of Thales prediction itself. We also do not know what method Thales used to predict this eclipse which according to modern astronomy must have been the total eclipse of May 28, 585 BCE. Maybe, given that Thales lived in Asia Minor, he did benefit more directly from the astronomical knowledge of the Babylonians; many scholars believe so. Thales is also said to have discovered the solstices, which define the seasons. By inference, scholars think that it is likely that he would have deducted the length of a solar year to be 365 days.[5] There are a number of other astronomical observations credited to Thales and altogether they may justify why in the eyes of many he has earned the distinction of being the first of the Greek astronomers. Although, according to Herodotus, Thales may also have had some Phoenician ancestry. The first solid footing we can get in ancient astronomy is the universe as described by Aristotle (384 – 322 BCE), the fourth century BCE Greek philosopher. It is important to realize that nothing in the model that Aristotle proposed for the universe is grounded in measurements. Instead, it owes everything to what Greek philosophers were good at, particularly a man of Aristotle's caliber: using deduction and logic to explain their observations of the natural world. A great example for his keen observation, deduction, and logic is Aristotle's conclusion that Earth was a sphere.

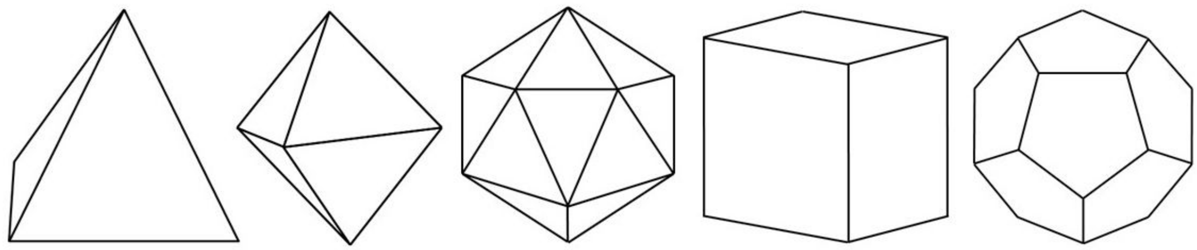


Figure 1.2: The five regular polyhedra referred to as the five Platonic Solids. Plato believed the first four to be the fundamental building blocks of matter and the last one used by the gods to arrange the constellation of the heavens. Later, Aristotle assigned the fifth solid to the ether. From left to right: the tetrahedron (fire), the octahedron (air), the icosahedron (water), the cube (earth), and the dodecahedron (universe / ether).

He reasoned that because Earth's shadow on the Moon is always circular, Earth had to be a sphere. Aristotle's model is geocentric: Earth sits at the center of the universe, surrounded by a large number of crystalline spheres revolving around it, most of which we will ignore here except the eight ones that matter.[6] Those eight spheres are, in increasing diameter, the spheres of the Moon, Mercury, Venus, Sun, Mars, Jupiter, Saturn, and the sphere of the fixed stars. This model would remain the foundation for our understanding of the universe for the next roughly 1,900 years. Fig. 1.1a gives a graphic depiction of Aristotle's model with Earth at the center and eight shells around it that contain the Moon, the then known planets, the Sun, and the fixed stars. Aristotle also assumed an outer sphere, called the sphere of the Primary Mover, the purpose of which was to keep all the shells contained within it revolving. If one switches the places of Earth and Sun and removes the Moon in this model as a separate planet, the ordering of the planets with increasing distance from the center would be correct. We do not know exactly when for the first time our ancestors understood the ordering of the outer planets. However, the fact that they did so shows that they must have made a connection between the observed times it took the planets to complete their perceived rotations around the Earth and how far away from Earth any given planet therefore was relative to the others. Like others before and after him, Aristotle also speculated on the elemental materials and substances that make up everything in the universe.

A little more than a generation before Aristotle, the philosopher Democritus (c. 460 – c. 370 BCE) had formulated his atomic theory of the universe. As maybe already the fifth-century BCE philosopher Leucippus before him, Democritus held the view that everything is composed of atoms, which he thought to be physically indivisible, and the empty space between them through which atoms move.[7] Like for so many of his fellow Greek thinkers, the worldview of the philosopher Plato (c. 428 – 348 BCE), a student of Socrates (c. 470 – 399 BCE) and Aristotle's teacher, was profoundly influenced by geometry. The ancient Greeks were certainly excellent mathematicians, but they were geometers first. One of the most famous theorems in all of mathematics that the ancient Greeks gave to us concerns regular polyhedra and it says that there are exactly five regular polyhedra, no more.[8] Fig. 1.2 shows these five regular polyhedra. Three of them are composed of equilateral triangular faces, the tetrahedron with four faces, the octahedron with eight faces, and the icosahedron with twenty faces. The dodecahedron has twelve sides made up of regular pentagons and the cube has of course six square faces. Plato thought that all matter was composed of them, and this is why we refer to them as the Platonic Solids. Plato and some of his fellow Greek philosophers, including his student Aristotle, believed the tetrahedron related to fire, the octahedron to air, the icosahedron to water, the cube to earth, and the dodecahedron to the universe and the ether.

The association of the dodecahedron with the ether in addition to the universe was Aristotle's contribution who conjectured that the universe as shown in fig. 1.1a consisted of two parts. An inner core universe including Earth and everything below the sphere of the Moon made from the four substances, earth, fire, air, and water, or combinations thereof; and an outer universe region, starting beyond the sphere of the Moon including the Sun, the planets and the stars, made from a fifth substance, the ether.[9] The introduction of this substance called ether, first made by Aristotle, will re-surface again in a different but related context in nineteenth century physics. And the five polyhedra would continue to intrigue astronomers for centuries to come, including one of the most famous, Johannes Kepler.

In the third century BCE, the Greek geometer and astronomer Eratosthenes of Cyrene (c. 276 – c. 195 BCE) gave us the first calculation of the Earth diameter. His result came close to within a few percent of the actual Earth diameter.[10] Aristarchus of Samos (c. 310 – c. 230 BCE) made the first systematic effort we know of to put the Sun at the center of the then known universe. However, Aristarchus may have been aware that already Philolaus of Croton (c. 470 – c. 385 BCE) referred to Earth as revolving around a central fire; the Sun. Aristarchus devised a method to measure the relative distances of the Sun and the Moon from Earth. In addition to proposing that Earth revolves around the Sun he also described Earth as rotating around its own axis. However, with exceptions such as the second century BCE astronomer Seleucus of Seleucia (c. 190 – c. 150 BCE), his ideas were largely rejected, and the Aristotelian model remained dominant.

Hipparchus of Nicaea (c. 190 – c. 150 BCE) is considered the greatest astronomer of antiquity. He developed accurate models describing the orbits of the Sun and the Moon, the first such models we know of.[11] Hipparchus is also credited with the discovery of the "precession of the equinoxes". Today this technical term is not part of the scientific vocabulary anymore, but the broader public continues to use it. What it describes is a shift of the positions on Earth's elliptic orbit where the equinoxes occur such that on average, they return to the same positions roughly every 22,000 years. This happens because Earth's rotational axis itself precesses around the normal to the ecliptic plane. As it does so, about every 26,000 years Earth's rotational axis going through the South and North Poles will point to the very same spots in the southern and northern skies that it is pointing at today. This axial precession combines with the rotation of Earth's elliptical orbit around the Sun which has a roughly 112,000-year period to produce on average the 22,000 years cycle of the precession of the equinoxes. As we will see, such astronomical cycles have a very direct impact on long-term climate change.[12]

In the second century CE, Claudius Ptolemaeus (c. 100 – c. 170), better known under his original Greek name as Ptolemy, gave us the *Almagest*. This book became one of the most influential scientific works ever written and cemented the Aristotelian view of the universe for the next one-and-a-half millennia. A major challenge astronomers faced was to give an adequate mathematical description of so-called "retrograde motions". Retrograde motion, the seemingly backwards movement in the sky of some of the planets at times during the year, is not supposed to happen in a model where all planets move on spherical shells around Earth. To resolve this, Ptolemy's *Almagest* introduced movements that are more complex. Instead of the planets fixed to concentric spheres that move them directly, he attached the planets to smaller circles, called epicycles, rolling on a sphere's surface as this sphere continued to rotate itself.[13] Initially this allowed describing the then observable retrograde motions of planets. However, as increasingly accurate measurements of planetary movements became possible, the gap between what astronomical models could explain and what astronomers could observe in the night sky began to open again. This led to the introduction of a second level of epicycles that carried the planets as the second epicycles rolled on the surfaces of the first epicycles. Naturally, with every such new iteration of epicycles the model became more and more cumbersome. In addition, to explain the observed speeding-up and slowing-down of planets as they seemingly orbited Earth, Ptolemy had to introduce another mathematical concept, the equant. The equant he defined as the point in the orbital plane of a planet seemingly orbiting Earth from which viewed the angular velocity of this planet remained constant. The concept of the equants, fictional points controlling the speed of planets then believed to orbit Earth, worked remarkably well. During the decline of the Roman-Hellenistic world and the final demise of the Roman Empire, much of the knowledge of the ancient world was lost and so was Ptolemy's work.

We must be ever grateful to the Muslim scholars to whom we owe the rescue of many of the ancient Greek works that we still have today, including the *Almagest*. It took until the fifteenth century before an accurate Latin translation of the *Almagest* became available again to Western scholars. Between the demise of the Roman Empire and the time first translations of astronomical texts into Greek or Latin from Arabic manuscripts became available to European scholars in the eleventh and twelfth century, Europe was in a true Dark Age regarding astronomy. However, not so in other parts of the former Roman Empire or areas adjacent to its past frontiers where astronomy continued to flourish in Muslim societies. As Muslim scholars translated astronomical texts into Arabic, among them the works of Ptolemy, they not only interpreted and analyzed them but also in many instances expanded or improved them. Among them we find in ninth-century Baghdad the Persian scholar al-Khwarizmi (c. 780 – c. 850) whose work on linear and quadratic equations, the *Al-Kitāb al-mukhtaṣar fī ḥisāb al-jabr wa'l-muqābala* (*Compendious Book on Calculation by Completion and Balancing*) was published in the early ninth century and translated into Latin in the twelfth century. It is the translation and reception of this work that gave us the word *Algebra* (*Al-jabr*). Another work of al-Khwarizmi only survived in its twelfth-century Latin translation, introducing Hindu-Arabic numerals and their use to the Christian West. Al-Khwarizmi was an eminent mathematician and he used this skill to improve astronomical tables that found their way into European lands through Muslim Spain.

For the ninth-century Arab or Persian scholar al-Farghānī, mabye better known under his Latin name of Alfraganus, we have no good estimated dates for when he was born or died. He was a famous astronomer in the Abbasid court in Baghdad and his *Kitāb jawāmi' 'ilm al-nujūm wa uṣūl al-ḥarakāt al-samāwiya* (*Book of Generalities of Astronomy and Bases* *of Celestial Motions*) built on Ptolemy's *Almagest*. Its twelfth-century Latin translation remained a major source for access to Ptolemy's work. Another Arab scholar in the court of Baghdad was Thābit ibn Qurra (c. 836 – c. 901) who eventually also became court astronomer. He is most well-known for his translations of Greek mathematical works into Arabic, among them Euclid (c. 325 – c. 270 BCE) and Archimedes (c. 287 – c. 212 BCE) but also re-translations of Ptolemy.

The work of the Arab astronomer al-Battānī (c. 858 – c. 929), also known under his Latin name Albatenius and occasionally referred to as the Ptolemy of the Arabs, would influence European astronomers up to Copernicus time. Among other achievements, al-Battānī refined astronomical tables and produced more accurate descriptions of the movements of the Sun and the Moon than Ptolemy. This was mostly because of his adoption of trigonometry for his astronomical calculations instead of geometrical methods. The fifth-century Indian astronomer Aryabhata (c. 476 – c. 550) had already used this method for his calculations but al-Battānī may not have been aware of this. Al-Battānī's observations and calculations also allowed him to improve on the accuracy of the length of one year and calculate the precession of the equinoxes and the inclination of Earth's axis with respect to the ecliptic plane. The ninth-century Arab mathematician and astronomer Ibn al-Haytham (c. 836 – c. 901), sometimes referred to in the West as Alhazen, is most well-known for his work on optics, prominent among it his *Kitāb al-manāẓir* (*Book of Optics*). Western Europe would not reach the level of understanding of optical phenomena he achieved until well into the Renaissance and in parts it would only accomplish this later than that. In mathematics, much of his work focused on the study of Euclid. Ibn al-Haytham's most famous work on astronomy is his interpretation and critique of Ptolemy's *Almagest* given in his *Hay'at al-'ālam* (*On the Configuration of the World*) and *Shukūk 'alā Baṭlamyūs* (*Doubts against Ptolemy*).

The eleventh-century Persian astronomer Abd al-Rahman al-Sufi (903 – 986), known in the West as Azophi, merged the Greek observational astronomy record as condensed in Ptolemy's star catalog with Muslim observational astronomy data. The result was the most comprehensive catalog of stars and constellations of his time, the *Kitāb suwar al-kawākib* (*Book of Fixed Stars*), published in 964. This work remained highly influential for more than half a millennium. Abd al-Rahman al-Sufi also has the distinction of being the first to record observations of two objects that we know today are not stars in the night sky but galaxies: the Large Magellanic Cloud, currently thought to be our third closest neighbor galaxy, and the Andromeda Galaxy, the closest spiral-galaxy to our own Milky Way Galaxy.[14] He was first to describe the Andromeda Galaxy as a nebulous object. As the telescope would not be invented for another roughly five hundred years, all of these observations were made using the naked eye.

The eleventh-century Persian Abu Ubayd al-Juzjani (died 1070), a pupil of the famous Ibn Sīnā (938 – 1037) better known under his Latin name Avicenna, similarly critiqued Ptolemy. He contended that the motions of celestial bodies should be uniform and not as described in the retrograde motion where at times planets seemingly moved backwards in their orbits. This led him towards the search for alternatives to Ptolemy's equants, which he summarized in his *Kitāb kayfiyyat tarkīb al-aflāk* (*The Manner of Arrangement of the Spheres*). His and Ibn al-Haytham's are the earliest critiques of Ptolemy's equants but they did not provide a solution. In another attempt, the twelfth-century

Spanish-Arab astronomer Nur ad-Din al-Bitruji (died c. 1024) formulated an alternative to Ptolemy's astronomical model that tried to avoid epicycles. While he succeeded in doing so, the calculations of planetary positions in his model were less accurate than predicted by Ptolemy's model and therefore the latter would reign supreme for another five hundred years. Abu Ubayd al-Juzjani's teacher Avicenna, the preeminent philosopher and physician of his time, had a wide range of interests, among them also astronomy. His contributions there include an extensive commented summary of Ptolemy's *Almagest*, the *Taḥrīr al-majisṭī* (*Summary of the Almagest*) where Avicenna added what he thought was missing in the original.

The last two Muslim scholars we mention here lived in the thirteenth century, the Arab astronomer al-Urdi (c. 1200 – c. 1266) and the Persian astronomer Nasir al-Din Tusi (1201 – 1274). Each of these two came up with a mathematical theorem, both of which were not included in any way in one of the translations such as of Euclid or Ptolemy that would have made them available to European scholars by the time of Copernicus. Because it is thought that the mathematical structure of Copernican astronomy could not have been built without these two theorems there must have been another mode of transmission; one that we do not know of yet. Certainly, we should expect that what we know about the transfer of such knowledge is likely incomplete as it reflects the historical record that is available to us. The latter of course has many gaps.

It is for a good reason that the period from the ninth to the thirteenth century is often called the Islamic Golden Age. Muslim scholars expanded star tables and while they revered Ptolemy's model they were certainly aware of its weaknesses. However, in the end, they did not correct it in any major way. Probably much more important than the Arab contributions to astronomy itself is their transmission of the astronomical knowledge of the Hellenistic World back to Europe, passed along with another major gift, mathematics. Muslim scholars made significant contributions to mathematics, adding to what they had learned from the Greeks. In doing so, they benefited from intimate contact with the Indian sub-continent. As much as Muslim scholars may have treasured the heritage of the Hellenistic World and learned from the ancient Greeks, they seem to have absorbed the knowledge of the East just as well. Astronomical thought on the Indian sub-continent has a long history dating back possibly before the first millennium BCE. In the fourth century BCE, following the short-lived conquests of Alexander the Great in India, a much longer and more fertile interaction between Greek and Indian cultures also had a profound influence. The fruits of which must have in turn enriched the interaction between Muslim and Indian astronomers and mathematicians as we can tell from what Western Europe inherited from both cultures.

At the time, mathematics in the West, as Europe had inherited it from the ancient Greeks, was essentially geometry; mathematical arguments and discussions were almost exclusively geometry based. The Roman numbering system was cumbersome to use for anything beyond simple calculations and really not well suited for astronomy. It is hard to imagine that medieval European scholars would have made much progress in astronomy, or any other science that requires mathematics, if they had to do their calculations using Roman numerals. Muslim scholars also built on the geometry-based mathematics of the ancient Greeks but as the Muslim conquest reached the border of the Indian sub-continent in the early eighth century, they were also exposed to the mathematics developing in parts of India. The result of this cultural exchange was the creation of the numbering system we still use today, based on the Hindu-Arabic numerals. This included the number zero, which scholars believe to have been "invented" in seventh-century India. The zero is such a known quantity to us, no pun intended, that it might be difficult to imagine how people could ever do without it; and why it took so long; and why the otherwise so smart ancient Greeks did not have the idea of introducing a zero. What it likely comes down to is a difference in philosophical outlook between the ancient Greek and the ancient Hindu world, which made Hindus much more comfortable with the concept of nothingness, as that is essentially what zero expresses. Of course, that is exactly what it remains if put in front of a number but put it behind a number and the zero will quickly reveal its power. Nothing divided by zero remains the same nothing. However, try to divide another number by zero and you will enter the roam of infinity. In a sense, the zero is as much an invention of philosophy as of mathematics. Arab scholars seem to have quickly understood that and put the zero to good use. The words *Algebra*, *Arithmetic*, or *Algorithm* reflect cultural transmission from Muslim to Christian countries and if the vocabulary for geometry had not already been defined so completely by the Greeks, based on Muslim and Hindu contributions in this area we could be using today Hindu or Arabic derived words instead of Greek ones such as trigonometry. In science, Europe had nothing even comparable to offer during this period. Scholarship in the west survived on what had been saved through the centuries from the works of the ancient Greeks and the Hellenistic World. However, most of the original Greek texts had been lost in the turmoil of the Roman Empire succumbing to the onslaught of the Germanic tribes invading the western half of it.[15] Because of that, much of the history of intellectual life such as it continued in the western half of the former empire, and to some extent in its still existing eastern half, reflected a diminished version of the intellectual traditions and schools of thought, as they had existed for hundreds of years throughout the prime of the empire. Much of the scientific knowledge of antiquity was more deeply rooted in the Hellenistic World of the East than in the Roman World of the West. Romans were great engineers, but one would be hard pressed to come up with the name of a truly Roman scientist; every so often when one believes to have found one, he turns out to be a Greek after all. The Romans may have regretted that from time to time, like when they for example besieged Syracuse from 213 – 212 BCE, which was home to the Greek mathematician and physicist Archimedes of Syracuse. According to legend, it was Archimedes genius in designing innovative weapons for the defense of his native city that prevented the Romans from taking it for much longer than they had anticipated. Eventually they succeeded and while plundering the city they killed one of the most eminent scientists of antiquity. With the demise of the Roman Empire, much of the Hellenistic World was lost to Western Europe. Few traditions of learning survived through the centuries of turmoil following the fall of West Rome, often referred to as the Dark Ages. They provided the intellectual canvas of Western Europe up to the early Medieval Period. Over time, scholars added to it through original contributions, of which there were few, or the rediscovery and interpretation of classical texts.[16]

We have to thank Anicius Manlius Severinus Boethius (c. 477 – 524), commonly referred to as Boethius, for translating several works of Aristotle into Latin in the sixth century. He seemingly also translated Ptolemy into Latin but nothing seems to have survived.

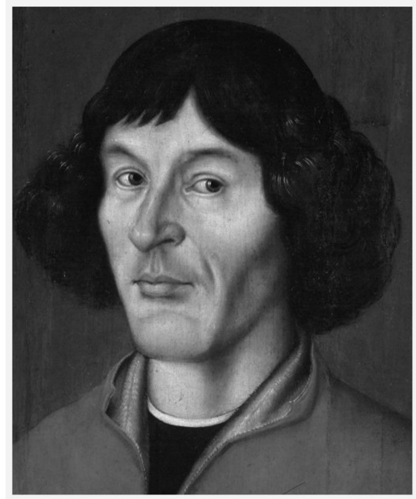
Boethius legacy was absorbed into the Catholic Church and with that, its philosophical and scientific outlook became heavily dominated by Aristotle; specifically as his philosophy was becoming increasingly important to medieval theology.[17] The result was that much of Aristotle's worldview as it was adopted by the Catholic Church became eventually religious dogma thereby cementing the geocentric model of the universe.

After the fall of Rome in the fifth century, on the surface, the work of Ptolemy remained the most important astronomical treatise for over a millennium. However, its reception in Europe only started again in parts in the twelfth century when Gerard of Cremona (c. 1114 – 1187) translated the *Almagest* from Arabic manuscripts into Latin. A first print edition based on this translation was published in Venice in 1515. Therefore, we can assume that by the late Medieval Period and the early Renaissance, European scholars cannot have scrutinized the *Almagest* for more than a couple of centuries. Fig. 1.1b reproduces a drawing of Ptolemy's model from the end of this period, taken from the *Cosmographicus liber* also known as *Cosmographia*, a famous work published in 1524 by Petrus Apianus (1495 – 1552).[18] Petrus Apianus book provided instructions in astronomy, geography, cartography, navigation, and instrument-making. It was highly acclaimed at the time and earned him the patronage of the then Holy Roman Emperor Charles V. The *Cosmographia* also included one of the first maps of the Americas, which must have pleased the emperor, as he was also the King of Spain and as such the largest land owner in the newly discovered continent.[19]

As fig. 1.1b shows, Ptolemy put the Earth firmly at the center of his universe and grouped the then known planets and the Sun around it, reaffirming Aristotle's model of the universe. As indicated above for Aristotle's model, if one exchanges the places of Sun and Earth and removes the Moon as a separate planet in this schematic, it would display the correct sequence of the planets in terms of their distance from the Sun, as we know it today. The innermost planet Mercury followed by Venus, Earth, Mars, Jupiter and Saturn. The outer planets Uranus, Neptune, and Pluto were not known in Ptolemy's time or at the time of Peter Apianus.[20] Only the assumptions that the Earth and not the Sun was at the center of the solar system and that the Moon was not an Earth satellite but a separate planet, kept Aristotle and Ptolemy from getting it right. Those two erroneous assumptions are connected because if one puts the Sun at the center of the universe instead of the Earth, one cannot continue to perceive the Moon as orbiting the Sun. As for the stars, Ptolemy's model puts them all into their own celestial sphere enclosing the nested spheres of the planets and the Sun with Earth at its center. Ptolemy's system had no understanding of the physical forces that governed the movements of the planets and the Sun and certainly not of the stars. We should not fault Ptolemy for that because nobody at the time had any good idea what governed the motions of the planets; and that was not to change until the age of Kepler and Newton. We must remember that for a long time in human history the astrological properties of planets and stars were much more important than their physical or astronomical properties. It was an important consideration into which planetary configuration a person was born. In the perception of many people back then, being born under Saturn, Mars, Jupiter, Venus or Mercury had its consequences; for some among us, it seems to be still of paramount importance even in our time as today's thriving astrology business and human susceptibility to it clearly show.

Before we look into the seminal changes about to happen at the dawn of the modern age, a few words on the importance or rather the inconsequence of flat Earth conceptions. While there is evidence that before what we can call the Greek enlightenment some cultures did believe that Earth was a flat object, we have already seen above that Aristotle thought that Earth was a sphere; and he could provide proof of that. His view of a spherical Earth gained prominence quickly. However, there was no understanding yet of gravity as a force attracting two masses with the direction of the force aligned along the axis between the respective centers of gravity of those masses. Hence a lot of confusion remained as to how exactly on the so-called antipodes of the Earth, we refer to them now as North and South Poles, people could walk on the Earth with neither one falling off. It seems that most at the time speculating about the shape of our planet intuitively understood that falling off the Earth could not be an issue. It must have been obvious that the peoples inhabiting the two respective antipodes would likely hold different views as to who was supposed to fall off the Earth. However, for those who did not come to this conclusion, a flat Earth model must have strengthened the logic of an up and down world model because different from a disk, a sphere really has no up or down orientation.[21]
Some still do hold the Catholic Church responsible for keeping alive and propagating the flat Earth model. However, that is a misconception. The Catholic Church's cosmological view took shape over centuries, articulated by a number of scholars before it became dogma. Like by Thomas Aquinas (1225 – 1274), the leading church scholar of his time, whose objective in interpreting Aristotelian philosophy was consistency with theology. That included a spherical Earth, firmly placed at the center of the universe. Others had arrived at this conclusion even much earlier than that, starting with the reception of Aristotle through the translation of Boethius in the early sixth century and confirmed again later by eminent scholars such as Beda Venerabilis (672 – 735). The Venerable Bede was a pillar of the seventh-century Catholic Church and like Aquinas later promoted to sainthood. The Catholic Church has been consistent in its view that the Earth is a sphere, the orb of the world as the Bible calls it. With regard to cosmological matters, the Catholic Church was steadfastly clinging to the Aristotelian geocentric cosmological view, including a spherical Earth sitting at the center of the universe. It used its full power - which at the time was also a very real secular power - to uphold its cosmological view against anyone questioning this divine order. It did so with Giordano Bruno who defended and expanded the Copernican heliocentric worldview with new ideas. For this offense, they burnt him on the stake as a heretic. It did so again later with Galileo Galilei when he pronounced the Earth to be orbiting the Sun. Galileo Galilei was luckier than Giordano Bruno was and escaped with his life, but otherwise he was destroyed. One can surmise that the cruel and harsh treatment that proponents of the geocentric worldview were subjected to by the Catholic Church would also have been meted out to proponents of a flat Earth theory; but alas, there were none. Somehow, later generations must have assumed that an organization that was fighting the worldview that put the Sun at the center of the universe so stubbornly would also cling to likewise outdated views regarding the shape of the Earth, such as the flat Earth model. However, far from the truth, the Catholic Church was only defending the Aristotelian cosmological view, which had become church dogma and a flat Earth model was not part of that but a spherical Earth at the center of the known universe was.

From the Renaissance to the Modern Age



Nicolaus Copernicus
1473 – 1543

Models such as Ptolemy's were exercises in geometry and trigonometry to explain the observed movements of the planets and the Sun in pure geometrical terms. This is the best Ptolemy, his predecessors, and for roughly one-and-a-half millennia, all of his disciples could do. We should not fault them for that but applaud their continued effort to improve and expand their observations of planetary motions resulting in ever better and more accurate data. While these incremental improvements achieved over generations were nothing like a straight-line progress, it eventually led to the realization by Nicolaus Copernicus and then Galileo Galilei that the Sun was the center of the universe.[22] Towards the end, it culminated in the data sets and understanding of astronomers such as Tycho Brahe and Johannes Kepler eventually enabling Kepler to formulate his three laws that first correctly described the movement of the planets before finally Isaac Newton discovered the laws of physics that govern their motions. But let's take one step at a time. With Nicolaus Copernicus, we are firmly in the Renaissance Age and like for the humanities and arts this period brought a new vigor to science, including astronomy. Only nineteen years after Peter Apianus had summarized the astronomical knowledge of his times in the *Cosmographia*, Nicolaus Copernicus published his most important work *De revolutionibus orbium coelestium* (*On the Revolutions of the Heavenly Spheres*) in 1543. Nicolaus Copernicus was born into a German family but his native city, a Hanseatic League port on the Vistula river, became part of the Polish Kingdom in 1466, thirteen years before Copernicus was born.[23] Copernicus pursued an ecclesiastical career and in 1495 became a Canon in Frauenberg, which today is Frombork in Poland.[24] However, his interests in astronomy were clear from early on as he studied the subject in Bologna and it would stay with him throughout his life. He started circulating his ideas early but not publicly, only among people he could trust. His *Commentariolus* (*Little Commentary*), an early summary of Copernicus initial ideas regarding the heliocentric worldview, is a ten page manuscript now kept in the Austrian National Library. This manuscript likely dates to before 1514 and never saw publication in Copernicus lifetime. Copernicus finished the manuscript of *De revolutionibus orbium coelestium* in 1532 but he continued editing it and another ten years passed before his work would finally be published in 1543. By then, Copernicus had passed away. In *De revolutionibus orbium coelestium*, Nicolaus Copernicus proposes an alternative to the system of Ptolemy by placing the Sun at the center of the known universe. The heliocentric universe that he proposed had seven spheres, i.e., less than the models of Aristotle and Ptolemy: six for the planets rotating around the Sun and one for the fixed stars.[25] As shown in fig. 1.3a, Copernicus placed the planets in the right sequence. Earth, which he calls Telluris, sits on the revolving sphere placed correctly between the spheres of Venus and Mars.[26] The Moon is not included anymore as a separate planet of our solar system.

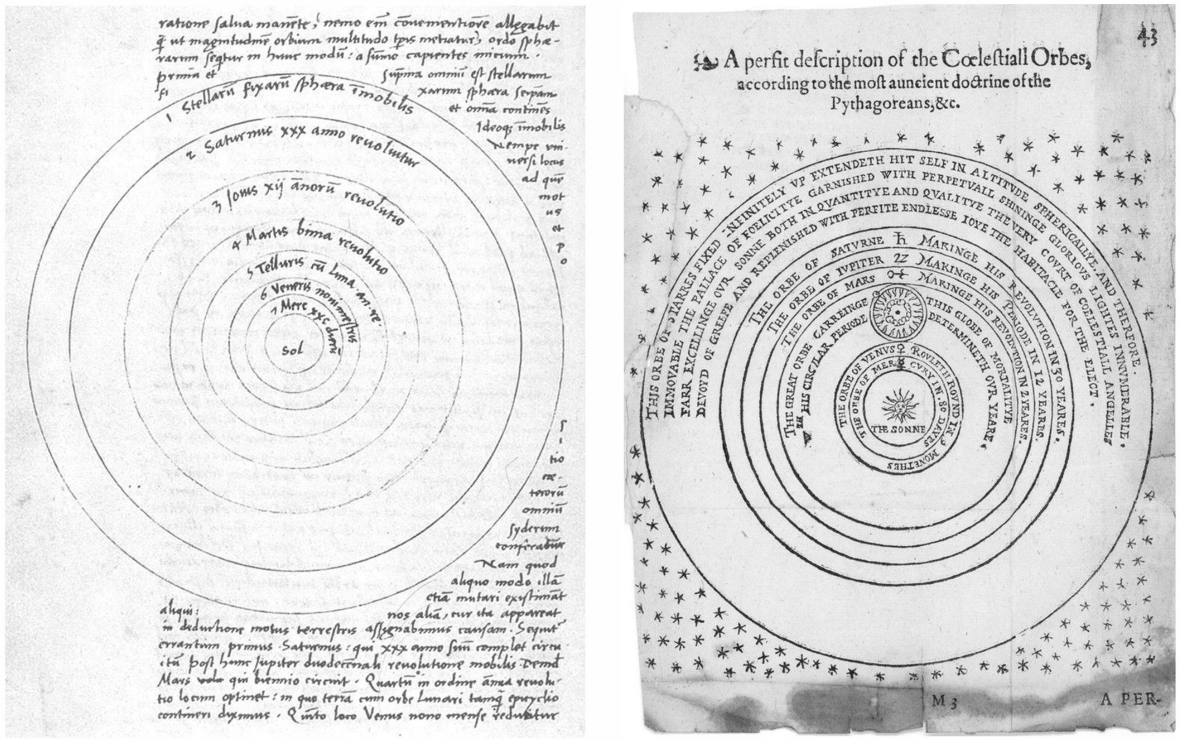


Figure 1.3: Nicolaus Copernicus heliocentric cosmos: (a) Copernicus original depiction of the cosmos as published in his *De revolutionibus orbium coelestium* in 1543; (b) Thomas Digges extension of Copernicus model published in his 1576 treatise *A perfit description of the caelestiall orbes* replaced the outermost shell of fixed stars with an infinite number of stars in a boundless cosmos.

Instead, the description in fig. 1.3a explains that it orbits our planet Earth. Copernicus realized that the apparent daily revolutions of the Sun and the fixed stars around the Earth as described in Ptolemy's system must actually be attributed to Earth's daily rotation around its own axis. It seems that what convinced Copernicus of the heliocentric worldview was not so much any new data but the increasing complexity Ptolemy's system required to keep it consistent with observations that a heliocentric model could explain in a much more simple way. This is a position surprisingly similar to modern science where, if observed facts are explained equally well by a simple and a complicated model, the simple model is usually seen as the correct one, and most often, though not always, this proves to be the right choice. However, we should not be surprised by this seemingly modern view because it is not modern at all. The very same rationale that is one of the key governing principles underlying the advances of science today, must have been quite familiar to a scholar of Copernicus caliber; albeit under a different name - Ockham's razor. The latter suggests that if there are multiple possible explanations for a given observation the simplest one is to be preferred to all others. William of Ockham (1285 – 1347), a fourteenth-century Franciscan friar is credited with formulating this principle. However, we likely underestimate the human capability for reasoning if we do not assume that this commonsense principle has been in use long before the times of

William Ockham whom we credit with formulating it. Copernicus was careful in sharing his heliocentric worldview, which he developed in his late thirties and forties. It was only just before his death at age seventy, that he completed his *De revolutionibus orbium coelestium*. While the Catholic Church got wind of his worldview and started inquires as to exactly what position he held regarding the Sun at the center of the universe and the Earth rotating around it, the church inquisitors never got to him because he died just as his work was being published.[27]

With the publication of *De revolutionibus orbium coelestium* we are looking at a major shift of the Western European worldview. The *Commentariolus*, distributed decades earlier, is likely to have already prepared the ground for this transition. With good reason, we speak of a *Copernican Revolution*, because it was nothing short of that as it represented a fundamental change in the understanding of the universe with all the consequences that would follow from that. Historians of science call this a paradigm shift; and it was not so much a scientific paradigm shift, which it was of course also, but much more of a philosophical paradigm shift. For the better part of a millennium, philosophy had been the handmaiden of theology and the latter determined what proper philosophy was and what heresy was. Because theology had expanded its scope into natural sciences, claiming authority over what was right or wrong, the philosophical dispute became inevitably a religious one. In the sixteenth century, the religious landscape was changing just as dramatically as the scientific and philosophical landscapes. Martin Luther (1483 – 1546), the German monk who started Reformation in Germany was an almost exact contemporary of Copernicus. Luther's reaction to Copernicus ideas was not favorable but some of his leading disciples saw things differently. The Catholic Church position remained unclear as it really had no official position on the topic. Copernicus theory was not a heresy, but it was also not something that religious astronomers were comfortable with. Ordained protestant as well as catholic astronomers continued to believe in Ptolemy's cosmology as this was consistent with their interpretation of the Bible. However, they clearly understood the superiority of Copernicus model. Sixteenth century astronomers, not only those who were ordained members of their churches but also lay astronomers would use Copernicus system but they could not publicly propagate it because the Bible was teaching something different. This also included the most famous of sixteenth-century astronomers, Tycho Brahe, whom we will meet shortly.

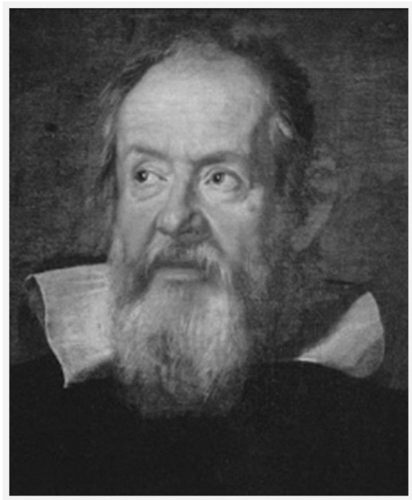
Aside from religious and philosophical concerns, there were other important reasons that slowed the adoption of Copernicus new cosmos model. They had to do with calculations. Copernicus heliocentric model was conceptually attractive as it provided simpler explanations for what was observed in the heavens. However, it did not make astronomical calculations easier and in some cases even more difficult. Copernicus still had to use epicycles because the planets seemingly did not move at uniform speeds as they orbited the Sun. In addition, he even had to use new epicycles to make up for removing the equants, which Ptolemy had introduced to account for the non-uniform speeds of planets as they seemingly revolved around Earth. Today we automatically assume that the ultimate purpose of astronomy is the scientific discovery of the heavens so we can better understand our universe. But in Copernicus times and long after that, this was not the prime motivation for people to look towards the planets and the stars. The most important function sixteenth-century astronomy had to serve was to calculate accurate positions of astronomical objects at any given time, as they were needed for navigation but much more importantly, for calculating horoscopes. Such computations, referred to as ephemerides calculations, were used to produce printed tables for the positions of astronomical objects at defined time intervals.[28] Ephemerides calculations go back to Babylonian times betraying their astrological roots. Unfortunately, the ephemerides calculated using Copernicus model were not more accurate than the ones calculated using Ptolemy's model. That was not owed to a weakness in Copernicus model but to a combination of factors. The most important one was that even though Copernicus heliocentric model had the correct understanding of the arrangements in our solar system, the new physics underlying this model would not be discovered for quite some time and so he had to use the old constructs to explain the motions of the planets. Posthumous, Copernicus became a widely renown astronomer but his heliocentric model largely remained a mathematical hypothesis. It was as such that Copernicus model of the cosmos would sometimes be part of the university curriculum. The Aristotelian view of the universe, adapted and molded into Christian theology did not leave any space for a different worldview as it saw no need for another explanation. Everything that needed explaining, the Christian-Aristotelian synthesis already did fully explain; and that was certainly expected to be also the case for anything new that would come along and require explaining. The *Copernican Revolution* was more an evolution than a revolution as it would play out over more than a century. Only with the discoveries of the early seventeenth century and the development of theories that could explain and predict the movements of the planets in our solar system would the *Copernican Revolution* be complete. It would eventually fall to Johannes Kepler and Isaac Newton to complete what Nicolaus Copernicus had begun. Some thirty years after Copernicus had died, the English mathematician and astronomer Thomas Digges (c. 1546 – 1595) published the first English translation of Copernicus work. But he did much more than that because he also added his commentaries on Copernicus model and his very own modification. Thomas Digges discarded the outermost shell carrying the fixed stars, which Copernicus had inherited from Ptolemy's model unchanged, and asserted instead an infinity of stars in a boundless universe (see fig. 1.3b). The drawing shown here was first published in his 1576 treatise *A perfit description of the caelestiall orbes*. With it, Thomas Digges shared his discovery of the infinite universe or as he put it in his own words, inscribed on the drawing just outside the orbit of Saturn:

> *"this orbe of starres fixed infinitely up extendeth hit self in altitude sphericallye, and therefore immovable the palace of foelicitye garnished with the perpetuall shininge glorious lightes innumerable farr excelling our sonne both in quantitye and qualitye the very court of coelestiall angelles devoid of greefe and replenished with perfite endlesse ioye the habitacle for the elect"*

Copernicus heliocentric model may not have been as convincing to his contemporaries as he may have wished, and it would take longer for it to be adopted than he likely would have hoped. But there, only some thirty years after his death, his worldview had opened the door for someone else to grasp the infinity of the cosmos for the very first time.

Giordano Bruno
1548 – 1600

The Catholic Church could accept the Copernican worldview when it came to practical matters. For some time it took no formal position regarding the Copernican heliocentric system. However, catholic dogma firmly asserted that the Bible provided the only correct worldview and negating Earth's position at the center of the universe was not acceptable; even more so if such a proposition came from within the church's own ranks. Maybe this is why the Catholic Church came down so hard on Copernicus first real and notable proponent in its own ranks, Giordano Bruno.[29] As a Dominican friar, Giordano Bruno was a greater danger to Catholic Church authority than a faraway German-Polish professor was, even though he was a Canon. Giordano Bruno was born in 1548, only a few years after Copernicus died, and the Catholic Church burnt him on the stake in Rome in 1600. The two philosophical works which eventually had the Inquisition come down on him were *La Cena de le Ceneri* (*The Ash Wednesday Supper*) and *De l'infinito universo et mondi* (*On the Infinite Universe and Worlds*), published in 1584 / 85 and 1584, respectively.[30] Both works are in dialogue form and lay out his arguments for the Copernican heliocentric worldview and against the celestial sphere model of Ptolemy. However, it is not for propagating the Copernican worldview that the Catholic Church prosecuted Bruno. His offense was far more serious. What likely made him most dangerous to Catholic Church authorities was the relativism he introduced with his expansion of the Copernican model. This expansion particularly included the notion of an infinite universe and of many other worlds that he put in the sphere of the fixed stars, which he interpreted correctly as other suns.[31] Clearly, the presumption of an infinite universe with many worlds that circle around stars just like Earth does around the Sun was an unacceptable relativism that would have jeopardized the religious authority of any church seeing itself as the exclusive source of all religious authority; and for the Catholic Church at that time, also of all secular authority.[32] Just to think of it, who was to tell if there were potentially people living in those other worlds that are also part of God's creation and as such must have been revealed the same religious truth as people on Earth? They could have their own churches, their own Popes. That cannot be, such worldviews must not take hold, and this is exactly why the Catholic Church condemned Bruno as a heretic and burnt him alive.[33] Giordano Bruno was not the first to suggest an infinite universe, as we just saw, this honor belongs to Thomas Digges, a layperson in protestant England, far from Rome and out of reach for its Inquisition. Bruno was a Dominican monk and his views were undermining the very authority of the Catholic Church and its Pope. The conflict between Bruno and his church was not a scientific one; it was a philosophical and theological one. He was not an astronomer but a clerical scholar with his own doctrine, which happened to differ significantly from the views his church held on the same matters. Bruno was no less fanatic than his church superiors were but he was in a weaker position; and so, he lost.



Galileo Galilei
1564 – 1642

In some aspects, Giordano Bruno may have anticipated Galileo Galilei. Like with his realization that relative motion is an important concept to understand the physical world around us. But there was certainly also one big difference: Galileo Galilei was an observational astronomer who came with new data to support his arguments. Galilei was, as we would say today, a scientist first and foremost. He was not driven by religious zeal like Giordano Bruno but by his desire to learn and know. In many ways, we can look at Galilei as the first scientist in the modern sense. Aside from that, like most of his contemporaries, he was certainly a religious person. The increasing rift opening between religion and the nascent modern sciences would define his life and his career. Born in 1564, Galileo Galilei was a late contemporary of Giordano Bruno. He must have known Giordano Bruno's works and arguments at least to some extent; and he certainly learned about his fate. Galilei was a scientist of many talents but here we are only interested in him as an astronomer. Galilei did not invent the telescope, this honor goes to the German-Dutch Hans Lipperhey (1570 – 1619) who invented it around 1608. However, he significantly improved it into what became known as the Galilean telescope. This he started to put to good use with his mapping of the Moon's surface, and the first observations and descriptions of the four moons of Jupiter and the phases of Venus.[34] The first two of these accomplishments revealed that celestial bodies do have structured surfaces as captured in his famous Moon drawings and that having an orbiting moon is not unique to Earth. The discovery of the phases of Venus however, was a true milestone in convincing other scholars that the heliocentric worldview was the correct one as such an observation was not compatible with a geocentric cosmology but could be well understood in the context of the Copernican heliocentric model. In 1610, ten years after the execution of Giordano Bruno, Galileo Galilei published his defense of the heliocentric worldview in his short treatise *Sidereus Nuncius* (*The Starry Messenger*).[35] And that was where his troubles with the Inquisition started.

The Medici were Galileo Galilei's patrons for most of his career and he named the four moons of Jupiter that he discovered in their honor the Medici planets; though the naming would not stick and they are rarely referred to by this name today.[36] The Medici were one of the most powerful families of late Renaissance Italy and they could protect him for a long time from the far reach of the Inquisition. For that, it was helpful that members of the Medici family also occupied high church offices, which was not unusual at the time for such a powerful family. In the case of the Medici, that also included a number of popes. However, when the Medici's influence on the Church weakened, or the Medici needed the Catholic Church more than vice versa, Galilei became vulnerable. In the year 1616, it became a formal heresy to hold, propagate, or defend the view that the Sun is the center of the universe and that the Earth revolves around it. Finally, the Catholic Church had taken an official position with respect to the Copernican heliocentric model, a little more than seventy years after the publication of Copernicus' *De revolutionibus* *orbium coelestium*. Why make it a heresy now, why not much earlier? Maybe a good guess is that with Galilei's work the experimental evidence started to accumulate that the Bible actually may be wrong. Before Galilei's observations, Copernicus model was only a mathematical hypothesis that could easily be discarded with some Christian-Aristotelian scholastic. Galilei's data started to confirm Copernicus hypothesis and to contradict the geocentric model, this is where the problem lay. Following this declaration of the Catholic Church position with respect to the heliocentric worldview, Galileo Galilei received an express warning on behalf of Pope Paul V (1550 – 1621) with regard to this: he was not to discuss the heliocentric worldview in any way, orally or in writing. When Galilei received this warning, he was also told that the Inquisition was not investigating him.[37] Pope Paul V passed away in 1621 and after the two-year papacy of Pope Gregory XV (1554 – 1623), Pope Urban VIII (1568 – 1644) took his place in 1623. Initially things started to look better for Galilei but that would change quickly. The Catholic Church in Galilei's time was a very different body than it is today. We need to appreciate that back then, the Catholic Church was as much a secular power as it was a religious one; and it was as much a political body as it was a church. Therefore, if a viewpoint could be useful in achieving the political objectives of the Catholic Church, religious aspects often were secondary concerns.

The relationship between Galilei and the Pope must be seen in this context to understand that the Pope initially was looking at the Copernican theory as something that could be useful to him, or could be used against others. Galilei was not a politician and failed to understand that the Pope was not supporting him because of a genuine belief that Galilei's arguments where right and that for that matter the Copernican heliocentric model was correct. The Pope supported him because Galilei and his arguments could potentially be useful to him; until they were not anymore. Galilei misread the Pope and may have interpreted Urban VIII willingness to listen to his arguments as an indication the Pope may eventually change the churches position with respect to the heliocentric model. Rather than looking at Pope Urban VIII as a patron and friend who turned on Galilei, we should look at the Pope as a political opportunist. As such, he initially allowed Galilei to write about the Copernican heliocentric theory as long as he treated it as a mathematical hypothesis. However, then Galilei got into trouble with another scientific publication that had nothing to do with astronomy but advocated an atomistic worldview church officials saw as undermining the transubstantiation of bread and wine into Christ's flesh and blood, the holy sacrament of the Eucharist. While the Inquisition cleared Galilei in its investigation in this specific instance, it could not have made his already strained relations with powerful church members any better.

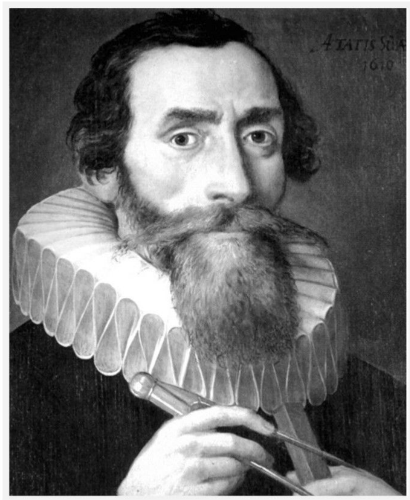
When in 1630 Galilei completed his major work on defending the heliocentric worldview, the now famous *Dialogo sopra i due massimi sistemi del mondo* (*Dialogue Concerning the Two Chief World Systems*), he was initially allowed to publish it in 1632. However, when Pope Urban VIII at first approved the book he did ask Galilei to include arguments for and against the heliocentric worldview and warned him not to advocate the latter. But Galilei must not have heard the warning or maybe he chose not to listen as he pretty much ignored the Pope's request. Nevertheless, he did honor another papal request, which was to include the Pope's viewpoint. Albeit the way in which he did it must have made the Pope only more ill disposed towards Galilei. As the title of the book indicates, Galilei presented the arguments for and against the differing worldviews in dialog form, a literary style that enjoyed wide popularity at the time. It could include two or more participants allowing the author to put his arguments into the mouth of one of them while having others offer contesting voices or alternative views. Of the three protagonists in Galilei's Dialogue, Simplicius is the defender of the Aristotelian view, essentially the Pope's position. The way Simplicius argues, he does come across in some respects as a simpleton. Exposing the perceived ignorance of Simplicius position was part of what the reader was supposed to learn from such a dialogue; in turn this would be followed by an educational conversion of Simplicius to the enlightened view, which for Galilei was of course the Copernican worldview. Galilei put the arguments of Pope Urban VIII in the mouth of Simplicius. While that was bad enough, choosing the name Simplicius for this character was another unforced error on Galilei's side, if it indeed was an error in the first place. While Galileo states in the preface of the dialogue that Simplicius is named after a famous Aristotelian philosopher, it is only a thought away from the Italian version of the name, Simplicio, which to this day carries the connotation of simpleton.[38] Exposing what could become powerful enemies to such ridicule was not a good idea. It took not long before Pope Urban VIII and the Catholic Church stopped the distribution of the *Dialogo* and summoned Galileo Galilei before the Inquisition.

At the time Galilei was summoned before the Inquisition, he was already sixty-six years old. He suffered from multiple ailments and cannot have been in a condition to put up much resistance when he was formally interrogated in the spring of 1633. After an extended interrogation, Galilei eventually confesses that he may have made the Copernican case in the *Dialogo* too strong. He offers to refute the Copernican case in his next book. However, at that point the Church officials in charge, including the Pope, seem to have decided to go all the way. His examination continues under the dual threats of indefinite imprisonment and torture. It eventually ends with a prison sentence, religious penances, and in what must have been a most humiliating experience, Galilei was forced to abjure his errors in a formal ceremony before the Inquisition.[39] At some point between this ceremony and his following house arrest in Siena, legend has it that Galilei murmured the words *"e pur, si muove - even so, it does move"*; implying Earth after all moves around the Sun and not otherwise. We do not know to whom he may have spoken the words or if they ever were spoken at all. For all we know they may be apocryphal but even though, most of us likely wished he said them. But little do we appreciate how powerful and dangerous the Inquisition was in Galilei's time. It is doubtful that anyone would have been able to spite the Inquisition and live to tell it.

Under house arrest in Siena he begins to write his *Discorsi e Dimostrazioni Matematiche Intorno a Due Nuove Scienze* (*Discourse on Two New Sciences*) which was published in 1638 out of reach of the Inquisition in Leiden, South Holland.[40] Galileo remained under house arrest until his death in 1642. The story of Galileo Galilei rings familiarly modern given the kind of disputes we continue to have in our societies today. Sometimes it seems we have not learned all that much from the past. Then as now, when it seems opportune and serves powerful interests, we are all too ready to bend facts and distort reality to fit our personal belief systems.

Tycho Brahe
1546 – 1601

Tycho Brahe was born a scion of noble families in Denmark in 1546, eighteen years before Galileo Galilei. Many know the portrait which shows him with an artificial nose. This is because he lost part of his nose as a student in a duel with another nobleman; and not as sometimes is insinuated due to syphilis. Even without his prosthetic nose, he was certainly a colorful figure. Much more important, Tycho Brahe was one of the greatest astronomers of his time and he is rightly famous for the copious and accurate empirical astronomical data he gathered over his lifetime. Tycho Brahe built his observatory from 1576 to 1580 on the island Hven in the Øresund, back then a Danish island, today part of Sweden. He named it Uraniburg after Urania, the muse of astronomy. Part of it was an underground facility he called Stiernburg, to minimize external disturbances such as winds distorting his measurements. The land on the island Hven on which he built Uraniburg was given to him by King Frederick II of Denmark to whom he served as astronomer as well as astrologer. When Frederick II died in 1588, the regency council that took over on behalf of the adolescent new king was less sympathetic to Tycho Brahe's work. As was Christian IV of Denmark, once he took the reins himself in 1596. Eventually, Tycho Brahe's position in Denmark became untenable and he went into exile in 1599. Before he left his homeland, he was able to finish the compilation of his thousand-star catalog, the *Stellarum Octavi Orbis Inerrantium Accurata Restitutio* (*The Exact Restoration of the Fixed Stars*) in 1598, but a complete print version of this work would not be available until 1627. However, in 1598 Tycho Brahe did publish his *Astronomiae Instauratae Mechanica* (*Instruments for the Restoration of Astronomy*), a detailed description of the astronomical instruments he had used in Uraniburg and Stiernburg. About a year after he left Denmark, he became the imperial court astronomer of Emperor Rudolph II in Prague. The significant contributions Tycho Brahe made to astronomy include a change in perspective based on an improved understanding of the astronomical distances involved. He applied this knowledge to comets and objects he interpreted as new star formations or *Stellae Novae* as he called them in Latin. We now know that what he observed was actually not the birth of a new star but the death of a star, what we call today a supernova.[41] Tycho Brahe was neither comfortable with the cosmos as Ptolemy had explained it nor with the heliocentric model of Copernicus. He acknowledged Copernicus as the greatest astronomer of modern times but he could not find his way through to accepting the vast distances at which Copernicus model put the fixed stars. Tycho Brahe's fame rests in his accurate observations and not in his astronomical theories. His own version of a heliocentric worldview that fell short of the Copernican heliocentric model was quickly forgotten once his famous assistant Johannes Kepler had made his mark. Hiring Johannes Kepler in Prague as his assistance quite literally changed the world. It would be Johannes Kepler, using Tycho Brahe's extensive data, who would eventually describe the planetary motions correctly within the Copernican heliocentric model.



Johannes Kepler
1571 – 1630

Johannes Kepler was born in 1571 in Southwest Germany. Unlike Tycho Brahe, Kepler came from a poor family and initially saw himself destined to become a protestant minister but his interests in astronomy and mathematics prevailed. Nevertheless, protestant theology and philosophy would strongly influence his scientific worldview. After completing his studies, Kepler moved to the city of Graz in Austria in 1594 where he taught astronomy and mathematics at the local protestant school. It was there that he completed his first work on astronomy, *Mysterium Cosmographicum* (*The Mysteries of the Cosmos*) published in 1597. Kepler believed the geometry of the Copernican universe to be shaped by the five Platonic Solids shown in fig. 1.2. Simply put, he believed that the distances between the shells carrying the planets in Copernicus model were to be found by nesting these solids into each other in just the right way. To achieve that they had to be used in the correct sequence and scaled to the right size with respect to each other. The resulting scaling ratios were critical numbers that just like the frequency spacing between musical notes would ensure a harmony of spheres. The search for a divine cosmic order was not unique to Kepler but he certainly took it most seriously and his *Mysterium Cosmographicum* speaks to that. The Counter Reformation forced the protestant Kepler to leave Graz. So, in 1600, already an accomplished mathematician, astronomer, and astrologer at the time, he moved to Prague where he became Tycho Brahe's assistant. When Tycho Brahe died in 1601, Kepler was chosen as his successor to complete the astronomical tables that Emperor Rudolph II had commissioned from Tycho Brahe. This work, building on Tycho Brahe's singular lifetime achievement, his star catalog containing the most accurate position measurements of stars and planets of the time, was published by Kepler in 1627 and is known today as the *Rudolphine Tables*.[42] It took Kepler almost thirty years to complete this work, which he did reluctantly as it required many tedious calculations that he felt kept him from more important and enjoyable pursuits such as his philosophical speculations. Kepler's philosophical speculations are all but forgotten today but he will be forever remembered as the first person who gave an accurate mathematical description of how the planets move around the Sun in our solar system.[43] Kepler derived his equations that describe the planetary orbits doing the very kind of tedious calculations that he disliked so much when he had to do them for the *Rudolphine Tables*. It was access to these tables built on Tycho Brahe's original compilations, which provided the basis for Kepler's seminal discoveries. After studying the planetary motion of Mars for a decade, Johannes Kepler published *Astronomia Nova* (*New Astronomy*) in 1609. This work was nothing short of revolutionary as it did away with celestial spheres to which the movements of planets were thought to be tied until then and introduced the concept of planets moving freely through space. *Astronomia Nova* included the first two of Kepler's laws. After another decade of observations, Kepler published what became his third law in his 1619 treatise *Harmonices Mundi* (*Harmonies of the World*). The three laws of Kepler are still part of the physics curriculum in high schools all around the world today. Put into simple words they state that:

- The planets move in elliptical orbits with the Sun at one focus.
- The speed of a planet changes such that the time elapsed between two positions is always proportional to the area swept out in its orbital plane between those.
- The ratio of the cube of the length of the semi-major axis of each planet's orbit to the square of time of its orbital period is the same for all planets.

Fig. 1.4a shows the elliptic orbit of a fictional planet with the Sun at one of the focal points. Most planetary orbits are much less elliptical than shown here. Earth's orbit around the Sun is much closer to a circle than to an ellipse. If drawn accurately, using the real relative length difference between the minor and major axis of Earth's elliptic orbit, the curve would very much look like a circle. In fact, on the largest scale that for practical matters is available in a book it would look indistinguishable from a circle. Hence, almost all textbook drawings show Earth orbit much more elliptical than it really is because otherwise it would convey the impression that Earth was circling the Sun in a circular instead of an elliptic orbit. A parameter called eccentricity measures how elliptic an ellipse really is. It is defined as the ratio between the length difference of major and minor axis and the length sum of those two axes. If they are of the same length, then the eccentricity is zero and we have a circle. The larger the length difference between those two axes, the closer the value of the eccentricity will be to one, but it will never reach it because a conical section with an eccentricity of one does not describe an ellipse anymore but a parabola.[44] The eccentricity of the orbit of Earth is currently 0.017, so it is very close to a circle. The major axis of Earth's elliptic orbit around the Sun is less than 0.015% longer than its minor axis. Among the planets in our solar system, the orbit of Mercury, the planet closest to the Sun, has the largest eccentricity with a current value of 0.206 implying that its minor axis is about 2.19% shorter than its major axis. However, the first detailed observational data for Mercury became only available with the improved instrumentation of the eighteenth century. Among the planets known in Kepler's time, Mars has the second largest eccentricity with a current value of 0.093, implying that the minor axis is about 0.44% shorter than the major axis of its elliptical orbit.[45] So Kepler picked well and his decade long gathering and evaluation of Mars orbital data was rewarded with insights that would change how humans perceive the cosmos.
Kepler's second law is schematically illustrated in fig. 1.4a. The distance vector between the Sun and the planet in the exaggerated ellipse shown for a fictional planet sweeps over the two areas **A** and **B**, which are equal in size, in the same amount of time. Consequently, a planets orbital velocity is highest at the perihelion, where it is closest to the Sun, and lowest at the aphelion, where it is farthest from the Sun. Because the orbit of Earth is much more like a circle than an ellipse the respective areas **A** and **B** would look very similar and the orbital velocities of Earth at the positions of perihelion and aphelion in its orbit around the Sun are not very different, 30.29 km/s vs. 29.29 km/s, or only a $\sim$1.7% difference. Kepler's third law is best illustrated by an example for which we will use the orbits of Earth and Mars using number values accurate to two decimal places.

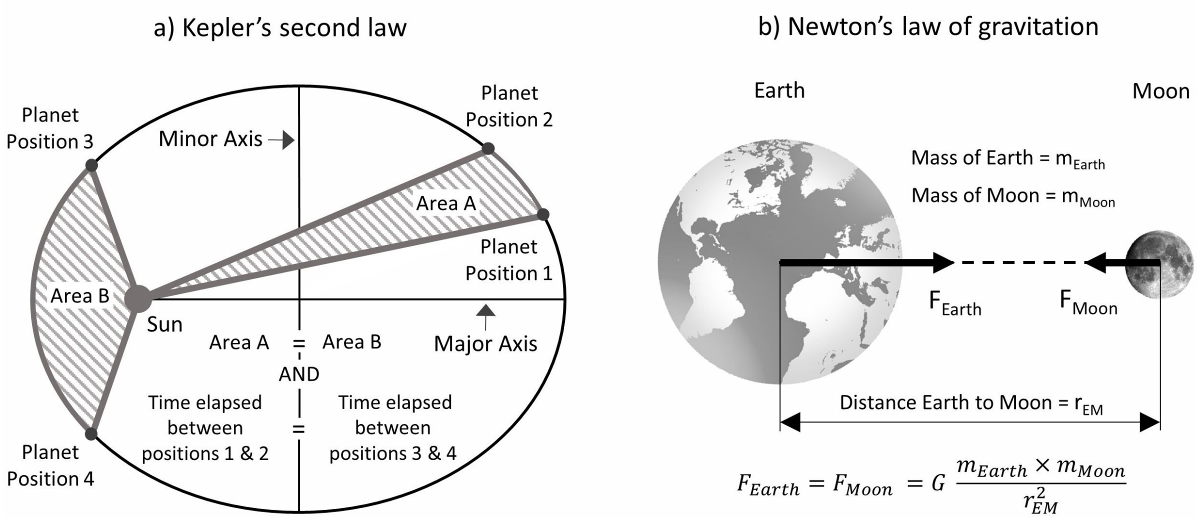
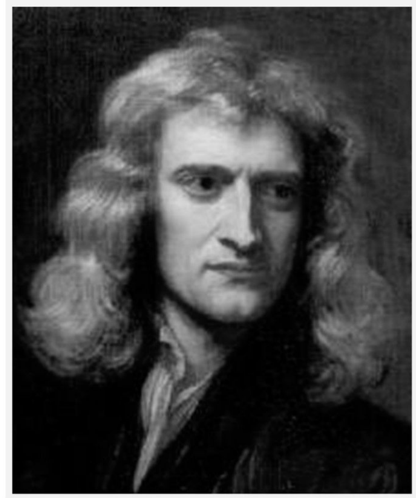


Figure 1.4: Discovering and understanding the laws governing the motions of celestial bodies: (a) Kepler's laws gave the first correct description of planetary orbits; (b) Newton's law of gravitation explained what makes planets move according to Kepler's laws.

The length of the semi-major axis, half of the major axis of Earth's orbit around the Sun, is 149.6 million kilometers. It takes Earth about 365.3 days to complete one 360-degree revolution in its orbit around the Sun. For Mars the respective numbers are 227.9 million kilometers for the semi-major axis and 687 Earth days for its orbital period. Kepler's third law states that if we multiply the length of a planets semi-major axis three times by itself and divide this number by its orbital period multiplied by itself, the result is the same for all planets. Doing this exercise with the numbers for Earth and Mars and rounding the results to two digits accuracy, we indeed get for both planets the same value of 25.09 for this ratio. The beauty of Kepler's third law is that it allows to determine a planets distance from the Sun if one knows its orbital period and the orbital period and distance from the Sun of another planet orbiting our star. However, no such distances from the Sun were known for any of the planets in Kepler's time. Therefore, he could only calculate the relative distances of planets from the Sun. But even that was far more than what was possible before Kepler's discovery. We can illustrate this by going back to the above example of Earth and Mars. By just knowing the orbital periods of Earth and Mars, which both were known in Kepler's time, he could calculate that Mars must be about 1.5 times more distant from the Sun than Earth. In the same way, he could also calculate the relative distances of other planets for which orbital periods were known. One such planet was Saturn, the orbital period of which is about 29.4 Earth years. Kepler therefore could calculate that Saturn must be roughly 9.5 times more distant from the Sun than Earth. Kepler's third law enabled astronomers for the first time to understand the relative distances within the then known universe. Kepler derived his three laws solely from observations and tedious mathematical calculations. As for the force between the Sun and the planets orbiting it, he assumed its strength to be inverse proportional to the distance of a given planet from the Sun. Kepler died some twelve years before the man was born who would proof this to be wrong: Isaac Newton.



Isaac Newton
1643 – 1727

Isaac Newton, born in 1643, changed our understanding of the laws of physics and the universe like no one before him. Newton's talent was recognized early, and he entered Cambridge Trinity College at age nineteen. Eight years later, at twenty-seven he became a professor of mathematics at Trinity and would stay there for almost three decades. His seminal work, the *Philosophiæ Naturalis Principia Mathematica* (*Mathematical Principles of Natural Philosophy*) published in 1687, not only laid the foundations of classical mechanics but of modern science as we still very much define it today. Newton's accomplishments were not just many; they were almost all revolutionary, in physics as well as mathematics. For some, their importance became apparent almost immediately, for others it took longer but with no less impact.[46] Like many a natural philosopher in his time and age, Newton was an experimental as well as a theoretical natural philosopher, or physicist as we would say today. As he worked on his theory of gravitation that eventually enabled him to derive Kepler's laws from first principles, Newton, being the astute astronomer he was, also developed the first practical reflecting telescope and we call such telescopes to this day Newtonian telescopes. What set him apart from his predecessors and many of his contemporaries was that he developed and improved such scientific instruments not so much through tinkering but through fundamental understanding of the laws of physics, as first laid out in his *Philosophiæ Naturalis Principia Mathematica* or later in his *Opticks or A Treatise of the Reflexions, Refractions, Inflexions and Colours of Light*, published in 1704. It looks like by the late seventeenth century time was ripe for a theory of gravity to emerge which could explain Kepler's laws. The hypothesis that the force exerted on planets by the Sun should follow an inverse square law had become common knowledge in Newton's time. It can be traced back to the work of the French astronomer and mathematician Ismaël Boulliau (1605 – 1694) and his *Astronomia Philolaica* (*Astronomy according to Philolaus*) published in 1645; a fact that Newton acknowledged in his *Principia*. The famous story of the apple falling on Newton's head igniting the spark of genius that resulted in Newton formulating his law of gravity may be apocryphal. But even if the story is not true, it sheds light on Newton's real accomplishment: he realized that gravity governs the motions of the Moon, the planets, the Sun, as well as the apple. It is this level of abstraction into a universal law of gravity and the ability to transform this concept into physics models using the language of mathematics, which differentiate Newton from earlier natural philosophers. A concept of gravity where the motions of apples and planets alike are subject to the influence of other massive objects without being physically in contact with them was nothing short of revolutionary. It is understandable that Newton's contemporaries were struggling to understand and accept this. However, they did so quickly because Newton's law of gravity together with the infinitesimal calculus, independently invented by Newton and the German philosopher and polymath Gottfried Wilhelm Leibniz (1646 – 1716), enabled them to calculate the motions of heavenly as well as earthly bodies and allowed them to make accurate predictions as to how those objects would move in the future.[47] Newton's law of gravity as illustrated in fig. 1.4b describes an attractive force acting between two bodies that is proportional to the masses of those bodies, and inverse proportional to the square of their distance. The distance measure is between the respective centers of gravity of the two objects and the force acts as if the respective masses of the bodies were concentrated in their centers of gravity. If that distance doubles, the force will weaken by a factor of four, if that distance if halved, the force will quadruple. The factor $\boldsymbol{G}$ in the equation is the universal gravitational constant also known as Newton's constant. Newton knew that $\boldsymbol{G}$ existed but, in his time, there was no way to measure it. He died in 1727 without being able to use his equation to calculate the mass of Earth. In 1798 the British natural philosopher Henry Cavendish (1731 – 1810) determined the value of $\boldsymbol{G}$ in a very demanding experiment using a torsion balance to measure the attraction between a small mass and a large mass; only then could Newton's equation be used to calculate the mass of Earth. The result that Cavendish reported to the Royal Society in 1798 was within 1% of the modern value for the mass of Earth.[48] In turn, this allowed the calculation of the mass of the Sun and then of the other planets. This shows the power of Newton's simple equation which from then on could be used to calculate the mass of any celestial object orbiting another one, be it in our solar system or somewhere else in the universe. Newton's law of gravity has one peculiarity that already troubled Newton and his contemporaries and that is action at a distance. Because there is no medium transmitting the force of gravity, the mutual interaction of the two masses is instantaneous even if they are arbitrarily far away from each other; any slight change in the position of one planet would for example have to affect the other planet immediately in what would later be called action at a distance. In Newton's own words:[49]

> *"… tis unconceivable, that inanimate brute matter should (without a divine impression) operate upon & affect other matter without mutual contact: as it must, if gravitation be essential and inherent in it."*

We know today that Newton and his contemporaries were right in harboring suspicions regarding action at a distance. However, Newton's equation worked exceedingly well and does so to this day throughout the galaxy where it is still a good enough approximation to solve most problems with it. We have to thank Newton's friend Edmond Halley (1656 – 1742) for helping to publish the *Philosophiæ Naturalis Principia Mathematica*, a watershed in modern science, as the Royal Society at the time did not have the funds to print Newton's work. Halley was an accomplished scientist and explorer and was the first to realize that not only planets but also comets travel on elliptical orbits around the Sun. Those orbits are just much bigger and so it takes the comets a long time to return, but return most of them do. Halley analyzed the available data and realized that recorded comet sightings in 1456, 1531, 1607, and 1682 were observations of the same comet, which he predicted to return in 1758. And so, it did, awaited and observed sixteen years after Halley had died, Halley's Comet, named in his honor since then. Undoubtedly, Newton's law of gravity governed the planets orbiting our Sun, the comets in the sky and the motions of everything else in the universe. Given sufficient data, in principle scientists could calculate the motions of every object in the universe, or so it seemed.

Frederick William Herschel
1738 – 1822

The eighteenth century was an age of great discoveries in many areas and astronomy was certainly one of them. And its greatest astronomer was Frederick William Herschel. Born in the Electorate of Hanover in Germany, Herschel left for Great Britain at age nineteen where he lived as a music teacher and organist, before his interests in astronomy took over in the late 1770's. Eventually, he was appointed court astronomer by George III, the ruler of Great Britain and Hanover. What earned Herschel this honor was the 1781 discovery of the planet Uranus. Most certainly, astronomers such as the English John Flamsteed (1646 – 1719) and the French Pierre Charles Le Monnier (1715 – 1799) observed Uranus independently before, maybe others did as well. At the time of discovery, Herschel thought he had found a comet and continued to describe his discovery as a comet even when other astronomers had already begun to believe it was rather a planet. The German astronomer Johann Elert Bode (1747 – 1826) realized that the earlier observations by Flamsteed were observations of the same object that Herschel had found. He used Flamsteed's data and combined it with Herschel's new observations to calculate the orbit of the suspected new planet; and he gave it a name – Uranus. By 1783, Herschel acknowledged that his discovery was actually a planet but by then he had missed the chance to name it. He had proposed to name it *Georgium Sidus* to honor his benefactor George III but the name did not stick and from then on the planet Uranus it was to be.[50] Herschel's contributions to observational astronomy and to the improvement of mirror and telescope making are even more impressive when one keeps in mind that he was already in his early forties when he became a professional astronomer. Another aspect is at least equally remarkable. In his astronomical pursuits, Herschel had a trusted partner and confidant, his sister Caroline Herschel (1750 – 1848). She regularly explored the night sky with him and became an astute observational astronomer in her own right, which included the discovery of several comets. While some may point to Hypatia of Alexandria as the first female astronomer, maybe we should be more inclined to give that title to Caroline Herschel. Hypatia of Alexandria lived in fourth-century Egypt, at the time a province of the East Roman Empire and was a revered scholar and philosopher. However, we know little of either herself or of her astronomical accomplishments, and all we know is only through other sources as none of her work has survived. For Caroline Herschel this is different, we know her contributions including the comet she discovered, still carrying her name today. Therefore, she seems to be more deserving of the title of first female astronomer. By the time William Herschel discovered Uranus, the distance between the Sun and the Earth was already known. This finally had been achieved through observations of what are called the transitions of Venus. A planetary transition describes the event when the orbits and rotational periods of two planets align in such a way that the inner planet, the one closer to the Sun, passes directly between the Sun and the outer planet.[51] If that happens, someone on the outer planet can make measurements to calculate the distances of both planets from the Sun. An observer on Earth must wait for a transition of Mercury or Venus, the two innermost planets, to calculate the absolute distance of Earth and the respective inner planet from the Sun in this way. For Venus those transitions occur in pairs separated by eight years, and the separation of any successive pairs itself alternates between 113 and 130 years. The last such pair of transitions happened in 2004 and 2012, the next pair will occur 2117 and 2125, and the next after that in 2247 and 2255. The orbital period of Mercury is 88 Earth days, which is much shorter than the 225 Earth day orbital period of Venus. Therefore, transitions of Mercury are more frequent than transitions of Venus. They occur at repeating intervals of roughly 10, 3, 13, and 7 years, restarting with a 10-year interval after that. The last Mercury transition just happened in late 2019 after a 3-year interval so the next Mercury transition will be in 2032. Hence, if you have not observed a Venus transition yourself yet, you missed your chance as it is unlikely that any of us will live to be around in 2117; but you may still have a good chance to observe a Mercury transition.

Because Venus is closer to Earth and bigger than Mercury, a Venus transition is easier to observe than a Mercury transition. This was even more so in the eighteenth century as optical instruments available to professional astronomers back then pale in comparison to what an ambitious hobby astronomer can order on the internet today. The English astronomers Jeremiah Horrocks (1618 – 1641) and William Crabtree (1610 – 1644) were the first to observe a Venus transition in 1639. Based on his *Rudolphine Tables*, Kepler had predicted a Venus transition for 1631 with the next one to occur in 1761; however, he missed the 1639 transition due to an error in his tables. It was Horrocks who shortly before the 1639 transition caught this error and correctly predicted it. It seems the first to suggest the idea of using a planetary transition measurement to determine the distance between Earth and Sun was the Scottish mathematician and astronomer James Gregory (1638 – 1675). However, he would not see his idea realized. The most recent pair of Venus transitions in his time had just occurred in 1631 and 1639 and the next pair of transitions would not occur for another 130 years. Finally, it was Edmond Halley, who after observing a transit of Mercury from St. Helena in 1677 suggested a method of measuring the transit of Venus from different observation sites on Earth to calculate the Earth-Sun distance. And so it happened when multiple astronomers around the globe observed the 1761 and 1769 Venus transits and calculated the average distance of planet Earth from our Sun to about 150 million kilometers.

Kepler's third law and knowing the orbital period of Uranus allowed astronomers to estimate that Uranus is on average 19.1 times further out from the Sun than Earth. Herschel's discovery of Uranus had dramatically expanded our universe. Compared to that, Saturn, the outermost planet before Uranus discovery, was 9.5 times more distant from the Sun than Earth. With the discovery of Uranus, the solar system just had increased by a factor of two; and with the just determined distance between Earth and Sun, scientists now could calculate the absolute size of the known solar system to an average radius of close to 2,900 million or 2.9 billion kilometers. One can only wonder what this distance must have felt back then. The fastest speed of travel overland was still on horseback and sailing ships took many weeks or months for their journeys to the New World or Asia. Almost three billion kilometers must have seemed an impossible distance to cover. It still is a very large distance to go today, but we have already done it.

On the twenty-fourth day of January 1986, after an eight-year-and-five-month journey, the spacecraft Voyager 2 reached Uranus. Voyager 1 and its sibling Voyager 2 have gone further than any other spacecraft. By mid-2019, Voyager 1 was more than 21.7 billion kilometers away from the Sun and Voyager 2 about 18 billion kilometers. Voyager 1 is already outside our solar system in interstellar space and Voyager 2 is still in the heliosphere on the very fringes of our solar system. Both spacecraft keep moving away further every day and one can track their actual distance from Earth or Sun online [1]. We are talking about very large distances now and it is about time that we upgrade our ruler to the unit of length used in astronomy to measure distances in our solar system and beyond, which is the astronomical unit or short AU. One AU is the equivalent of the average distance between the Earth and the Sun, so one AU is about 150 million kilometer.[52] As of Herschel's discovery of Uranus, the size of the solar system had increased to about 19.1 AU. In mid-2019, Voyager 1 was at a distance of ∼145 AU from the Sun and Voyager 2 at ∼120 AU.

After the discovery of Uranus in 1781 it took a while until astronomers came upon the eighth planet in our solar system, Neptune. However, when its existence was confirmed in 1846, it was not so much the discovery itself that was remarkable but how it had been made. In hindsight, it became clear that Neptune had been observed many times before but not been recognized as a planet. It seemingly would have been only a question of time before someone recognized what she or he were looking at. But this is not what happened as the existence of planet Neptune was predicted before it was observed and confirmed officially. What led astronomers to predict a planet with Neptune's orbit were irregularities observed in Uranus orbit that indicated according to Newton's laws the presence of another massive body. Several astronomers were working on calculations to determine the orbit and position where to look for this planet. In the end, it turned out to be the French mathematician Urbain Jean Joseph Le Verrier (1811 – 1877), who first predicted the position of Neptune in June 1846, which was confirmed at the Berlin observatory by Johann Gottfried Galle (1812 – 1910) in September 1846. This represents the first major example where we see astronomical theory and experiment starting to work together as we take it for granted in today's science. Humans had developed a scientific model of the heavens allowing them to predict the presence of celestial bodies that had not been seen before. With the discovery of Neptune, the known size of the solar system had increased to about 30 AU, which is an increase of roughly 56% since the last time the solar system had grown when the planet Uranus was discovered sixty-five years earlier. It took another eighty-four years until the American astronomer Clyde William Tombaugh (1906 – 1997) discovered the planet Pluto in 1930. Astronomers had suspected the existence of another planet beyond Neptune, a so-called trans-Neptunian planet, for some time. The reason they suspected the presence of another planet was similar to what eventually had led to the discovery of Neptune. While the presence of Neptune could account for much of the observed perturbations in Uranus orbit, it could not do so fully. Therefore, astronomers concluded there must be another planet beyond Neptune awaiting discovery. It took a while to find this planet as with a 39.5 AU average distance from the Sun, Pluto is much farther out than Neptune. The discovery of Pluto increased the size of the solar system by another ∼32%. However, by then it was already clear that the size of the universe was much larger than our observable solar system.

The astronomical community decided in 2006 to degrade Pluto from a planet to the status of a dwarf planet. What many of us may not appreciate, for a celestial object to be called a planet, astronomers require it to meet three criteria. First, it must of course be in an orbit around its star; second, it must have sufficient mass so its own gravity pulls it into a spherical shape; and third, it must clean out its orbital neighborhood from all other large competing objects. Pluto meets the first two criteria but not the third one. It is not the first time that astronomy encountered a celestial object much larger than an asteroid but not quite a planet. Before 2006, astronomers used to classify such bodies as minor planets, defined as planetary objects in direct orbits around the Sun, but neither a planet nor a comet. The first such planetary body designated as a minor planet was Ceres, the largest object in the so-called Asteroid belt between the orbits of Mars and Jupiter, home to numerous asteroids and several minor planets. It was the Italian astronomer and priest Giuseppe Piazzi (1746 – 1826), who discovered Ceres in 1801. While originally considered a planet it quickly became clear that this was not the case. With its 945-kilometer diameter, Ceres is at the same time the largest asteroid and the only confirmed dwarf planet within the Asteroid belt. Other large celestial bodies discovered in the Asteroid belt and classified as minor planets include Vesta, Pallas, and Hygiea. However, none of them made the cut for a dwarf planet when the International Astronomical Union reclassified minor planets as either dwarf planets or Small Solar System Bodies (SSSB). SSSBs include all comets and minor planets other than those that are dwarf planets. The list of minor planets is a very long one and includes many of the objects in the Asteroid belt and in the Kuiper belt, which we will discuss in the next paragraph. Ceres qualifies as both, a dwarf planet and a SSSB, the only celestial object to do so. Even so not official anymore, the designation of minor planet is still in use but the correct designation today for planetary bodies that are neither planets nor SSSBs is dwarf planet.

The redesignation as a dwarf planet may have been disappointing for many who cherished Pluto's planetary status. However, the now dwarf planet Pluto, though barely, still holds pride of place as the largest celestial object in the region of space we know as the Kuiper belt. The chilly expanse of the Kuiper belt is full of icy bodies, which astronomers believe are remnants of our early solar system. Pluto just happens to be one of the largest representatives of such icy bodies. There are other large objects out there, which astronomers refer to as Kuiper Belt Objects (KBO). The Kuiper belt's extension stretches from just beyond Neptune at 30 AU to approximately 50 AU from our Sun. The American astronomer David C. Jewitt (born 1958) and the Vietnamese-American astronomer Jane X. Luu (born 1963) discovered the first KBO in 1992. Their observation of this first trans-Neptunian object besides the dwarf planet Pluto and its moon Charon marks the discovery of the Kuiper belt itself. This region of space derives its name from the Dutch-American astronomer Gerard Kuiper (1905 – 1973) who suggested in 1951 that some comets might originate from there. Because the British astronomer Kenneth E. Edgeworth (1880 – 1972) already suggested in 1943 that comets and larger bodies might exist beyond Neptune this region of space is sometimes also referred to as the Edgeworth–Kuiper belt.

Examples of KBOs similar in size to Pluto include other dwarf planets like Sedna or Eris, about three quarters the size of Pluto and almost the size of Pluto, respectively.

Both are recent discoveries. A team of astronomers at the California Institute of Technology led by the American astronomer Michael E. Brown (born 1965) discovered Eris in 2005. He and his team also were co-discoverers of Sedna in 2004 working with a team of astronomers from the Palomar Observatory including his co-discoverers, the American astronomers Chadwick A. Trujillo (born 1973) and David L. Rabinowitz (born 1960). Sedna and Eris are much farther out in our solar system than Pluto is. At an average distance of 39.5 AU, it takes Pluto 248 years to complete one orbit around the Sun. By comparison, it takes Eris 558 years to complete one orbit around the Sun and Sedna's orbital period is roughly 10,500 years. Sedna's orbit is by far the most elliptical orbit of any planetary object we know today with its closest position to the Sun, the perihelion, around 76 AU and its furthest position from the Sun, the aphelion, around 936 AU. The Kuiper belt is home to a great number of icy objects, among them other dwarf planets such as Quaoar, Haumea and Makemake, the first discovered in 2002, the latter two in 2008, again by astronomer teams from the California Institute of Technology and the Palomar Observatory. We can count on terrestrial or Hubble astronomers to continue finding more such dwarf planets. However, science is also pursuing other ways to explore the Kuiper belt.

In 2006, NASA launched the spacecraft New Horizons to explore the outer regions of our solar system and then follow the Voyager probes into deep space. After a nine-and-a-half-year journey spent in hibernation, astronomers activated New Horizons shortly before its Pluto flyby to record high-resolution images of Pluto's surface. New Horizons started taking images of Pluto in January 2015 when it was about 200 million kilometers away from the dwarf planet. In mid-July 2015 it reached its closest point of approach in its Pluto flyby at a distance of less than 20,000 kilometers; many of our Earth satellites orbit Earth at larger distances. The high quality of the images New Horizons gave us of the dwarf planet Pluto completely changed our understanding of this far away world and there was certainly nothing even remotely dwarfish in those images [2]. Nothing on Pluto was like astronomers and astrophysicist had thought it would look like. As the spacecraft New Horizons passed Pluto and continued its journey through the Kuiper belt, it explored other remarkable KBOs such as Ultima Thule.[53] An object only some 35 kilometer in length, Ultima Thule looks like a larger elongated and a smaller round potato stuck together. Astronomers believe these "potatoes" are primordial ice rocks stuck together by a process resembling the first steps in planetesimal accretion. The theory is that Ultima Thule coalesced from a cloud of rocky, icy material far from the Sun where it would remain, never to move beyond this early stage of planetary formation.

We are getting ahead of ourselves here and we need to look at what happened in the eighty-four years that had passed between the discoveries of Neptune and Pluto. Put simply, science had started to appreciate that the roam of the stars, fixed in Ptolemy's cosmos to the outermost shell of his model, in reality vastly exceeded the size of our solar system. However, this process began long before the discovery of Neptune. Already in 1576, Thomas Digges had discarded Ptolemy's outermost shell of fixed stars and imagined a cosmos where fixed stars could be at infinite distances, a concept that most of us have a hard time to grasp. A little more than one hundred years later Edmond Halley had discovered that the fixed stars where not fixed at all if one observed them for long enough, meaning for centuries. Evaluating astronomical data recorded through centuries, he could show that the positions of fixed stars had moved since the ancients had first observed them. Not only did it become increasingly evident that the universe was much larger than what astronomers had assumed, it was also much less static. Not only planets and comets were moving, but stars also moved, however imperceptibly, during a human lifetime.

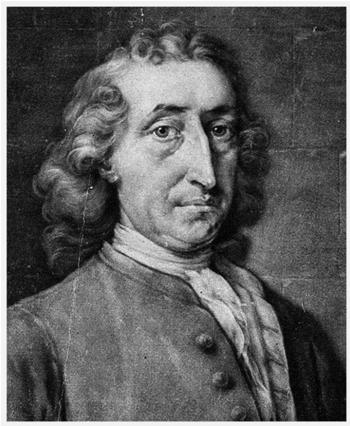
Another important step towards understanding what we actually see when we look up into the night sky was the discovery that the speed of light is finite. In 1676, the Danish astronomer Ole Christensen Rømer (1644 – 1710) was the first to succeed in measuring the speed of light. He found that the speed of light is finite and from his measurements, he estimated it to be around 200,000 kilometers per second.[54] Our more accurate measurement methods today tell us that the speed of light is 299,792 kilometers per second so Rømer's estimated value was by about one-third too low. However, the implications were still the same and these were certainly acknowledged by astronomers such as William Herschel. Herschel understood that the stars we see in the night sky were essentially suns like ours and that they must be quite far away for them to appear so tiny to us. He also understood that when we look into the sky we actually do look back into the past because the speed of light as Rømer had found was finite. While the light reflected from the Moon takes about one second to reach Earth, the light of our Sun takes about eight minutes to reach us. Therefore, what we see in the sky is always the Moon as it was one second ago and for the Sun how it looked like eight minutes ago. Translated into a distance perspective, the Moon is about one light-second and the Sun is about eight light-minutes away from Earth. For the stars and galaxies visible to us in the night sky, it became clear that the light reaching us from them today had to travel much farther than that. While Herschel could not have known how immense the observable universe actually is, he had a good idea that it was quite enormous.

From early on, humans realized that they could use the stars for orientation and navigation and later to plan for the agricultural year, but the heaven and the stars always also had mythical and religious meanings. Throughout history, these latter aspects often dominated human cosmology and at times some people lost their lives because their view of the universe was not compatible with the one ordained by those in power. Such was the case with Giordano Bruno, executed by the Catholic Church for believing the universe was different from the one the Catholic Church mandated. However, only a few generations after the execution of Giordano Bruno, it had become a clear and indisputable fact that the universe was quite different from what religions would have it to be. Humans had invented modern science with its methodical and critical research of experimental data and theoretical predictions. In matters of the universe, theological arguments had given way to astronomy, physics and mathematics. The Catholic Church, and more broadly, religion and mythology had lost their dominance over what humanity was allowed to know about the universe for good. This was the status at the end of the nineteenth century and many at the time may have thought that everything important there was to know about the universe was already known. They could not have been more wrong. Over the roughly past one hundred years, our picture of the universe has changed even more dramatically than throughout all of human history before. This is where we will turn our attention next.

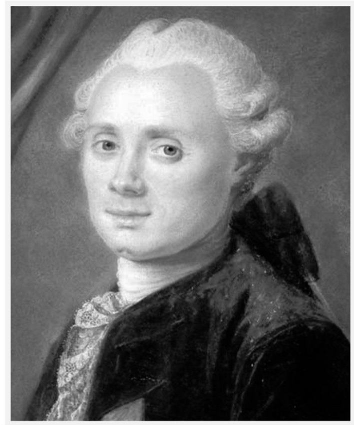
Discovering the Universe

The great astronomical discoveries of the eighteenth century were owed to the significant improvement of optical instruments since the German-Dutch spectacle-maker Hans Lipperhey had applied for a patent for his invention of the telescope in 1608.[55] With ever increasing telescope resolution it did not take long until some astronomers noticed that the night-sky not only had point-like light sources such as stars. These other objects in the night sky seemed to be quite different from stars, as they appeared dimmer, smeared out, and diffuse. The latter was the case because even though telescopes were now much better than two hundred years ago, they could not yet resolve the details of these structures which due to their shapes, sizes, and appearances became known as *nebulae*.[56] Some nebulae are visible with the naked eye if one knows where to look for them in the night skies. One can only wonder if for example the Babylonians may have already observed some of them. They kept meticulous star catalogs and star constellations such as Orion played an important role in the mythologies they inherited from the Sumerians. Within the constellation of Orion, one can for example find the Orion Nebula which is visible to the naked eye.[57] However, the Babylonians left us no such records of nebulae; and neither did the ancient Egyptians or Greeks. As we saw earlier, the Persian astronomer Abd al-Rahman al-Sufi recorded his observations of two cloudy objects in the night sky sometime around the middle of the tenth century, which are today known as the Large Magellanic Cloud and the Andromeda Galaxy. The Large Magellanic Cloud cannot be observed in most of the northern hemisphere so European astronomers would not see it. However, to a skilled observer it is visible in the night sky when viewed from places in the Arabian Peninsula such as Yemen. The second cloudy object that Abd al-Rahman al-Sufi observed, the Andromeda Galaxy, is also visible in the northern hemisphere without the help of a telescope. The Orion Nebula, one of the brightest nebulae in the night sky, is visible to the naked eye in both hemispheres, best observed during northern hemisphere winter or southern hemisphere summer. Therefore, it is not surprising that it was right after the invention of the telescope that astronomers such as the Frenchman Nicolas-Claude Fabri de Peiresc (1580 – 1637) already studied it in 1610. In the 1660's the Dutch physicist and astronomer Christiaan Huygens (1629 – 1695) invented the telescope eyepiece and with this improved telescope was able to determine that the Orion Nebula's bright interior held more than one star.[58]

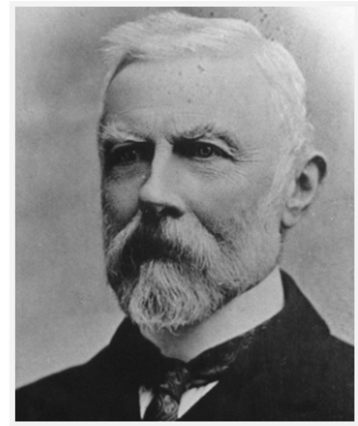
The first to formally catalog nebulae as such was the French astronomer and passionate comet observer Charles Messier. In addition to the comets, he was searching for, just like his fellow astronomers, he spotted fuzzy objects, the nebulae, which he knew were not comets. He cataloged those objects for his fellow astronomers and himself so that they would not be confused with comets. The Crab Nebula became the first object of his *Catalogue des Nébuleuses et des Amas d'Étoiles* (*Catalog of Nebulae and Star Clusters*) which was published in 1774. It included a total of forty-five objects designated M1, object Messier 1 for the Crab Nebula, through M45, object Messier 45 for the Pleiades.[59] The final Messier catalog included more than one hundred objects, among them a number of nebulae. Messier's catalog was followed in the late nineteenth century by the *New General Catalogue of Nebulae and Clusters of Stars*, originally compiled by the Danish-Irish astronomer John Louis Emil Dreyer.



Thomas Wright
1711 – 1786



Charles Messier
1730 – 1817



John Louis Emil Dreyer
1852 – 1926

This new catalog, now generally referred to as the NGC catalog, included thousands more nebulae.[60] The basis for the NGC catalog was the work of the Herschel astronomers, first William Herschel and Caroline Herschel who between themselves compiled some 2,500 objects that were various kinds of nebulae; only much later would these be differentiated into galaxies, star-clusters, or true nebulae. Later, John Herschel (1792 – 1871), William's son and an eminent astronomer in his own right, joined William and Caroline Herschel's efforts. He contributed an additional ∼1,700 nebulae objects to the Herschel catalog.
The English mathematician and astronomer Thomas Wright published his treatise *An original theory or new hypothesis of the Universe* in 1750. With it, he was the first to describe the shape of the Milky Way as an optical effect, created by the fact that our Milky Way Galaxy is a spinning disk of stars, planets, and true nebulae.[61] Because our solar system is located about two thirds out from the galaxy center, this spinning disk of stars appears to us as a band in the sky.[62] Observational confirmation came from no other than William Herschel who was the first to measure star densities in the Milky Way Galaxy and to provide a first map of our galaxy in 1785.[63] Thomas Wright's treatise inspired speculation by others about the nature of nebulae, prominently among them Immanuel Kant, the German philosopher. Less known about Kant is the keen interest he took in astronomy and that he suggested a nebula hypothesis. With it he contented that nebulae were gas and dust clouds, which under the force of gravity would collapse into spinning disks out of which stars and planets would form. Kant may have borrowed from Emanuel Swedenborg (1688 – 1772), the Swedish theologian and philosopher scientist who formulated a similar idea maybe some twenty years earlier. When later it became clear to Kant and others that some of the nebulae did have spiral arms, he believed them to be galaxies in their own right. Independently, the French mathematician and physicist Pierre-Simon Laplace developed a more detailed model of the formation of solar systems from nebulae, his *Exposition du systeme du monde* (*The System of the World*), published in 1796. However, this model, sometimes referred to as the Kant-Laplace model, could not account for the angular momentum distribution, essentially the rotational energy distribution, between the Sun and the planets. The solution of this specific problem would have to wait for the twentieth century.

Immanuel Kant
1724 – 1804

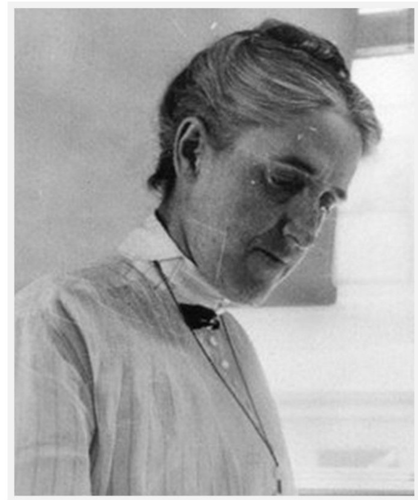
Pierre-Simon Laplace
1749 – 1827

Heber Doust Curtis
1872 – 1942

Until then, the Kant-Laplace model remained the dominant theory for the formation of solar systems as the basic ideas behind it are correct and our modern theories of solar system formation retained much of them. Were nebulae the origins of solar systems within the confines of our own Milky Way Galaxy or were they galaxies in their own right? Or rather, to stay with the language of the time, island universes as astronomers referred to them back then. Much of the confusion about what nebulae really are came about because the actual size of our universe was unknown as back then there was no way to measure how distant the nebulae actually were. Our closest sister spiral-galaxy, the Andromeda Galaxy, the star catalog object M31 or NGC 224, was already identified as a nebulous smear by the Persian astronomer Abd al-Rahman al-Sufi in the mid-tenth century. In the early twentieth century, the American astronomer Heber Doust Curtis observed a number of supernovae in the Andromeda Nebula and compared their brightness to supernova events in our own galaxy. He found the supernovae in the Andromeda Nebula to be much fainter and concluded that the spiral Andromeda Nebula was not located in our galaxy but was a separate island universe, another galaxy that is. He estimated the Andromeda Galaxy to be around 500,000 light-years away from our own Milky Way Galaxy. We now know that the distance between our galaxy and the Andromeda Galaxy is about 2.5 million light-years, about five times larger than Curtis estimated. Curtis discovery ensued a few years of intense debates among astronomers. On one side where those who maintained that the Milky Way Galaxy was the whole universe and the Andromeda Nebula was an object in it. This required the Andromeda Nebula to be small so its distance could fall within the then assumed size of the Milky Way Galaxy of about 300,000 light-years in diameter, which as we know today is about one third larger than our galaxy really is.[64] On the other side were Curtis and his supporters who believed that the Andromeda Nebula, based on the magnitude of the observed supernovae within it, must be very distant which in turn meant it would have to be a very large object, in size comparable to our own galaxy. The debate was finally settled for good in Curtis favor in the early 1920's when Edwin Hubble entered the discussion. Armed with measurement data for the distance to the Andromeda Nebula he would conclusively demonstrate that the spiral nebulae are far away galaxies.

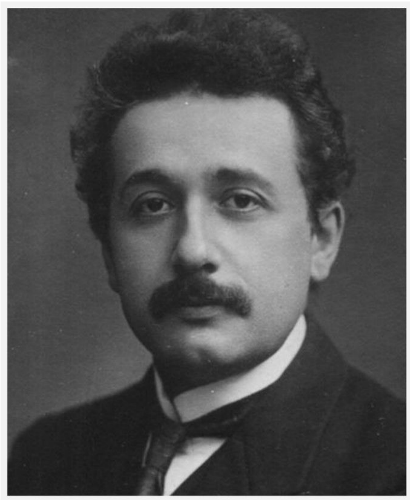
Edwin P. Hubble
1889 – 1953

The contributions of the American astronomer Edwin Hubble expanded the size of the known universe from the confines of our own Milky Way Galaxy to the size of the universe as we know it today. Edwin Hubble was born in Marshfield, Missouri in 1889 and nothing in his childhood pointed in any respect towards him becoming the most famous astronomer of his time. Hubble studied astronomy at the University of Chicago's Yerkes Observatory and his 1917 dissertation, *Photographic Investigations of Faint Nebulae*, showed where his interests lay. The opportunity to make his mark on astronomy in general and nebulae specifically arrived in 1919 with the offer of a staff position at the Mount Wilson Observatory just outside Los Angeles, when it was about to receive the most advanced telescope ever built, the 100-inch Hooker telescope.[65] Hubble started his work at a critical juncture in astronomy. By the time he began his observations at the Mount Wilson Observatory, astronomers had already compiled data and information on nebulae for decades. Part of Edwin Hubble's genius was that he trained his telescope onto objects that he knew that only he could hope to get a better understanding of because of the singular resolution power of his instrument. At the same time, he understood well that these measurements if successful would help to address some of the fundamental questions and disputes in astronomy. When in 1923 he trained the telescope onto the Andromeda Nebula he discovered that is was full of stars and more importantly, he also found that it contained so-called Cepheid variable stars. They were first discovered in the star constellation Cepheus, hence their name.[66] Cepheid variable stars are changing brightness in a very predictable way. They are pulsating stars, and their pulse frequency directly correlates with their observed brightness. The American astronomer Henrietta Swan Leavitt who worked as a *"Computer"* at the Harvard Observatory was first to discover this relationship. Regrettably, she never really received the recognition that should have been awarded this important discovery, a fate sadly shared by many women at the time contributing to science. Leavitt's discovery of the Cepheid variables in 1908 and her meticulous and extensive analysis of thousands of Cepheid variables in the Magellanic Clouds published in 1912 changed the course of astronomy.[67] The Cepheid variable ruler was just the tool Edwin Hubble needed to measure distances to the Andromeda Nebula.[68] His estimate for the Milky Way - Andromeda distance based on the Cepheid variables he found in the Andromeda Nebula was one million light-years, which he published in early 1924. This distance was evidently too large for the Andromeda Nebula to be located within the confines of the Milky Way Galaxy and so it became clear that Andromeda was itself a galaxy far away from us. Edwin Hubble's first distance estimate was twice the estimate of Curtis but still a factor of 2.5 short of the actual distance as we know it today. Over the next years, Hubble discovered Cepheid stars in other nebulae and could make similar distance estimates. By the end of the 1920's his accumulated data had convinced most astronomers that our Milky Way Galaxy is only one galaxy of very many in the universe.



Henrietta Swan Leavitt
1868 – 1921

Today, the order of magnitude estimate for the number of galaxies in our universe is approximately 1,000 billion or one trillion. The order of magnitude for the average number of stars in a galaxy is around 100 billion. For our Milky Way Galaxy the estimates range between 200 – 400 billion stars, for the Andromeda Galaxy the estimated number of stars is about twice as high. Edwin Hubble's discoveries did not only make our universe home to many more galaxies and stars, they also made it very much bigger. For a long time astronomers had used the Doppler Effect to determine if an object was moving away from us or moving towards us.[69] We are all familiar with this effect when it comes to acoustic waves, like when we hear an increased pitch from a firetruck siren as it approaches and a lowered pitch as it moves away from us. The same works for light waves where one observes the shift in the spectral lines of the optical emission spectrum of molecules or atoms, the so-called Fraunhofer spectra.[70] If a stellar object moves away from us, the Fraunhofer lines of its optical spectrum shift towards the red - longer light wavelength equals lower pitch; if it moves towards us, they shift towards the blue side of the spectrum - shorter light wavelength equals higher pitch. The faster an object moves towards or away from us the larger the observed line shifts are. Hubble studied the spectra of a number of galaxies and compared the shifts to line spectra observed in our own galaxy. This enabled him to calculate the relative speeds at which galaxies were moving with respect to each other. What he found was remarkable and completely unexpected. The farther apart galaxies are, the faster they seemingly move away from each other. There was only one conclusion, the universe was expanding in every direction, and it was expanding at increasing speed. This relationship, today known as Hubble's law, can be drawn up in a diagram where the velocity of an object as calculated from spectral line shifts is plotted against the distance measured for that object. Data entered in such a diagram clusters around a line the slope of which is the so-called Hubble constant, a measure for how fast our universe expands. More recently, towards the end of the twentieth century, astronomers were able to use other rulers than Cepheid star variables that allowed them to measure the receding speed of even farther away galaxies. The ruler they used is the brightness of so-called type Ia supernovae. Because those supernovae are known to all have the same intrinsic brightness their observed brightness can be used to measure distance. The resulting measurements not only extended the velocity versus distance data to greater distances. More importantly, they conclusively demonstrated that the expansion of the universe is accelerating. At the time Hubble made his discovery, a revolution of our understanding of time and space and consequently of our universe had been going on for almost two decades. This new way of looking at time and space and how they interact with gravity produced the first coherent models of our universe, allowing scientists to make predictions about its past and future; provided one had some additional information, such as was coming from Hubble's data.



Albert Einstein
1879 – 1955

Albert Einstein was born in 1879 in Ulm, a small city in what is today the German state of Baden-Württemberg and back then was the Kingdom of Württemberg in the German Empire. No scientist, except Isaac Newton, has changed our understanding of the universe as fundamentally as did Albert Einstein. Until his mid-twenties, Einstein's career gave little indication of what was to come and that eventually his name would almost become synonymous with genius. It all started with Einstein's so-called *"annus mirabilis"*, the miracle year of 1905. In this year Einstein published four revolutionary papers, each one of which would have made a physicist's career. The first paper explained the photo-effect using energy quanta and thereby reasserted the corpuscular nature aspect of light first identified by Isaac Newton [3]. It made Einstein one of the fathers of quantum mechanics. The second paper explained Brownian motion, the random movements of particles in a liquid, and essentially established that atoms and molecules were real [4].[71] However, the ones we are most interested here are the third and fourth paper which laid out the principles of the special theory of relativity and the equivalence of energy and matter in the famous equation $\boldsymbol{E = m \cdot c^2}$, where the symbol $\boldsymbol{E}$ stands for energy, $\boldsymbol{m}$ for mass and $\boldsymbol{c^2}$ for the speed of light multiplied by itself [5,6].
Let us look at the special theory of relativity. What is the fundamental change it brought about and how did it eventually change our perception of the universe? In his special theory of relativity, Einstein argued that the speed of light is a universal constant that must be the same regardless from which reference frame it is measured. Before Einstein came to the rescue with his unique way to think about the fundamental nature of time and space, physics had been facing a quite serious challenge that its brightest minds were seemingly unable to resolve. On the one hand, there was the fundamental premises of the constancy of the speed of light. On the other hand, there was the so-called relativity principle. The relativity principle in its simple form states that the laws of nature should be identical in two reference frames that move relative to each other at uniform and constant speed. Such reference frames that experience no acceleration, neither in linear or rotational form, are called inertial frames or inertial systems. According to the relativity principle, the laws of nature should produce the same results regardless of where in the universe they are studied as long as the respective measurements are done in inertial reference frames. Now, how could this principle be at odds with the constant speed of light throughout the universe? That is where the addition of velocities comes in. Since Galileo Galilei, physicists had used what is known as the Galilean transformation to convert coordinates in one inertial reference frame to coordinates used in another inertial reference frame that was moving at constant speed relative to the first one. It connects the coordinates in two inertial reference frames to each other by adding or subtracting the distances inertial frames travel relative to each other.
The example that Einstein frequently used to explain the conundrum one faces when adding velocities is that of a train passing through a station, moving with at constant

and uniform speed $\boldsymbol{v}$ parallel to the station platform. He considers a man in this train walking through the train compartments with a constant and uniform speed $\boldsymbol{w}$ in the direction the train is moving. How fast is the man in the train moving relative to an observer standing on the station platform? What is the speed this man moves at, as seen from the station platform? If the man in the train would for example not move for one second, how far would he travel relative to the platform? Well, within this second, he would travel relative to the platform exactly the distance $\boldsymbol{v}$, the speed at which the train is moving multiplied by one second. However, the man is also walking at the speed $\boldsymbol{w}$ towards the front of the train, so he covers in one second the additional distance $\boldsymbol{w}$. So, the total distance the man in the train covers in one second relative to an observer standing on the platform is $\boldsymbol{W} = \boldsymbol{v} + \boldsymbol{w}$. So far so good.

Then, Einstein considered a somewhat modified *Gedankenexperiment*, the German word for a thought experiment. Einstein frequently used such thought experiments to explain the essence of his thinking to fellow physicists but also to people who had no mathematics or physics background but whom he thought well capable of understanding the basic ideas involved. In this modified *Gedankenexperiment* Einstein imagined a beam of light traveling parallel to the station platform which serves again as the reference frame for measuring its velocity, the latter being of course the speed of light $\boldsymbol{c}$. Then he asked the question, what is the speed of light relative to the moving train? The answer is easy as the beam of light simply replaces the man walking in the moving train in the first *Gedankenexperiment*. Therefore, the speed $\boldsymbol{W}$ of the man moving in the train as observed from the station platform is being replaced by the speed of light, or $\boldsymbol{W} = \boldsymbol{c}$. Now using this equality and substituting $\boldsymbol{c}$ for $\boldsymbol{W}$ in the equation of the first *Gedankenexperiment*, $\boldsymbol{W} = \boldsymbol{v} + \boldsymbol{w}$, and rearranging, we get for the speed of light in the train $\boldsymbol{w} = \boldsymbol{c} - \boldsymbol{v}$, which is less than the speed of light $\boldsymbol{c}$. Now that is a problem, because this is a violation of the relativity principle as the speed of light should come out the same, regardless if measured on the station platform or in the moving train.

How can one solve this dilemma? It was clear that the speed of light had to be constant. The theory of electromagnetism required it, and all measurements confirmed it. So this was a firmly established fact. Einstein was suspecting that many of his colleagues who certainly would have been aware of the problem were rather inadvertently relinquishing the principle of relativity. Something he considered part of the bedrock of physics and was not willing to give up. Through systematically analyzing the true nature of time and space, he found that the contradiction between a constant speed of light in all inertial frames and the principle of relativity was artificial. Artificial, because scientists just had not understood the proper nature of time and space and this is what his special theory of relativity changed. Einstein's special theory of relativity resolved the conundrum of either having to give up the constant speed of light or the principle of relativity. Both are pillars of our understanding of the physical world today. However, for that we had to give up our concepts of absolute time and space.

Einstein showed us that time and space are both relative and they are not independent but make up the four coordinates of what came later to be called the space-time continuum. Our perception of reality had to change to accommodate for time dilation, the fact that moving clocks slow down, which leads to the so-called siblings-paradox. Surprisingly, instead of confusing people and being rejected, these implications of Einstein's special theory of relativity were actually embraced by the popular culture back then. Here is how it goes. Of two identical twins one remains on Earth, and one leaves on a rocket ship coasting for years in interstellar space at a uniform high speed, only to return to Earth to find his twin much more aged or even deceased if her or his trip took too long. The special theory of relativity provides the explanation for this seeming paradox. Einstein's later general theory of relativity helps the sibling on Earth to age a tiny bit slower than he would without gravity, but it cannot compete with the anti-aging effect his sibling traveling in space experiences. More than anything else, the siblings-paradox reflects a fascination of popular culture with what are seemingly science fiction aspects of science. But there is no science fiction here as many measurements have confirmed that moving watches really do go slower.

Another prediction of the special theory of relativity is length contraction. This is actually something that the Dutch physicist Hendrik Antoon Lorentz (1853 – 1928) and the Irish physicist George Francis FitzGerald (1851 – 1901) had been postulating earlier in a different context.[72] In this context, Hendrik Lorentz had come up with what we know today as the Lorentz transformations. They replace the above-mentioned Galilei transformations, which are only good approximations for adding low velocities, which is still the case for most of what matters in our daily lives. However, if we look below the hood so-to-say, some of the important technology we have come to rely on would not work without accounting for special relativity. Einstein derived the Lorentz transformations as well but in his case they were the logical consequences flowing from the assumptions of the principle of relativity and the constancy of the speed of light.

Time dilation and length contraction have measurable and real consequences and they are complementary, contingent on the inertial frame in which one considers a situation. A good and often used example is the muon, one of the two larger siblings of the electron we know today.[73] Muons are not stable and their half-life, the time after which half of the originally present muons in a sample decayed into other particles, is 2.2 µs or put differently 2.2 millionth of one second. Importantly, this half-life is as measured for the muon at rest or said differently, in the muon's own inertial frame. Muons are constantly generated in the upper atmosphere at altitudes of around fifteen kilometers. Muons have mass so they cannot move at the velocity of light but let us assume for the sake of argument that they had no mass and could travel at the speed of light. Even in that case, they would only be able to travel about 2.2 µs times 300,000 km/s, which is 660 meters.[74] So how can they reach Earth's surface where many more muons reach detectors that register their arrival than what could be expected from their short half-life? The answer is either time dilation or length contraction depending into which reference frame we put ourselves.

We will look at time dilation first. Muons are known to travel at about 99% of the speed of light. If one calculates how much slower at that speed time in the muon's rest frame should pass, a quick estimate will give a factor of about 50. That means the actual lifetime of muons as measured by an observer at rest on Earth is 50 times as long as the half-life of the muon in its rest frame. Therefore, muons can travel close to 50 times the 660-meter distance estimated above before half of them have decayed. Being able to travel about thirty-three kilometers from the perspective of an observer at rest on Earth, muons can easily reach the surface after being generated at about fifteen-kilometer altitude. Now we will look at length contraction, for which we are putting ourselves in the rest frame of the muon while it moves towards Earth at 99% of the velocity of light. For someone riding on top of a muon moving with 99% of the speed of light towards Earth the distance between the muon generated at fifteen-kilometer altitude and Earth surface is very much contracted. But by how much? Well, the fifteen-kilometer distance to Earth surface from the muon's perspective have shrunk by the same factor 50 we calculated above. Fifteen kilometers divided by 50 equals 300 meters, which is a much shorter distance than the roughly 660 meters the muon can travel in its rest frame where its half-life is only 2.2 millionth of one second. Therefore, muons have no problem reaching Earth surface as measured from their inertial frame either. The muon has become kind of a textbook example illustrating the practical consequences of the special theory of relativity. There are many more and often very practical examples that have demonstrated the reality of time dilation and length contraction since Einstein wrote his 1905 paper.

Another consequence of the special theory of relativity is somewhat harder to demonstrate but we have been at it for quite some time. Over the past century, we have built larger and larger particle accelerators to study the fundamental particles of nature and to explore physics under conditions, as they presumably existed very close to the beginnings of time when our universe was born. Einstein's special theory of relativity predicted that no massive particle can move as fast as the speed of light. The so-called rest mass is the inertial mass of an object in its own reference frame where it is at rest. As an object is accelerated to ever-higher speeds, another type of mass comes into play, called the dynamic mass. It is the mass of the moving object as observed from another inertial frame and it relates to the observed increase of its kinetic energy as the object is being accelerated. The higher the speed of the object, the less of the added increase in kinetic energy goes into increasing its speed and the more goes into adding to its dynamic mass. Eventually, all the added energy goes into increasing the dynamic mass of the object. The consequence is that no massive object can move faster than light and only massless objects such as photons, the quanta of light, can move at the speed of light. To accelerate massive objects with rest masses greater than zero to the speed of light would require infinite amounts of energy. Now we can look at what that means for the muon in the example above. Its rest mass is about 207 times the mass of an electron. But as observed from Earth when the muon moves with 99% of light speed towards the surface its dynamic mass is about 50 times its rest mass. If the muon were moving at 99.9% of the speed of light its dynamic mass would become about 500 times its rest mass, and 5,000 times its rest mass if it were to move at 99.99% of the speed of light. The closer we would bring the muon's speed to the speed of light the more its dynamic mass would increase, and it would become infinite if the muon actually could hit the speed of light, which as a massive particle it cannot.

Inertial frames, reference systems that move at uniform and constant speed not subject to any acceleration, are of course an idealization. In practice, everything in the universe is subject to forces resulting in acceleration or deceleration. Einstein knew he needed a more general theory allowing for transformations not just between inertial frames but between frames moving arbitrarily with respect to each other, including any type of acceleration.

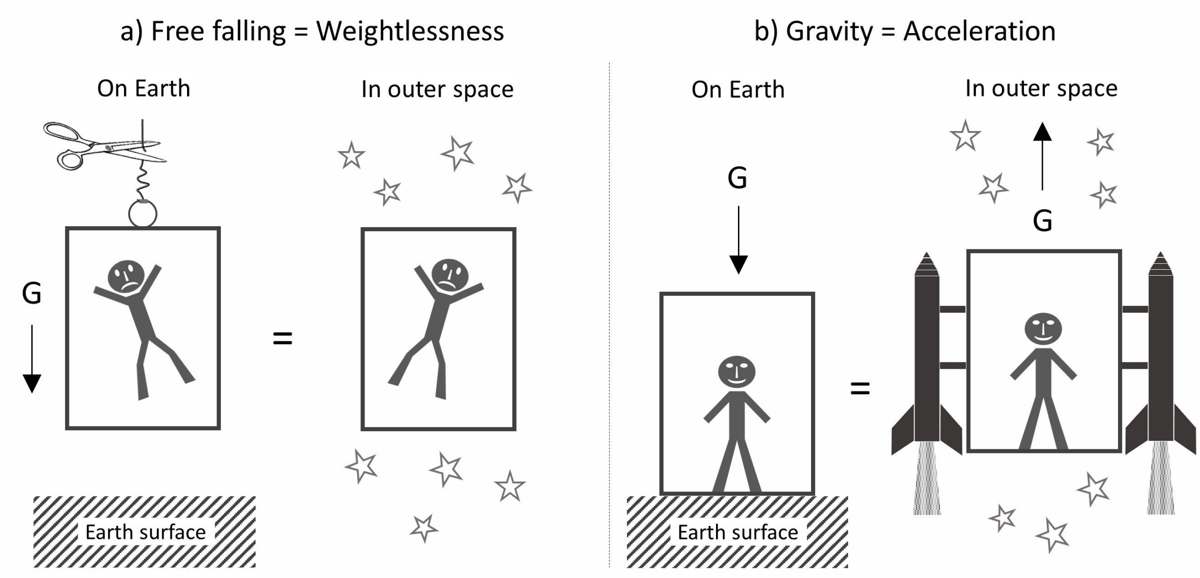


Figure 1.5: An illustration of Einstein's *Gedankenexperiments* leading him to his famous principle of equivalence. (a) A person floating weightless in a closed elevator cannot know if she or he experiences weightlessness because of the elevator free falling in Earth's gravity field or because the elevator is in outer space with zero gravity. (a) A person pinned down to the floor by gravity in a closed elevator resting on Earth's surface cannot distinguish his or her situation from a person experiencing the same force in such an elevator accelerated at one $\boldsymbol{g}$ in outer space.

Another *Gedankenexperiment* convinced Einstein that any general theory allowing for transformation between mutually accelerated frames of reference, so-called non-inertial frames, would have to be a theory of gravity. In this *Gedankenexperiment* Einstein first imagined that someone unlucky enough to find her- or himself in a free-falling elevator would experience a situation equivalent to being in zero gravity (fig. 1.5a). If this person did not know that she or he was in a free-falling elevator there was no way this person could say if she or he experienced quasi zero gravity like in outer space or was unfortunately stuck in a free-falling elevator in Earth's gravity field. Next Einstein considered the case when the elevator is not moving (fig. 1.5b). In this case, a person's feet are pinned down to the elevator floor by Earth's gravity. But then again, this situation also cannot be distinguished from an elevator in outer space with zero gravity but being accelerated in such a way and direction that the persons feet are similarly pinned down to the elevator floor. This convinced Einstein that gravity and acceleration are indistinguishable. This insight, that one cannot experimentally distinguish the effects of a gravitational field from those experienced in an accelerated frame of reference is Einstein's famous principle of equivalence.
Einstein went one step further. The next twist to his *Gedankenexperiment* was that Einstein thought about what would happen if one made a little hole in the side of the elevator to let a light beam pass through. If the elevator is greatly accelerated in the zero-gravity environment of outer space, let's call the direction upwards acknowledging there is no

up or down in space, it will travel a certain distance up before the light beam entering through the hole on one side can reach the elevator wall opposite to the hole. During the time it takes the light beam to travel this distance the elevator will have moved. Because of that, the light beam will hit the wall opposite the hole at a lower point than the hole drilled on the wall letting the light beam in. If one would now draw an imaginary line from the hole where the light beam entered, following its path to where it hit the opposite wall it would not be a straight line, but it would look bent, bent downward to stay in the picture. The result is that acceleration seems to bend light.[75] And because the first part of the *Gedankenexperiment* showed that acceleration and gravity were indistinguishable it became clear to Einstein that gravity would have to bend light also.

But how can light be bent? As it turned out, the only way to achieve this is to bend space-time. Light will always travel in what is the equivalent of a straight line in any given geometry. Because of that, Einstein's general theory of relativity is a geometric description of how space-time curves in the presence of matter and how matter moves in a curved space-time. Before Einstein published his theory in 1915, Euclidean geometry geometry had been the bedrock of all pure and applied sciences based on mathematics. In Euclidean geometry, the shortest connection between two points on a planar surface is a line and the internal angles of a triangle on a planar surface add up to 180 degrees. In non-Euclidean geometry, as for example on the surface of spheres such as Earth, the shortest connection between two points is also a line but it is a curved one; and the angles of a triangle drawn on a sphere add up to more than 180 degrees. On a playground slide, the shortest line between two points along the ride downwards is also a line but its curvature is bent different from that of a line drawn on a sphere. If we go down the playground slide neglecting any possible friction that may affect our ride, we will slide down or essentially do a guided fall along a line that represents the shortest connection on this surface between our starting point at the top of the slide and the point where we stop. In a similar way, a skydiver jumping from a plane would follow the shortest path to Earth surface if there were no atmosphere. This shortest path would trace the bent space-time generated by Earth mass. This was Einstein's critical insight. Very large masses bend space such that the space-time geometry around them has everything sliding towards their center. Einstein's revelation was that the force of gravity that we experience is nothing else than a geometric property of space-time. We and everything else in this universe inadvertently obey the law of gravity by traversing space-time in the equivalents of straight lines in any given space-time geometry, no matter how much or how little that space-time geometry is bent. To stay in the playground slide analogy, the farther away one moves from Earth the less space-time is bent and when one is far enough away from Earth or any other large mass, space-time becomes flat again and there is no more sliding. This is so to say where the top of our playground slide is located, and this is where we can push off to start the ride down the curved space-time playground slide shaped by Earth's mass. However, our playground slide does not have the sloping shape that on real playground slides gently slows down the ride. Our slide just continues straight downwards towards the center of the Earth. Which of course we will never reach as Earth's surface will stop us first, just as it does on real playground slides.

Einstein spent the decade after his 1905 annus mirabilis to understand what gravity really is and how it works. He finally published his general theory of relativity in 1915.

The theory, as its name implies, is a generalization of the special theory of relativity. It could rightly also be called the theory of gravity and sometimes is. It completely changed physicists understanding of how gravity works. It solved the critical problems that had been pointed out as inherent to Newton's laws of gravity but which physicists had been unable to address. Foremost among them was that Newton's model of gravitational attraction assumed instantaneous interaction across vast distances, essentially requiring the gravitational interaction between two masses to propagate at infinite speed. For quite some time it had been known that nothing could travel faster than the finite speed of light, so there was no practical way for Newton's instantaneous interaction to work. At the same time, Einstein's profound insights into the nature of gravity very much highlight the phenomenal achievement that Newton's discovery of the law of gravity some two hundred thirty years earlier really represents. Without knowing anything about how matter and space are intertwined, Newton's seventeenth-century law of gravity still remains a good enough approximation for how stellar bodies move throughout the universe and certainly for how apples on Earth fall from a tree.

Einstein quickly became an even greater celebrity than he already was when some of the key predictions he made to further verification of his theory showed him correct. The first verification of the theory came not from new experimental data or observation but by being able to account for something that Newton's theory could not. All planetary orbits around the Sun are affected by what is called the precession of the perihelion. This effect describes how the closest point of a planetary orbit around the Sun, called the perihelion, shifts over time. Newton's theory of gravity, and this is a testament to how good an approximation it is to reality, was able to correctly predict the magnitude of the effect for all planets, expect for Mercury. Because Mercury is so much closer to the Sun than all other planets, there are effects that Newton's theory of gravity cannot account for and therefore its predicted value for the precession of Mercury differed from the observed result. Einstein's new theory however, produced the exact value observed for this effect also for the planet Mercury. It was the English astronomer and physicist Arthur Eddington who in 1919 first experimentally verified Einstein's prediction that large masses bend light. Arthur Eddington was not only one of the leading astronomers of his time; he was also instrumental in communicating and explaining Einstein's work to the English-speaking world. After all, Einstein's work had been published in the midst of the First World War in a German language scientific journal. One way to verify Einstein's prediction of gravity bending light was to observe the apparent positions of stars during a solar eclipse in the region around the Sun and compare it with the positions of those stars as measured in the night sky when the Sun is out of the way. This is what Eddington's expedition to the island of Príncipe off the west coast of Africa set out to do in 1919. The result was clear, Newton's theory could not account for the observed shift in star positions due to their light-rays passing close to the Sun being bent, but Einstein's theory did.

Another prediction of Einstein's general theory of relativity was gravitational red-shifting. This phenomenon describes the fact that clocks in a higher gravitational field go slower as compared to the same clocks in a lower gravitational field. The clocks in the case of the gravitational red shift are represented by photons which are observed as they come to us from different parts of the universe. Photons are oscillating at frequencies all over the spectrum producing the color spectrum that we can see or measure. Higher photon frequencies translate into shorter wavelength - the blue part of the visible light spectrum - and lower photon frequencies translate into longer wavelength – the red part of the light spectrum.[76] Photons observed on Earth coming from a part of space with a higher gravitational field will have their wavelength red-shifted, as a slower ticking clock is equivalent to a lower photon frequency, which in turn means a longer wavelength and thus a shift to the red of the color spectrum. Correspondingly, photons observed on Earth coming from lower gravitational fields will have their frequency shifted relatively less to lower frequencies and thus have their wavelength shifted less towards the red side of the color spectrum. Scientists started the first measurement efforts to detect this effect in the 1920's but it took until the 1950's before gravitational red shift was finally confirmed for good. Other aspects of Einstein's theory had to wait even longer for confirmation. It was not until the development of atomic clocks before gravitational time dilation, the source of the gravitational red shift itself, could be verified directly. Measuring the passing of time with atomic clocks at different altitudes confirmed that clocks go faster at higher altitudes, which is equivalent to lower gravity. The difference is not much so you will only add a few nanoseconds over a lifetime by living at sea level instead of living at high altitudes. However, Earth gravitational field strength is low as compared to the gravitational field that other objects in the universe can generate. As we will see, there are places in the universe where the gravitational field is so immense that time stands still. In our modern age, Einstein's theory started also to have very practical uses. One of them is the satellite based global positioning system, or GPS, which most of us use on a daily basis, knowingly or not. Without correcting for the relativistic effects predicted by Einstein's theory, GPS would never have worked.

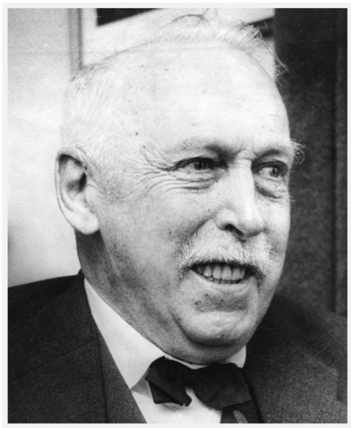
Only recently another prediction of Einstein's theory was confirmed – the existence of gravitational waves. Gravitational waves were first observed in 2015, a century after Einstein published his general theory. Like ripples in a pond, gravitational waves propagate through space-time, modulating it as they pass through. The enormous gravitational disturbances producing them are equivalents to the proverbial stone thrown into a pond that generates waves rippling through the ponds surface, away from the location where the stone entered the water. Similarly, gravitational waves are generated by giant gravitational disturbances of space-time propagating away from their point of origin at the speed of light. Extremely massive objects getting too close to each other can cause such a disturbance, thereby sending gravitational ripples through space-time. Today we know that this can happen in the merger of two black holes, first observed in 2015, or when two neutron stars merge to form a black hole as detected in 2017; how such objects like neutron stars and black holes form we will discuss in the next section. Hubble's discoveries and Einstein's theories led the way towards our understanding of the large-scale structure of the universe. Many others contributed to this effort as well, but these two scientists certainly stand out as the changes they brought about in our understanding of the universe were nothing short of revolutionary. One could certainly call the years between 1905 and 1920 a second *Copernican Revolution* in our understanding of the universe. But before we can take a closer look at how our understanding of the cosmos evolved since then we have to take a look at the life and death of stars.[77]

The Life and Death of Stars

Anyone ever having the good fortune to view the night sky on a cloudless night, far away from the lights of civilization, is likely to be amazed by the countless number of sparkling stars; and the band of stars that make up our galaxy, the Milky Way, visibly stretching across the night sky. Stars seem to be everywhere, and they seemingly shine forever. Since humans have been looking up into the night sky, they have been witnessing the birth of stars and the death of stars, unwittingly though. For a long time, our Sun was not perceived as a star. That started to change in 1838 when the German astronomer and mathematician Friedrich Wilhelm Bessel (1784 – 1846) for the first time measured the distance to a star. The technique he used is referred to as measuring a star's parallax and we will discuss it in more detail in a few paragraphs (see fig. 1.6).[78] The star he chose was located in the constellation of the swan and he calculated it to be about 10.3 light-years away from Earth, which is accurate to within 10% of this distance as measured in modern times.[79] Using this enormous distance to estimate the actual brightness of the star, it was quickly discovered that such far away stars must be as bright as the Sun. Or else, turned around, that the Sun was a star in its own right. Now, with a star at their doorstep, scientists could start to study star properties from a much closer distance than they seemingly had been able to do previously.

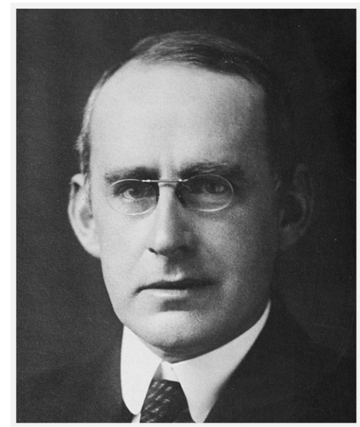
Astronomers studied stars using the instruments available to them at the time which allowed them to gather data on their color, luminosity, essentially how bright they are, size, and elemental composition by using spectroscopic techniques such as looking at Fraunhofer spectra. The accumulating amounts of collected and analyzed data eventually triggered the creation of what is probably the most important graph in astrophysics, the Hertzsprung-Russel diagram. Two scientists developed it independently between 1910 and 1912, the American astronomer Henry Norris Russell and the Danish astronomer Ejnar Hertzsprung. They found that when star luminosities are plotted against the surface temperature of stars most of them fall on a smooth curve, sloping from high luminosity or high temperature stars on the upper-left side of the graph down to low luminosity or low temperature stars on the lower-right of the graph. They had discovered the curve describing the life cycle of what today's astronomy refers to as the main sequence stars (see the Main Sequence curve in fig. 1.7).

Two physical star properties determine the position of a star in the Hertzsprung-Russel diagram, surface temperature and brightness. A third parameter used to classify stars is their color. Our star, the Sun, is a yellow star but many other stars are of different colors. The color of a star depends on its temperature, so it is really a derived parameter and often is the best and only way to measure temperature. Even with the naked eye, we can see that some stars seem to have a bluish hue and other stars a reddish one, but this is not what astronomers mean when they speak of star color. Using color filters one can much more clearly distinguish star colors than with the naked eye and this is what scientists refer to when they speak of star colors. The hotter or cooler a star is, the more its color spectrum is shifted towards the blue or red parts of the visible spectrum, respectively. Once a star's position in the Hertzsprung-Russel diagram has been determined, it also tells us something about the star's relative size and mass as compared to our Sun and how long the star is likely to exist.



Ejnar Hertzsprung
1873 – 1967

Henry Norris Russell
1877 – 1957



Arthur Eddington
1882 – 1944

Before we discuss the diagram itself - a few pages down fig. 1.7 shows the version used here - we need to get a basic understanding of the physics properties underlying it and the nomenclature used to describe it. First, this requires a brief digression into astronomy's history. Astronomers have classified the brightness of stars by how bright they do appear as seen from Earth for a long time. This measure is the so-called apparent visual magnitude of a star. It is not exactly clear who started classifying stars this way but Hipparchus of Nicaea and Ptolemy, both of whom we already encountered earlier, are frequently credited with classifying the brightness of stars by their apparent visual magnitude. The Ancient Greek word for size or greatness is *mégethos* which when used in mathematical sciences has the meaning of magnitude. The ancient Greeks started out to measure stars by their apparent size as they appeared in the night sky. The size of a star and its brightness are of course related which was unknown to the Greeks. Therefore, the Greeks got the relative brightness of the stars as seen from Earth correct, they just did not know that all those stars they assumed to be part of the sphere of fixed stars could be at very different distances from Earth and hence their absolute brightness or absolute magnitude could be quite different. The brightest stars in the night sky the Greeks classified as stars of first magnitude. The next brightest stars they then grouped into the class of second magnitude stars. The Greeks continued this classification to the dimmest stars they could see with their naked eyes in the night sky, which they then classified as stars of sixth magnitude. Translating these star classes into numbers resulted in magnitude = 1 stars, those of first magnitude, magnitude = 2 stars, those of second magnitude, and so on. Here we have the origin of a brightness classification system that assigns the lowest number value to the brightest stars and the highest number value to the least bright stars. Now, the problem is of course what to do if there are brighter stars than the stars of first magnitude. The only logical solution was to assign them even lower magnitude number values than magnitude = 1 which was assigned to first magnitude stars. For some time, this resulted in quite a few stars classified as magnitude = 0 stars but eventually the barrier broke and very bright stars with negative apparent visual magnitude values abounded. Once it was clear that the Sun was a star it also had to be assigned an apparent visual magnitude to classify its brightness. Of course, the Sun being the brightest object in the sky with a value of -26.74 has the lowest apparent magnitude of any star as seen from Earth. With telescopes becoming available, the number of dimmer stars astronomers could observe quickly increased. This led to many more stars on the positive side of the apparent magnitude scale classified higher than magnitude 6, with the dimmest observed stars having magnitudes exceeding 20 and the Hubble space telescope being able to detect the dim light from stars up to magnitudes of 30.

When the ancient Greeks first came up with six magnitude classes for stars, it was not clear how exactly those classes related to each other. Was a star of second magnitude half as bright as a star of first magnitude or was it a third, or a fourth as bright? It was only with William Herschel's systematic effort to measure star brightness quantitatively and matching data to Hipparchus and Ptolemy's magnitude groups that finally a ruler became available to underpin the apparent visual magnitude system. The ruler he defined was not of the linear type but a logarithmic ruler. In Herschel's system, an arithmetic difference of 5 in apparent visual magnitude of two stars translates into a brightness difference of a factor of 100. Or put differently, if one looks at stars with apparent visual magnitudes of 20, 15, 10, 5, 0, -5, -10 the relative brightness of two stars in that sequence would differ by a factor of 100 and a star with magnitude = -10 appears one million-million times brighter than a star with magnitude 20. Rulers measuring properties of our universe must handle enormous numbers and that is why many of them are of the logarithmic kind. The problem with apparent magnitude is that it is a relative and not an absolute measure of a star's brightness such as luminosity. A very bright star that is extremely far away will look very faint as seen from Earth and therefore would have a high apparent magnitude value. What is required instead is an absolute value for the visual magnitude of a star and not the apparent visual magnitude as seen from Earth. Determining a star's absolute visual magnitude becomes possible if one is able to measure very large distances in a precise way.

Measurements of very large distances often translate into measuring very small angles. This has been so in astronomy for a long time and is still frequently the case. Our only way to figure out how distant a faraway object is from Earth using geometric methods is to look at it from different vantage points and observe how the angle under which the object appears changes. Fig. 1.6 illustrates this so-called triangulation process of measuring distances. Figuratively speaking, the way it works is to put the object whose distance is to be determined at the peak of a triangle, the basis of which is defined by the points the astronomer chooses to make his observations from. By making this basis ever larger, astronomers can measure objects ever farther away, assuming of course a finite capability to measure small angles. As shown in fig. 1.6, the largest base that astronomers have available on Earth is the orbit of Earth around the Sun. By measuring the angle under which the star appears against the night sky at opposing points of Earth's orbit around the Sun one can use simple trigonometry to calculate the distance between the star and the Sun. Given the large astronomical distances involved, the movements of the star and the Sun relative to each other between such angular measurements separated by about six months, are negligible. Using this method very large distances become measurable indeed. To measure distance, astronomers use a definition based on a parallax angle as shown in fig. 1.6 of one arc-second also written as 1 *arcsec* or 1”.

If the parallax angle $\boldsymbol{p}$ equals 1" then per definition, the distance between the star and the Sun is exactly 1 parallax-second or $1\,pc$, also frequently referred to as 1 parsec. Simple trigonometry shows that $1\,pc$ is equivalent to 3.26 light-years.[80] The Earth atmosphere limits the resolution of Earth based telescopes for measuring stellar parallax to about 1/100 of an arc-second and hence Earth based telescopes can measure star distances to about $100\,pc$ or 326 light-years away. The diameter of our Milky Way Galaxy disk is on average roughly 170,000 light-years or about $52{,}000\,pc$.[64] Therefore, Earth based telescopes allow us to measure distances to only a small fraction of the stars in our galaxy. Space based telescopes improve this situation by a factor of 10 extending distance measurements using the parallax method to around $1{,}000\,pc$ or 3,260 light-years; but that still puts only a small fraction of the stars in our Milky Way Galaxy within reach. Using a star's parallax to measure distance was pioneered by Friedrich Bessel whom we have already met as the first to ever measure the distance to a star other than the Sun. To measure distances beyond what is possible with triangulation, astronomers developed the method based on Henrietta Swan Leavitt's work that uses the pulse frequency of Cepheid variable stars to measure distance. As we saw, this is what allowed Hubble to use a Cepheid variable star in the Andromeda Galaxy to measure the distance to our closest sister spiral-galaxy, a little more than 2.5 million light-years away from us. With this method it is possible to measure distances to galaxies that are on the order of 150 million light-years or some 46 million pc away from us. For measuring the distance to objects even farther away than that astronomers use a different candle, the light produced by stars exploding in a supernova, more specifically, the light produced in a type Ia supernova. By calibrating the relationship between distance and brightness of such supernova explosions in our galaxy and in near-by galaxies, using the parallax and the Cepheid variable star methods, astronomers can use type Ia supernova explosions observed in galaxies even farther away to determine their distances from us. The apparent brightness of these calibrated supernova explosions can then be used to measure distances exceeding 1,000 million pc. The edge of the observable universe is thought to be roughly 14,260 million pc or some 46,500 million light-years away from us so our ability to measure large distances is getting close to matching the scale of our universe. Unfortunately, as we will see later, our universe keeps expanding. Moreover, it does so at an accelerating rate. Therefore, even as we develop better and more accurate methods to measure distances, objects at the very edge of our universe will likely start to move out of reach for good.

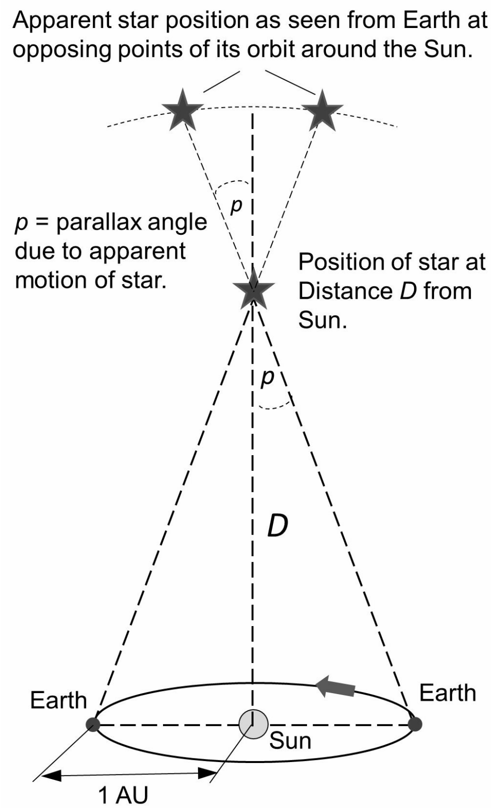


Figure 1.6: Measuring the distance of a star.

To convert the relative measure of apparent visual magnitude as seen from Earth to an absolute measure of visual magnitude, all brightness measurements must relate to a standard distance. Apparent visual magnitude is a measure of the light-flux from a star that reaches Earth. This flux of light is given by the star's intrinsic brightness, its luminosity, divided by the surface area of the virtual sphere that has the distance between the star and Earth as its radius, times the area that the solid angle under which Earth is seen from the star extends over this surface. When comparing the apparent magnitudes of two stars one has to know the value of four parameters, their respective luminosities and their respective distances from Earth. Because the intrinsic luminosity of a star does not change, converting the known apparent magnitude of a star at its real distance to the Sun to the apparent magnitude this star would have at any arbitrary distance requires the knowledge of only one parameter: the true distance of the star from the Sun. Therefore, if one can measure the true distances of stars from the Sun one can calculate the absolute visual magnitudes they would have at any other chosen distance from the Sun to get a true measure of their relative absolute visual magnitudes.[81] The fixed distance that astronomers use to do this star comparison is 10 *pc* which translates to roughly 32.6 light-years. We have used here the qualifier visual in apparent and absolute magnitudes because it is most often the apparent and absolute magnitudes in the visual spectrum that are being used, specifically for the Hertzsprung-Russel diagram. However, similar considerations also apply to other spectral bands as well as to measurements that include radiation at all frequencies, which are then referred to as bolometric magnitudes.[82]

The surface temperature of a star is a well-defined physical property but determining it is not easy for stars that are far away. Fortunately, star temperatures are directly related to their colors. Therefore, for most stars the surface temperature information comes from measuring their color properties. Scientists determine these properties by taking star color measurements in various parts of the visible spectrum. Such measurements need to be corrected for potential light absorption in the color spectrum of the star along the way as the star light travels from the star to Earth. Once the color data for different spectral emission bands are determined, a star's temperature can be extracted by calibrating its color spectrum against the known emission spectrum of a perfect black body.[83] There are also empirical methods where well established color-temperature data for a large number of stars allow scientists to derive a color-temperature calibration curve that they then can use to determine a star's temperature from color or vice versa. The data points for the selected stars shown in the Hertzsprung-Russel diagram in fig. 1.7 are plotted with their absolute visual magnitude against their temperature. We have so far discussed the star data with respect to two of the four labeled axes in fig. 1.7, surface temperature and absolute visual magnitude. The axis indicating the spectral star class is somewhat redundant because star color, as discussed above, refers back to star temperature as shown on the lower horizontal axis. The second vertical axis on the right side of the diagram providing the luminosity of a star relative to the Sun is an inferred measure and to it we will turn next.

Luminosity is defined as the total amount of energy emitted by a body at a given temperature per unit of time. A star's luminosity cannot be measured directly but requires knowledge of the star's surface temperature and of its distance from the Sun. Stars are spheres and for a sphere, the surface area scales with the sphere radius multiplied by itself. Therefore, a star that has twice the diameter of our Sun will have four times its surface area. If it has the same surface temperature the total energy radiated from its surface area, its luminosity, will be four times higher than that of the Sun. How luminosity changes with temperature is governed by what high-school students learn as the Stefan-Boltzman law. It states that the energy emitted from a black body is proportional to the fourth power of the temperature.[84] So if a star with the same size as the Sun has a surfaces temperature two times higher than our Sun it will emit $2 \cdot 2 \cdot 2 \cdot 2 = 16$ times more energy than the Sun.

By the very definition of the absolute visual magnitude M_V of a star, its absolute visual magnitude and its apparent visual magnitude are the same at the distance of $10\,pc$. If this same star where to be at a distance of $100\,pc$, a factor of 10 farther away, its M_V would be reduced by 5. If on the other hand we move this star from $10\,pc$ to $1\,pc$, reducing its virtual distance from the Sun by a factor of 10, then its M_V value would be raised by 5. Each factor of 10 distance increase or decrease will subtract or add the value 5 to the absolute visual magnitude M_V. In cases where a star has the same overall spectrum as the Sun another relationship holds that links the absolute visual magnitude M_V to the relative luminosity of the star as compared to the Sun.[85] A star with the same luminosity as the Sun has an $M_V = 4.83$. A star that has 10 times the luminosity of the Sun has an M_V value of $4.83 - 2.5 = 2.33$. A star that is 10 times less bright than the Sun has an M_V value of $4.83 + 2.5 = 7.33$. Each factor of 10 increase or decrease in luminosity relative to the Sun's luminosity will subtract or add the value 2.5 to the star's absolute visual magnitude M_V. Therefore, for stars with spectra similar to our Sun, M_V values of 10, 5, 0, -5, - 10 indicate relative luminosities of 1, 100, 10,000, 1,000,000 and 100,000,000. M_V is very useful to look at absolute star luminosity and distance changes in the Hertzsprung-Russel diagram. And this is made easier by choosing the ruler put next to the M_V axis to show increments/decrements of 2.5 in M_V as this either directly relates to factor of 10 luminosity changes; or if we take increments/decrements of 5, correspond to factor of 10 distance changes or factor of 100 luminosity changes.

How is star luminosity measured? As stated above, we cannot measure it directly, but we can determine it from two parameters of a star which we can measure: a star's size, which requires knowing its distance, and its surface temperature. We already discussed measuring a star's surface temperature, which is done through spectroscopic measurements and calculating the temperature an ideal black body would have to have in order to emit the same kind of spectrum. We also looked into how astronomers make parallax measurements to determine the distance of a star from the Sun (see fig. 1.6). Similar measurements but for even much smaller angles can be used to determine a star's angular diameter, the angle over which its disk is seen to extend from Earth. This is of course easier for very large stars such as supergiant or giant stars and very hard or next to impossible for smaller stars. For stars where the size cannot be measured, the relationships between apparent magnitude and luminosity can be used to infer star size. The final part of the vocabulary needed to make good use of the Hertzsprung-Russel diagram in fig. 1.7 is the classification scheme used to group stars into classes that share common properties. Scientific cataloging schemes develop over time as the understanding of the matter at hand improves. Therefore, it is no surprise that such classification systems, at least at first glance, may sometimes not seem very logical.

Table 1.1: Outline of the Harvard Spectral Classification system of stars.

Star Class	**Surface Temperature [K]**	**Color***
O	$\geq$ 30,000	Blue
B	10,000 - 30,000	Blue-white
A	7,500 - 10,000	White
F	6,000 - 7,500	Yellow-white
G	5,200 - 6,000	Yellow
K	3,700 - 5,200	Yellow-orange
M	2,400 - 3,700	Red
L	2,400 - 1,300	Red-brown
T	1,300 - 500	Brown
Y	$\leq$ 500	Dark-brown

**Note: Color as seen using spectral filters.*

To make matters worse, often multiple cataloging systems are developed in parallel or as alternatives to already existing ones; such has been the case for star classification. Although different versions are still in use today, the scientific community has essentially settled on using two of them. One of the two, developed at Harvard University in the late 1800's is the Harvard Spectral Classification system. The other one is the Yerkes Spectral Classification system, sometimes also referred to as the MK system after the initials of the scientists developing it in the 1940's at the Yerkes Observatory. Preceding the development of the Harvard system, stars were extensively cataloged just to keep track of them and one data point that was used to do that was the strength of the hydrogen spectral line in the Fraunhofer spectra of stars. As with all cataloging, the bins representing stars with similar hydrogen spectra had to be labeled somehow and for that an alphabetic scheme was chosen. When later it became clear that it made more sense to classify stars according to their surface temperatures the bins had to be reshuffled. This is why the Harvard Spectral Classification System does not come in alphabetical order but comes in the sequence of O, B, A, F, G, K, and M-type stars, with O-type stars being the hottest and M-type stars being the coolest in this sequence.

This milestone in the understanding of star classes that still is the basis of much of stellar classification was the work of three assistants to Edward C. Pickering (1846 – 1919), the director of the Harvard College Observatory: Williamina P. Fleming (1857 – 1911), Antonia C. Maury (1866 – 1952), and Annie Jump Cannon (1863 – 1941). Like Henrietta Swan Leavitt, these three women were part of Pickering's team of *"Computers"* to analyze huge amounts of stellar data. At the time, this was about the only kind of career open to women in astronomy and over time, Pickering employed more than eighty of them. Both, Cannon and Pickering, are credited today with the Harvard Spectral Classification system development, which was formally adopted by the International Astronomical Union in 1922. The system is of course more detailed than just the alphabetic grouping into spectral star classes. Each of the O, B, A, F, G, K, and M-type star classes is for example subdivided by adding number values to the letter ranging from 0 – 9, where a higher or lower number indicates a hotter or cooler star in a given class.

Table 1.2: Outline of the Yerkes Spectral Classification system of stars.

Star Class	Star Description
O or Ia^{+}	Hypergiants
I	Supergiants
Ia	Bright supergiants
Iab	Intermediate-size bright supergiants
Ib	Less bright supergiants
II	Bright giants
III	Giants
IV	Subgiants
V	Main sequence stars (dwarfs)
VI	Subdwarfs
VII	White dwarfs

There are more sub-classifications, but they are not of interest here. When cooler red stars were discovered and later also brown dwarf stars and methane dwarf stars, three more letters were added to the sequence designating L-, T-, and Y-type stars. Of course, classification never ends and from time to time, astronomers identify new stars that do not exactly fit into the existing classes. At some point, the astronomical community then decides to amend the classification scheme introducing new or finer classification details. There is no benefit here to go into the details of these new sub-classes, but the reader should be aware that classifications are not static but evolve over time. Tab. 1.1 shows a simplified summary of the Harvard Spectral Classification scheme, detailed enough for the purpose of understanding the Hertzsprung-Russel diagram but not too detailed to become confusing.

The Yerkes Spectral Classification system, or MKK system, was first named after the three American astronomers William Wilson Morgan (1906 – 1994), Philip Childs Keenan (1908 – 2000), and Edith Kellman (1911 – 2007) who developed it at the Yerkes Observatory. It was introduced in the early 1940's and when it was revised about ten years later, it became known as the MK system, named after Morgan and Keenan. The Harvard Spectral Classification system is one-dimensional as it derives from star surface temperatures only. The MK system uses luminosity for classification but because of that it also includes temperature. The width of the absorption lines observed as light passes through a gas - the Fraunhofer spectrum of the gas – do depend on the gas pressure. Scientists realized that the surface gravity of a star would have a similar effect on spectral lines as higher gravity at a star's surface also implies a higher pressure. This allowed them to use spectral measurements to infer relative luminosities of stars. From a star's spectrum one can already deduce its temperature. As discussed above, star luminosity only depends on two parameters, star temperature and star radius. That in turn allows scientists to determine the relative sizes of two stars if their masses are known. How does that work? To explain this let us assume that the two stars are of similar mass. If we then find in the spectrum of one of the stars more line broadening than in the spectrum of the other, we immediately know that the star with less line broadening must be bigger than the one with more line broadening. Why? The only way surface gravity for stars of similar mass can be different is for one star to be larger and hence have lower surface gravity than the other one.

Scientists refer to the spectral classes of the Yerkes Spectral Classification system as luminosity classes. The main classes are identified by Roman numerals I to V. Class I covers supergiant stars and is split into class Ia and class Ib, whereas classes II to V describe bright giant stars, giant stars, subgiant stars, and main sequence stars, respectively. As it so happens, over time, astronomers gather more data leading to new discoveries and improved understanding and eventually the original classification system needs to adopt to that. For the Yerkes system this led to a further split of star class I describing supergiant stars and has added the classes 0 or Ia^+ for hypergiant stars as well as classes VI and VII for subdwarf stars and white dwarf stars. Tab. 1.2 shows a simplified summary of the Yerkes system. As with the Harvard system, the Yerkes system also has further qualifiers to subcategorize the main classes into more detailed subcategories, specifically with respect to spectral characteristics, but that level of detail is not relevant in the context discussed here. In practice, scientists combine the two spectral classification systems from Harvard and Yerkes. The resulting short label used in a star's classification identifies its color, temperature, luminosity, and size. Our Sun is classified as a G2 V star. The translation of this designation is that the Sun is a yellow star (G-class) at the lower temperature range of its class (indicated by the number 2) and it is a main sequence star (V). Taking as another example the star Canopus we are looking at an A9 II star which means it is a star of white color (A-class), as it sits in the upper temperature range of its class (9) actually close to blue-white, and it is a bright giant star (II).

With that, we are now ready to discuss the details of the Hertzsprung-Russel diagram version shown in fig. 1.7. There are different ways to plot the data in a Hertzsprung-Russel diagram, which all have their specific advantages, but most of them look very similar to the one shown here. Some may contain more stars and details, some less. The objective here is not completeness but understanding and for that, the Hertzsprung-Russel diagram version shown in fig. 1.7 will suffice. But even at that level the diagram is quite busy containing lots of data. First, let us look at the four rulers shown around the diagram. The lower horizontal axis shows the surface temperature of a star, and it runs from a high temperature of 23,000 K on the left to 2,000 K on the right. This temperature axis is not drawn on a linear scale but a logarithmic one. The left vertical axis shows the ruler for the absolute visual magnitude of a star. As pointed out earlier, while it looks like a linear scale the absolute visual magnitude is an intrinsically logarithmic measure. The marks on the scale have been set at ± 2.5 arithmetic increments but each of these increments or decrements corresponds to a change in brightness by a factor of 10; an arithmetic change of 5 in absolute visual magnitude corresponds to a factor-100 change in brightness. Because of the nature of the magnitude ruler, a high brightness corresponds to a lower magnitude and vice versa. Therefore, the absolute visual magnitude axis runs from the lowest marker value of +17.5 magnitude at the lower end of the ruler to the highest marker of -10 magnitude on the upper end of the ruler where the brightest stars reside. The upper horizontal ruler shows the star classifications according to the Harvard Spectral Classification system with the temperature brackets of each class defined as outlined in tab. 1.1.

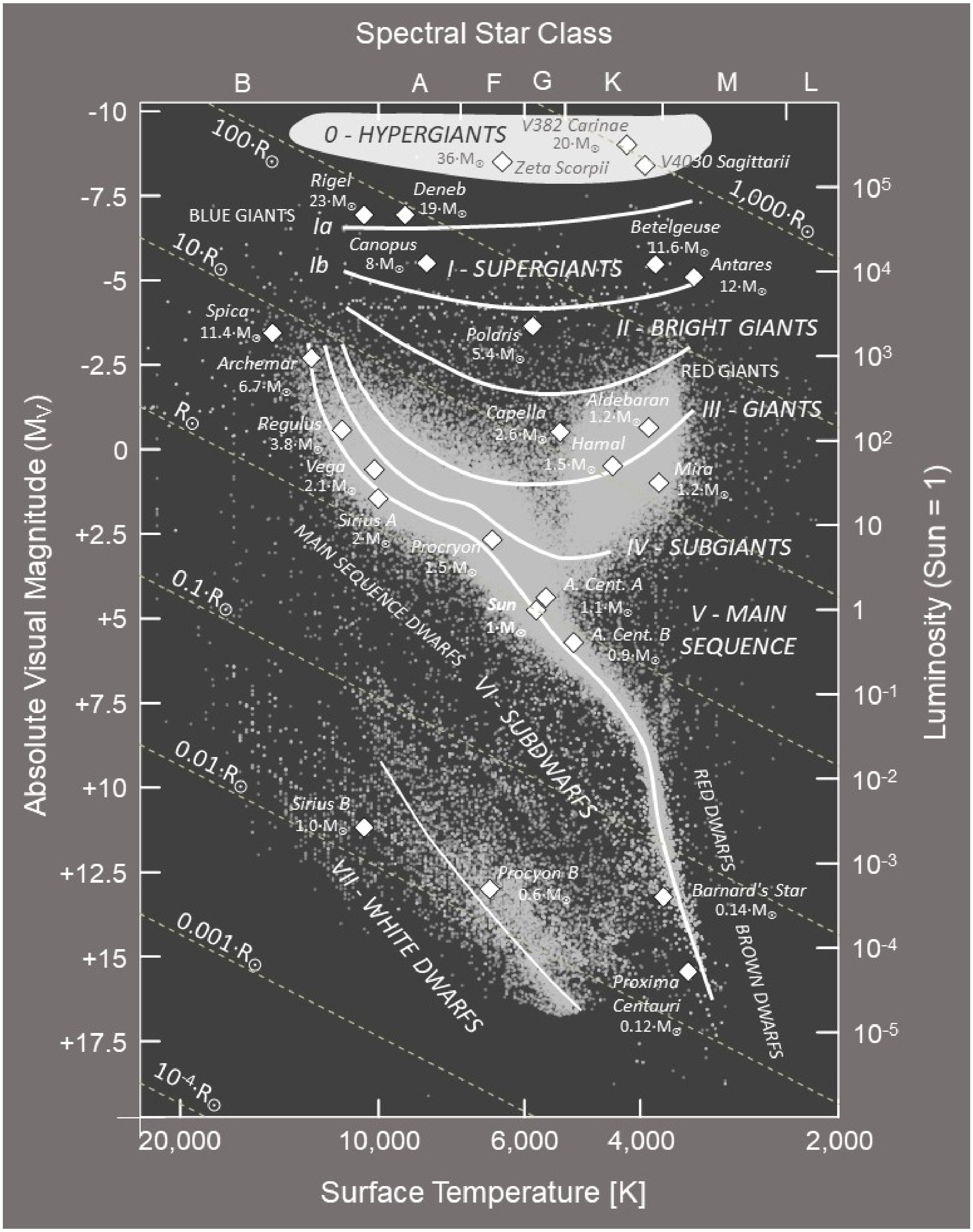


Figure 1.7: Hertzsprung-Russel diagram with close to 150,000 stars shown as tiny dots, sometimes so dense that they resemble clouds. Stellar data research for this graph used the SIMBAD database, operated at the Strasbourg astronomical Data Center (CDS), France; see Refs. [7,8].

And finally, the right vertical axis shows the ruler for luminosity, the other measure of a star's brightness discussed above. However, it does not show the absolute brightness of a star but the brightness of a star in units of the brightness of the Sun, $L_\odot$. Again, this relative luminosity scale is a logarithmic ruler. On the lower end it starts with stars 100,000 times (the 10^{-5} mark) less bright than the Sun and on the higher end it includes stars more than 100,000 times (the 10^5 mark) brighter than the Sun. In addition to the four rulers on the sides of the diagram, there is also a ruler within the diagram indicating the relative star sizes in units of the Sun's radius $R_\odot$. The marks of this ruler run through the width of the diagram from higher on the left to lower on the right side and represent lines of constant star radius in multiples or fractions of 10 with respect to $R_\odot$. Including this ruler into the diagram in the way shown here is possible because the luminosity ruler on the right vertical axis is a relative measure of a star's luminosity with respect to the Sun's luminosity $L_\odot$.[86] The many light-gray colored tiny dots in the Hertzsprung-Russel diagram represent a little less than 150,000 individual stars with their absolute visual magnitudes plotted against their temperatures. Some of the stars in fig. 1.7 are plotted as large diamonds and identified by their names. What is immediately clear from the data is that the stars are not randomly distributed over the graph. Some areas show prominent clustering resulting in dense light-gray clouds of stars where individual stars are not discernible anymore. One of these areas is the light-gray band of stars stretching from stars such as Spica and Archemar on the upper left to Barnard's Star and Proxima Centauri on the lower right. This band, highlighted with a white line going through the middle of it, is the so-called main sequence of stars with the Sun halfway in the middle of this curve. The reason this band of stars was named the main sequence is that the large majority of stars fall within this band. On the upper left side, this band includes stars much bigger, heavier, and many thousand times brighter than the Sun; on the lower right side, it includes stars much smaller, lighter, and with less than one-ten-thousandth of the Sun's luminosity.

Measuring a star's mass is not as straightforward as measuring its temperature through spectroscopy. The mass of the Sun can be calculated quite simply because Newton's gravitational force must balance the centrifugal force Earth is subject to when orbiting the Sun.[87] In a similar way, if a star has a companion and we can observe their orbital movements we can infer their masses. Many stars are indeed binary systems with the stars rotating around their center of gravity. One such pair are Sirius A and Sirius B, both called out in the Hertzsprung-Russel diagram in fig. 1.7. Sirius A is a main sequence star and one of the brightest in the night sky whereas Sirius B is a white dwarf star and much dimmer. Humans have known Sirius A since ancient times, but Friedrich Bessel only predicted the existence of Sirius B in 1844. The American Astronomer Alvan Graham Clark (1832 – 1897) confirmed the existence of Sirius B and with that also became the first to observe a white dwarf star, although, he could not know this back then. In his time, nobody knew about the existence of white dwarf stars. Sirius B, a dim companion to the much brighter Sirius A, was barely visible with the instruments available at the time. Using the orbital and rotational period data from observations of Sirius A and Sirius B one finds that Sirius A is a little more than two times as massive as our Sun, while Sirius B has about the same mass as the Sun. Similar measurements for binary systems have provided a large amount of stellar mass data.

Cecilia Helena Payne-Gaposchkin, 1900 – 1979

Eventually, this has revealed a mass-luminosity relationship for main sequence stars, which was first noted by the German-British astronomer Jakob Karl Ernst Halm (1866 – 1944): the higher the luminosity of a star on the main sequence curve, the greater its mass and vice versa. The most massive stars are on the upper left side of the main sequence curve and the least massive stars lie on the lower right side of the curve. Discovering the relationship between mass and luminosity and solidifying it through many tedious measurements since then is one of the great accomplishments of modern astronomy. In the mid-1920's, Arthur Eddington opened a new window into the life cycle of stars and as to what could potentially power their enormous energy output. Since the nineteenth century, scientists had struggled to find an explanation for the energy source that allowed stars such as our Sun to shine so brightly over very long times. Before it dawned on scientists that the Earth was much older than they thought, the problem had not existed. Now they had to look for mechanisms that allowed stars to shine not just for millions of years but for several billion years. Once this realization had sunken in, it became obvious that the then known physics held no explanation as to how the Sun could possibly shine for billions of years without running out of fuel. The longest star lifetime physics at the time could support would at most have kept the Sun shining through gravitational heating for an estimated one-hundred million years. As we already learned earlier, Eddington was very much excited by Einstein's theory and a particular feature of this theory was the equivalence of matter and energy, the famous $\boldsymbol{E} = \boldsymbol{m} \cdot \boldsymbol{c}^{\mathbf{2}}$ equation. Eddington knew that four hydrogen atoms are heavier than a single helium atom. By taking the small mass difference between one helium atom and four hydrogen atoms and multiplying it two times with the speed of light, he calculated the energy that could be set free in such a fusion reaction. He quickly realized how powerful an energy source the conversion of matter into energy was and in 1920, he was the first to propose nuclear fusion as the mechanism that powers a star's furnace. Physicists at the time did not consider anything like nuclear fusion, or fission for that matter, even as a possibility [9]. The elementary building blocks of the atomic nucleus such as the neutron were not discovered yet and most particles that are critical in nuclear reactions were still unknown. Although the energy that could be released through hydrogen fusion reactions was much greater than what any other known chemical or physical process could produce from the same amount of mass, there was a problem. It would require a lot of hydrogen to power the Sun, much more so than scientists at the time were willing to consider.

Then, in 1925, the British-born American astronomer Cecilia Helena Payne-Gaposchkin discovered that stars are mostly made of hydrogen. However, when she described her findings in her doctoral dissertation she was dissuaded by none other than Henry Norris Russell to qualify her milestone discovery as a spurious results because it went against the then accepted wisdom of astronomy.[88] However, by the end of the 1930's it was an established fact that for stars such as our Sun, most of their mass consists of hydrogen.

In addition to Einstein's theory of gravity the newly developed quantum mechanics, increasingly used by a growing community of nuclear scientists, eventually produced the understanding of what we call today stellar nucleosynthesis. This includes the discovery of the neutron, the theoretical understanding of much higher probability proton-proton interactions that would not have been possible within the classical framework, the discovery of the positron, the electrically positive charged twin of the electron, and the prediction of the neutrino without which many nuclear reactions would violate fundamental conservation laws of physics.[89]

In the late 1930's, scientists began to understand the different pathways for hydrogen to fuse into helium. The names associated with these efforts to understand the stellar nucleosynthesis of lighter elements are those of the German-American physicist Hans Bethe (1906 – 2005) and the German physicist Carl F. von Weizsäcker (1912 – 2007). Then, in 1946, the British astronomer and cosmologist Fred Hoyle (1915 – 2001) first formulated the theory of stellar nucleosynthesis including heavier elements and by the mid-1950's, the full picture of stellar nucleosynthesis had mostly emerged. The production of elements heavier than hydrogen via nuclear fusion reactions in stars up to the atomic mass of iron had been laid out.

Eddington made his conjecture that nuclear fusion could be powering stars in 1920 and reaffirmed it in his 1926 book *The Internal Constitution of the Stars* [9]. At the time, all of the above discoveries were still several years into the future. As it turned out, he was eventually proven correct. The understanding of star evolution that came with the concept of stellar nucleosynthesis is one of the greatest achievements of twentieth-century astrophysics and astronomy. Eddington also discovered the theoretical basis for the already noticed correlation between a star's mass and its luminosity. For a star to be stable, the radiation pressure produced by its hot core and the gravitational force exerted by the star's mass pressing down on the core must be balanced. This is the case if a star's luminosity is roughly proportional to the third power of its mass. Furthermore, the total energy available to a star is proportional to its mass and the rate at which it uses up this energy is proportional to its luminosity. Combining these facts allowed the prediction of a star's expected lifetime simply from its mass.[90] The result was surprising and positively counter-intuitive. A star with a much greater mass than the Sun would have a shorter life than the Sun and stars with less mass than the Sun would live longer. However, this is exactly what scientists discovered and since then has been confirmed many times over. Star clusters provide a good example of that.

Within some groups of stars, the average distance between stars in the group is much less than the average distance between stars in the Milky Way Galaxy as a whole. Such groups of stars are known as star clusters, and they are not only gravitationally bound together but they were also formed at roughly the same time in a shared stellar nursery. Many galaxies do contain such star clusters of which two major types have been observed. Globular clusters, as the name indicates, are spherical clusters of stars that can have from ten thousand to a few million stars packed in a space that can be as small as ten to thirty light-years across. More than 150 such globular clusters have been identified in our Milky Way Galaxy; the closest one is about 7,200 light-years away and was discovered in the mid-eighteenth century. Open clusters are the other major type of star cluster, they contain less stars, usually not more than a few hundred.

Table 1.3: Mass-luminosity relationship and lifetimes for main sequence stars.

Star Class	**Mass** *($M_\odot$ = Sun)*	**Radius** *($R_\odot$ = Sun)*	**Life Expectancy** *(Million Years)*
B	2.1 - 16$\cdot M_\odot$	1.8 - 6.6$\cdot R_\odot$	~100
A	1.4 - 2.1$\cdot M_\odot$	1.4 - 1.8$\cdot R_\odot$	~1,000
F	1.04 - 1.4$\cdot M_\odot$	1.15 - 1.4$\cdot R_\odot$	~3,000
G	0.8 - 1.04$\cdot M_\odot$	0.96 - 1.15$\cdot R_\odot$	~10,000
K	0.45 - 0.8$\cdot M_\odot$	0.7 - 0.96$\cdot R_\odot$	~50,000
M	0.08 - 0.45$\cdot M_\odot$	$\leq$0.7$\cdot R_\odot$	~200,000

The most well-known open clusters in our galaxy are the groups of stars known since ancient times as the Hyades, about 153 light-years away, and the famed Pleiades, with a distance of about 444 light-years from our Sun.[91,92] Most stars are not born alone but are created in star nurseries and at one time, our Sun is thought to have also been part of an open star cluster. Over time, because they are less tightly bound by gravity than the much closer spaced stars in a globular cluster, smaller star clusters can disperse due to gravitational pull from other objects. Maybe this is what happened to our Sun because within a sphere of thirty light-years that has the Sun at its center the number of stars is only around one hundred, maybe even less.

When astronomers plotted the measured star data of star clusters in a Hertzsprung-Russel type diagram they discovered something surprising - the more massive stars were missing on the main sequence curve. The curve seemed terminated towards higher mass stars and at the point where this happened, stars branched off to the upper right as if the main sequence curve had been bent and a red giant branch was developing. The point where this termination of the main sequence curve and the branching off to the upper right occurred, was not the same for different clusters. For some the missing stretch of massive stars was short and for others it was much longer. It all became clear once scientists understood that the older a star cluster was the more of its original heavier stars were missing on the main sequence curve as they entered the red giant branch.

Tab. 1.3 shows the mass-luminosity-lifetime relationship for stars on the main sequence curve where masses and star sizes are expressed in terms of the solar mass and the radius of the Sun. For most of the stars called out by name in the Hertzsprung-Russel diagram shown in fig. 1.7 the mass is given in terms of solar masses. Starting on the upper left side of the curve, we find the B-class stars Archemar and Regulus firmly on the main sequence curve with lifetimes on the order of one hundred million years. Quite a few species on Earth have been around for much longer than that. Maybe humans will one day witness the fate of those stars as they move off the main sequence curve. Among the A-class stars, we find Vega and Sirius A with lifetimes on the main sequence curve in the order of 1,000 million years. F-class stars stay a few thousand million years on the main sequence curve. After that, we enter the part of the main sequence curve where G-class stars such as our Sun reside. Here we can expect lifetimes on the main sequence curve of about 10,000 million years, like for our Sun, which is about halfway through its lifetime on the main sequence curve. The nearest star system to our Sun, Alpha Centauri, includes two G-class stars, Alpha Centauri A and B, which form a binary star system and have main sequence lifetimes similar to our Sun. As we move to the lower right part of the main sequence curve, lifetimes on the curve become extremely long. Proxima Centauri, another star in the Centauri system and our closest star neighbor, lies very much towards the end of the curve and has less than one-eighth of the solar mass. Because this star is so close, its size and mass are well known. Looking at tab. 1.3, a star like Proxima Centauri has an expected main sequence lifetime of about 200,000 million years. This is much longer than we know that our universe exists which is about 13,800 million years. Because K-class and M-class stars have such long lifetimes, many of them have been around for a very long time, some maybe as early as roughly a couple of million years after the beginning of our universe. A large percentage of the red dwarf stars on the main sequence curve and even more so of the brown dwarf stars, of which we still know very little because they are so hard to find, represent the oldest stars in the universe. Most stars we know, the estimate is around 90%, fall somewhere on the main sequence curve of Hertzsprung-Russel diagrams as shown in fig. 1.7. Of the remaining roughly 10% of stars, some fall on the blue side of the spectrum where the very bright, very large and heavy blue giant and supergiant stars reside; Rigel and Deneb belong to them. Above them, the hypergiant stars are even brighter and more massive. Then there are the supergiant and giant stars to the upper right of the main sequence with large sizes and low densities; and of course, there are the white dwarf stars below the main sequence, small stars with high densities and low luminosities. All stars moving off the main sequence curve enter different phases in the their life cycles.

For stars, mass is destiny. All stars we know started life somewhere on the main sequence curve with hydrogen fusion powering their initial phases. Very heavy stars do not stay for long on the main sequence curve. What blue giant stars have in common is that they were bound to move off the main sequence curve much earlier than less massive stars such as our Sun are obliged to do. However, that is where commonality ends. Blue giants are not a star class *per se* but rather a label applied to stars in a certain area of the Hertzsprung-Russel diagram, the upper left area above the main sequence curve. A star with a mass of about five times the solar mass will burn through the hydrogen fuel in its core very quickly and move off the main sequence curve to become a blue giant. As hydrogen fusion continues in shell layers outside what is now a helium core the star gets bigger. As the star eventually cools, its color changes towards yellow and red as it evolves into a red giant star. In the process, the star moves towards the upper right side of the Hertzsprung-Russel diagram where red giant stars live. At some point, gravitational pressure will reignite helium fusion in its core but when this energy source has also run its course there is nothing left to oppose the gravitational pressure and the star will contract. When it has contracted to a size where matter is squeezed such that electrons are getting to close to each other a new balance with gravitational pressure is found.[93] The star has become a white dwarf, having moved below the main sequence curve of the Hertzsprung-Russel diagram in pathways scientists understand but which are beyond what we can discuss here. Our north star, Polaris, is a good example for such a star. It once was a blue giant star and now is a white-yellow star evolving towards becoming a red giant; eventually, it will end up as a white dwarf.

Our Sun, a yellow dwarf star on the main sequence curve, faces a similar fate but will take a somewhat different path. When the Sun started its life cycle, its size was about 90% of what it is today, and it had about 70% of its current luminosity. Our Sun has already increased in brightness and in size and it is about halfway through its hydrogen fusion cycle with approximately five billion years left before it will start to move off the main sequence curve. At that time, about 10,000 million years into its life cycle the Sun will be twice as big as it is today, and its luminosity will have doubled. When finally, all the hydrogen in its core will have been converted into helium the temperature of the helium core will not be high enough to ignite helium fusion. At that point, the Sun will somewhat contract because the gravitational pressure exerted by the mass surrounding the core will not be balanced anymore by the radiation and thermal pressure generated through nuclear fusion in the core. The result will be that the core will heat up as it contracts under gravitational pressure and eventually will become hot enough for hydrogen fusion to start in the hydrogen rich boundary layers surrounding the helium core. So our Sun will have moved from hydrogen fusion at its core to hydrogen fusion in a shell around its helium core and it will have left the main sequence curve and entered the path to become a subgiant. As successive hydrogen layers will be used up, the helium core will grow, and the process will repeat with the hydrogen fusion layer moving into the still hydrogen rich outer layers above the growing helium core. As this hydrogen shell fusion progresses, the Sun will expand and because the outer layers of the expanding Sun will then be further away from the helium core with its hydrogen burning shell surrounding it, the outer layers of the Sun will cool. Eventually, the Sun will change its path from the almost horizontal transition it has been following towards where the subgiants reside and move onto a path leading it straight up to the red giant area. At the end of this journey, about 13,000 million years into its life cycle, our Sun will be about 2,000 times as luminous as it is today, and it will have grown to about one hundred times its current size. The final phase of the Sun's life cycle will start once its helium core has become hot enough for helium fusion to ignite. What follows will be the eventual conversion of its helium core into a carbon and oxygen core, which represent the end of the nucleosynthesis cycle for the Sun because it does not have sufficient mass to ignite carbon fusion. With the nuclear furnace of the Sun shut down for good, the Sun will blow off its outer layers. At that point, our star will take the appearance of a planetary nebula before after another ten-thousand years the remaining carbon-oxygen core of our Sun will end up as a white dwarf star.

Many stars heavier than our Sun eventually end up as white dwarf stars. However, if a star's mass exceeds our Sun's mass by about eight times, it has a very different fate and ends up as neutron star or a black hole, depending on its initial mass. Such heavy stars can burn up all their available nuclear fuel running through the complete cycle of nucleosynthesis to end up with an iron-nickel core. Without a nuclear furnace keeping gravity at bay the core of such a star will begin to contract. For lighter stars, electron degeneracy can stop a core collapse and the star ends up as a white dwarf star; not so for very heavy stars. Once a star's core mass exceeds about 1.4 times the solar mass, there is nothing that can keep it from imploding. Physicists refer to this mass limit of no more than 1.4 times the mass of our Sun for a star core to be stable as the Chandrasekhar limit, honoring the Indian physicist Subrahmanyan Chandrasekhar for his contributions to our understanding of the later stages in the lives of very heavy stars, including black holes. Collapsing star cores heavier than the Chandrasekhar limit will not be stable and end up as a neutron star or as a black hole, depending on their initial mass. As the core implosion of such an unstable star proceeds, the outer layers of the star follow suite and implode as well. This implosion happens with such force and energy that the star core heats up to a point where the atomic nuclei of iron and nickel are smashed and their protons convert into neutrons leaving the star with a neutron core.[94] At that point, an effect called neutron degeneracy, essentially the mutual repulsion of the neutrons compressed too closely, puts a sudden stop to any further implosion. The result is that the implosion of the outer layers turns into an explosive rebound, as the neutron core cannot be further compressed, thereby producing an enormous shockwave emanating from the star; this shockwave is what physicists call a type II supernova. It is in such violent explosion of a type II supernova that new heavier elements are created that a star's nuclear furnace cannot produce. These last stages in a massive star's life occur very fast. For example, look at a star with twenty-five solar masses. It takes the hydrogen fusion cycle several ten million years to burn through all the hydrogen of such a heavy star. Next, all the helium burns up to produce carbon and oxygen over roughly a million years. After that, the carbon fusion cycle produces mostly neon, sodium, magnesium, and aluminum. However, this carbon fusion cycle takes only around a thousand years or so. The next step is even shorter, as in only a few years the neon fusion cycle produces oxygen and magnesium. Once that has happened the oxygen fusion cycle runs through its supply in less than one year producing silicon, sulfur, argon, and calcium. Finally, the last stage, the silicon fusion cycle is complete in a mere matter of days producing nickel and iron. During this whole evolution, the stars temperature has increased from tens of millions of degrees to thousands of millions of degrees and its density has increased ten million times.

We have now encountered the two main types off supernovae explosions scientists have discovered so far. Type II supernovae are the result of the implosion of supermassive stars we just discussed while type Ia supernovae involve binary star systems, the stellar standard candles mentioned earlier. A white dwarf star by itself would not have enough matter to end its star life in the gravity driven implosion producing a supernova. However, it can steal mass from its binary companion until it exceeds the Chandrasekhar limit and is massive enough to undergo a supernova explosion of type Ia. As one can imagine, given the nomenclature classifying a type Ia supernova, there are also type Ib and type Ic supernovae explosions. Discussing the latter two in the context here does not add much to the more general understanding of a star's life cycle that we are looking for here. So it will suffice to remark that type Ib and type Ic supernovae are explosions of massive stars similar to a type II supernova. Let us just remember that there seems to be a number of avenues for stars to go supernova and scientists can learn much from each of these. The important fact is that all such star explosions do provide the conditions required to seed the universe with the chemical elements building the world around us, including those heavier than iron, which cannot form through nucleosynthesis in a star. Scientists call the process that can also produce the heavy elements in the supernova shockwave expanding through material that has essentially the density of an atomic nucleus, supernova nucleosynthesis. What remains after a star implosion-explosion as described above is a very small but very massive neutron star.

Susan Jocelyn Bell Burnell
born 1943

Two astronomers predicted the existence of such neutron stars in 1933, the German-American Walter Baade (1893 – 1960) and the Swiss-American Fritz Zwicky. The experimental confirmation came in 1967 from measurements by the Irish-British astrophysicist Jocelyn Bell Burnell and the British radio astronomer Anthony Hewish (born 1924). The radius of a neutron star is typically on the order of ten kilometers and its mass ranges between 1.4 and a little more than two solar masses.[95] Therefore, the density of a neutron star is on the order of half a million billion tons per cubic meter; for our Earth to have a similar density, all of its mass would have to be compressed from its current volume, a sphere with a mean diameter of 12,742 kilometers, into a sphere with roughly a 300-meter diameter. When these extremely dense and highly magnetized stars rotate around their axis many times a second, they emit radiation across a broad frequency spectrum, including strong radio pulses. Not every neutron star becomes such a radio pulsar, but every radio pulsar found to date is a neutron star. This was not known when Jocelyn Bell, a graduate student then, her thesis adviser Anthony Hewish, and their colleagues discovered the first such radio pulsar in 1967. However, within a few years it became clear that radio pulsars were neutron stars and in 1974 this discovery was awarded the Physics Nobel Prize. Jocelyn Bell was not included in the award, a decision that the Nobel Committee justified with her being a graduate student. Every way one looks at it, this was a highly dubious decision and some of the most prominent physicists said so at the time. Things may be different today but unfortunately, back then the Nobel Committee and its advisers being mostly older men could not see their way through to rewarding merit rather than seniority. Chances of receiving the prize were already diminished if you happened to be a female scientist but seminal discoveries seemed even less likely to be rewarded when you were a graduate student.

It is hard to imagine that there could be something even denser than a neutron star but there is: black holes. They form when a star is so heavy that gravitation can overcome the mutual repulsion of neutrons, pressing them together beyond the point where neutron degeneracy can stop the gravitational collapse of the star. The final stage in the life cycle of such a star is a black hole. A black hole gets its name from the fact that not even light can escape the pull of its gravity; if there would be a photon that wanted to escape a black hole, gravity would pull it back into the black hole and we would never see it. But we do not know if there are such photons at all because we have very little means to understand what is happening inside of a black hole. Surprisingly, an eighteenth-century reverend and natural philosopher, the Englishman John Michell (1724 – 1793), a man with wide ranging interests, already thought about bodies heavy enough such that their gravity would not allow light to escape. In a letter to Henry Cavendish, read in the Royal Society in 1783 and published in 1784, Michell reasoned that the gravity of a star could reduce the speed of light particles emitted by that star, which in Michell's deduction was a natural consequence of Newton's corpuscular theory of light.[96]

Karl Schwarzschild
1873 – 1916

Subrahmanyan
Chandrasekhar
1910 – 1995

Stephen Hawking
1942 – 2018

He further considered that if he could measure the reduction in the speed of light emitted by stars, this in turn would allow him to infer the respective masses of such stars. We know today that this latter conclusion was wrong because light always propagates only at one speed, a constant we call the speed of light, and the reason light is held back by a massive star is not a reduction of the speed of light but space-time being curved such that light cannot escape. However, what still stands is that Michell was first to propose the existence of what he called a dark star, we call it today a black hole, that is invisible because light emitted from it cannot escape its gravity. Michell was not the only one thinking about this. A little later towards the end of the eighteenth century, in his already mentioned 1796 publication *Exposition du Systéme du Monde* (*The System of the World*) Pierre-Simon Laplace also concluded that light could be trapped by objects with very high gravity.

Form the hypothetical existence of black holes considered in the late eighteenth century it took more than two hundred years before scientists discovered their potential existence as a solution of Einstein's equations for gravity. This happened almost as soon as Einstein had published his equations in 1915. The German physicist and astronomer Karl Schwarzschild was one of the first to derive exact solutions to Einstein's equations and he found them already in 1916. He described what scientists today refer to as the Schwarzschild radius or surface, the event horizon of a black hole, which describes the circumference around the black hole beyond from which nothing will ever reach us; and whatever crosses the event horizon into the black hole will not be able to return. Only much later did scientists discover that black holes were the remnants of very massive stars.

Once that was understood, the next surprise was that at the heart of our Milky Way Galaxy sits a supermassive black hole, like it seemingly is the case for every galaxy we know; devouring anything that comes too close to it, stars, planets, or spaceships alike. By their very nature, black holes are intrinsically difficult to observe directly. Their presence manifests itself through their bending of space-time around them. As objects move through this bent space-time, they do so on trajectories that cannot be explained unless a very massive object, though invisible, is present in their vicinity. The most recent evidence for the existence of black holes, as already mentioned earlier, came from the detection of gravitational waves here on Earth that were generated through the merger of two black holes. This merger of two black holes took place a long time ago, millions of light years away from Earth, but the gravitational ripples produced in this event just reached Earth in our time. There are only very few events we know of in the universe that can ripple space-time like that and send out gravitational waves for us to register on Earth. Our universe has become much more wondrous than any pre-twentieth-century astronomer could have imagined.

For a long time it looked to physicists as if a black hole would swallow up everything that came to close to its event horizon without anything ever making it back into the universe outside the black hole. All black holes looked much alike, indistinguishable except for their different masses, the latter reflecting the masses of celestial objects they swallowed over their lifetime. That in itself posed a problem, as it seemed as if the information stored with all the objects vanishing into a black hole would be lost forever. Eventually, scientists discovered that black holes are not complete one-way streets, where everything could go in, but nothing ever would come out.

In the 1970's, the British physicist Stephen Hawking proposed that black holes actually do emit radiation and thereby lose mass. Given enough time, black holes actually could completely evaporate through the emission of what is the so-called Hawking radiation. However, you would not want to sit around and wait for that to happen as the time it takes for a typical black hole to evaporate via emitting Hawking radiation is very much longer than our universe has existed so far. Therefore, for all practical purposes, black holes seem to live forever. What is more interesting than the seemingly endless lifetime of black holes is the fact that this Hawking radiation owes its existence to quantum mechanics whereas black holes live in the world of Einstein's general theory of relativity. It is one of the few instances where two fundamental physics theories scientists have not found a way yet to reconcile do connect: quantum physics describing the micro-cosmos of the very small and Einstein's general theory explaining our universe on the large scale. It also poses a conundrum, as this Hawking radiation carries no information about the black hole itself. Consequently, as the black hole loses mass through this Hawking radiation some of the information stored in the black hole would be lost. This is at odds with some fundamental principles of physics that assert information that existed in the past, as contained in the objects that the black hole swallowed, should be recoverable.

The study of black holes may seem for many to be just of academic interest but far from it. In all likelihood, it could give us a much better understanding of fundamental aspects of our universe and the world we live in. One really wonders, what would the ancient Greeks have thought about the existence of such strange objects as neutron stars or black holes? Even for many modern educated women and men they may seem to be more objects of science fiction than science. However, they are very real and as our technologies improve, they increasingly have become measurable objects. With respect to black holes, we have so far only scratched the surface; in a very literal sense though, as it is their event horizon, which we also call the Schwarzschild radius or surface, beyond which we cannot penetrate. However, even the little we know about black holes today has already completely changed our picture of the universe.

Since the first star was formed, countless billions more have been born. The oldest stars in our universe must have been around since the time when the first stars lit up in the early universe. We have not found one exactly that old yet, but we continue to find older and older stars. The first generation of stars had only hydrogen and helium available to them. As successive star generations followed, each of them seeded the universe with new elements, leaving behind their stellar cores collapsed into neutron stars or black holes, or lingering on as star dwarfs. With each generation of stars, the stellar nurseries, the nebulae, included more and more of the heavier elements. Our Sun is a third-generation star, and it therefore contains an abundance of chemical elements other than hydrogen or helium. The presence of these elements in the Sun reveals itself through the observation of their characteristic absorption lines, the Fraunhofer lines of solar spectra we discussed earlier. The respective elemental percentages of the total solar mass are 71% for hydrogen, 27.1% for helium, and 1.9% for the heavier elements. The balance between hydrogen and helium has shifted in favor of helium since the Sun started to fuse hydrogen into helium some 4,600 million years ago. But because the Sun is still in its hydrogen burning cycle and will be for a long time to come, none of the heavier elements present in the Sun today can have been produced in it through stellar nucleosynthesis. Therefore, the 1.9% mass fraction of the Sun that consists of elements heavier than helium must have been inherited from previous star generations. These earlier stars, as they ended their life cycles, seeded the universe with those heavier materials, including the stellar cloud from which eventually our Sun and the solar system formed. Of the 1.9% mass fraction of the Sun making up the heavier materials, 7.4% of this mass is iron, 51.1% oxygen, 21.1% carbon, 5.2% silicon, 5.1% nitrogen, 4% magnesium, 3% neon, and 2% sulfur with more than fifty other elements making up the remaining 1.1%.

The forging of new elements in a star seeds cold space with the critical elements for rocky planets such as Earth and life as we know it. We have to be a bit careful here because we have not said much about how that happens. Just think for a moment that stars ending up as white dwarfs can hold on to their heavy elements produced in their cores through stellar nucleosynthesis for a long time. So how exactly does a star disperse the chemical elements produced in its nuclear furnace through stellar nucleosynthesis? This is where the so-called planetary nebulae come in. Planetary nebulae are giant glowing shells of gas and plasma enveloping a star thereby giving it the appearance of a large planet. It was none other than William Herschel first describing them like that, nebulae resembling large planets. Planetary nebula is a misnomer because it is a shell of gas and plasma formed during a stars red giant phase once it has shed its outer layers. Planetary nebulae form in the late stage of a red giant star, last only for a few ten-thousand years, and are the mechanism by which a star disperses some of the chemical elements produced in its nuclear furnace throughout its lifetime.

Even once a star has moved from the red giant phase to the white dwarf phase it still can disperse the remaining chemical materials in its dense core into the universe. However, for that it needs help from a companion star. In a white dwarf star nuclear fusion has stopped and what is holding gravity at bay is electron degeneracy.[93] As we saw earlier, this only works as long as the mass of a white dwarf star does not exceed the Chandrasekhar limit of about 1.4 times the mass of our Sun. By accretion of mass from a companion star, a white dwarf star can eventually exceed the Chandrasekhar limit.

When that happens, the star's electron degeneracy pressure is not sufficient any more to hold back gravity. The result is the brief reignition of the star's nuclear furnace leading to a runaway nuclear fusion resulting in a type Ia supernova. Inadvertently, the punishment for a white dwarf star stealing mass from a companion star is immediate death once its mass exceeds the Chandrasekhar limit. The reward for the universe is the dispersion of all the stars chemical elements into the universe. Importantly for astronomers, because all such type Ia supernovae explosions have the trigger limit of about 1.4 solar masses, they make for ideal standard candles which they have successfully used to learn much about our universe. A white dwarf star without a companion star faces a different fate. Astronomers have a concept they call a black dwarf star, but they have not found such a star yet. If they should ever discover a black dwarf star, it would be the stellar remnant of a completely cooled off white dwarf star no longer emitting much radiation. However, for a white dwarf star to get to this point in its life cycle takes longer than our universe has existed. Because of that, it is unlikely that such lonely white dwarf stars ever will disperse their remaining chemical elements into the universe. All of the chemical elements except hydrogen and most of helium come from stellar nucleosynthesis and supernovae nucleosynthesis. The materials from stellar nucleosynthesis include all elements up to iron, distributed either through the planetary nebulae of late red giant stars or through explosions of white dwarf stars in type Ia supernovae. The latter also produce heavier elements than iron as do supermassive stars when they die in type II supernovae explosions. Similar to type Ia and type II supernovae, type Ib and type Ic supernovae explosions contribute as well to the dispersal of elements heavier than iron.

In the first half of the last century, scientist analyzing light spectra from interstellar clouds detected carbon-containing molecules in them. Today, we know that complex chemistry in stellar and interstellar clouds may include the formation of some of the basic building blocks of life, amino acids. It looks like complex organic molecules are floating through interstellar space in our galaxy and that may also be the case for amino acids themselves. We certainly know that meteorites can contain amino acids as scientists have found them in some meteorite samples. Generations of dying stars enriched interstellar clouds with new elements, making them the nurseries for future star systems with rocky planets like Earth, maybe even the nurseries for the chemistry of life itself. As we will see later, nature has other pathways that can produce amino acids, the building blocks of life. However, knowing that maybe the building blocks of life themselves and not just the atoms heavier than hydrogen and helium are owed to the life cycle of stars certainly does give us a very different appreciation for what we perceived for a long time as the cold lifeless vastness of outer space. In a very literal sense, life may be born out of stardust, and we certainly are made out of stardust. What holds true for the stars is also our fate only that our human life cycles are so much shorter. The universe is a much stranger place than we ever could have imagined. What scientists have learned about the life cycle of stars is nothing short of amazing. The study of stars and the universe itself have become their own branches of science and to close this chapter we will briefly go back and look at the universe at large, what we believe today we know about its past, present, and what we see in its future.

Our Universe at Large

For most of our recorded human history, dating back to the first inscribed clay tablets some 5,000 years ago, the perceived size of the universe we live in barely changed. The inquisitive ancient Greeks were foremost in giving order to our understanding of the universe. They did this so well that from antiquity to the Renaissance the accepted cosmological model continued to be essentially Aristotle's universe, including the seven then known planets, the Moon counted as one of them, and the sphere of the fixed stars as the outermost shell with our planet Earth firmly residing at the center of this universe. Then, with the *Copernican Revolution*, Earth lost its prominence of place to the Sun in a matter of decades. Following that, Kepler's laws gave us a better quantitative understanding of planetary motions and Newton's law of gravity enabled us to explain and predict the motions of celestial objects in the sky. Newtonian physics combined with improved telescopes rapidly increased our ability to observe, measure, and understand the heavens. This led to the discovery of all of our solar system planets and once it was clear that our Sun was just one among countless other stars, astronomers turned their attention to the stars in our Milky Way Galaxy and to the mysterious nebulae. In the first decades of the last century, this eventually culminated in Hubble's confirmation that the Andromeda Nebula really was a separate galaxy and his discovery of an expanding universe. We realized that just as our Milky Way Galaxy has countless stars, the universe is home to countless galaxies. Only a few hundred years after we had removed the Earth from the center of our universe, we finally started to comprehend that we live in an unimaginable enormous and expanding universe.

Parallel to these paradigm changes in observational astronomy in the first decades of the last century, a revolution in physics took place. On the cosmological scale it came from the insights provided by Einstein's general theory. On the very small scale, it came from the quests to understand what constitutes matter and the physical laws governing it, both of which rapidly increased our understanding of the sub-atomic world. Eventually, a good part of cosmology and particle physics became two sides of the same coin. First nuclear physics and then particle physics would in turn help us discover the physical laws that underlie the nature of stars, how they are born, live, and die, eventually providing us insights into how the early universe may have evolved.

In the first half of the last century, scientists struggled for quite some time to come to terms with the nature of our universe. Was it expanding, steady, or contracting? If it was expanding, would it do so forever or would it eventually stop expanding and start contracting? More recently, scientists have been pondering the existence of not just one universe but of multiple universes, a multiverse. How strange and far removed from the reality of everyday people this seems to be; and how confusing and irritating it must be for those convinced they hold the truth about the universe in their religious texts. Despite the revolution Einstein brought about in physics and contrary to many commonly held beliefs about him, at heart, Einstein was a conservative physicist. A steady state universe that was neither expanding nor contracting seemed most attractive to him. Therefore, when his initial cosmological model produced a contracting universe, he introduced a new constant, later called the cosmological constant. This new parameter counteracted gravity and prevented his universe from contracting, making for a steady state universe.

In essence, Einstein had introduced the equivalent of negative gravity or anti-gravity. When it became clear that we live in an expanding universe, Einstein called this his greatest blunder. He could not know it then, but the jury is still out on that and in a few paragraphs we will see why. At the time, others had offered Einstein solutions to his equations and all of them were describing an expanding universe. In 1922, the Russian physicist Aleksandr Friedmann (1888 – 1925) showed that Einstein's equations were suited for an expanding universe. The Belgian astrophysicist and priest Georges Lemaître came to the same conclusion in his 1927 thesis and actually sought to convince Einstein in a meeting that same year. However, Einstein was not convinced as his response to Lemaître purportedly was: *"Your calculations are correct, but your physics is atrocious"*. Lemaître was also ahead of his time with his thoughts about a *Big Bang*, for which he used a different name; he referred to it as the *Primordial Atom*, the seed from which our universe evolved.[97] He knew what he was talking about when he argued for an expanding universe as he had derived Hubble's law himself and estimated Hubble's constant before Hubble had published his results. When Hubble's results finally became public towards the end of the 1920's, Einstein gave in and finally adopted an expanding universe in 1931. Together with the Dutch physicist Willem de Sitter (1872 – 1934), he proposed a cosmological model for an eternally expanding universe in 1932. This Einstein-de Sitter universe eventually became the standard cosmological model that remained in use until the middle of the 1990's. However, without Einstein's cosmological constant.

Georges Lemaître
1894 – 1966

In the second half of the last century physicists continued their exploration of the fundamental nature of matter into ever-smaller components. To study those smaller particles which made up protons and neutrons or other particles that like the photon of the electromagnetic field are the intermediaries of the nuclear or weak forces required the building of ever-larger instruments. The reason for this is simply that in order to detect those particles one needs ever-greater energies to smash up elementary particles hoping that the searched for particle will be contained in the debris generated by such high energy collisions. Building large particle accelerators that could smash protons into each other at gigantic energies got physics ever closer, if only for an infinitesimal short moment of time, to producing conditions that may have prevailed in the very early universe, when the universe was just minutes or seconds old. In this way, particle physicist's efforts to understand the fundamental constituents of matter and the forces of nature that govern them, aligned with efforts of astronomers and astrophysicists to understand the early nature of our universe and its origin.

With the universe expanding in all directions, it was obvious that there must have been a point in time when the universe was much smaller than we observe it today. The consequence of this is of course that if we take this analysis through to its logical conclusion, so-to-say run the movie of the expansion of the universe backwards, then everything would converge in one common point of origin some 13,800 million years ago. Here was Lemaître's *Big Bang*. As we saw earlier, Cecilia H. Payne-Gaposchkin discovered in the mid-1920's that stars were mostly made out of hydrogen with helium coming in as a distant second and other elements accounting for only a small fraction of a star's mass. As most of our universe is made of stars and the interstellar clouds which birth them, it did not take astronomers too long to establish that about 98% of the total mass in our universe is made of hydrogen and helium and only 2% is made of the stuff that accounts for the bulk of a rocky planet such as our Earth. This meant that any *Big Bang* theory for the origin of our universe had to explain the relative abundance of hydrogen and helium we find today as well as their relative mass fractions when the universe was still infinitesimally small.

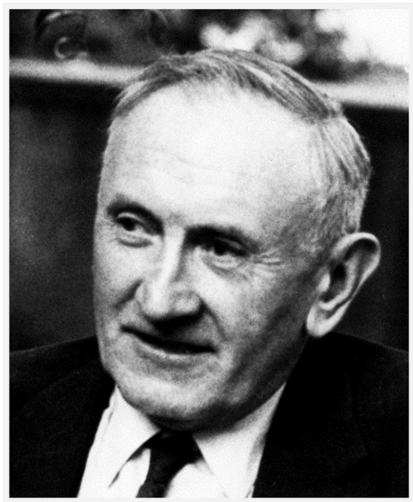
In 1948, The American physicist Ralph A. Alpher (1921 – 2007) and his doctoral adviser the Russian-American physicist George Gamow (1904 – 1968) were first to formulate such a *Big Bang* theory.[98] A couple of years earlier, Fred Hoyle had just published the first of his foundational ideas towards understanding the process of stellar nucleosynthesis. Hence, the part of Alpher's and Gamow's model explaining the formation of heavy atomic nuclei would become obsolete and some of the mechanisms they assumed to lead to the formation of the first atomic nuclei and helium would also be revised. However, their model did correctly predict the relative fractions of hydrogen and helium present in the universe only a few minutes after the *Big Bang*, the preponderance of these chemical elements, and their relative abundances we still find reflected in our universe today. The same year, Alpher and his American physics colleague Robert Herman (1914 – 1997) made another critical discovery. Alpher and Herman were first to understand that the *Big Bang* should have left a measurable imprint in our universe in the form of a universal background radiation. Because of our universe expanding for thousands of millions of years, they calculated that this background radiation should be detectable in the microwave frequency spectrum. It would take quite a while before the significance of this insight, the theoretical prediction of what we call today cosmic microwave background radiation, would become clear.

In the 1960's, the American physicists Arno A. Penzias (born 1933) and Robert W. Wilson (born 1936) were setting up a microwave receiver to detect radio signals bounced of balloon satellites in the stratosphere. While eliminating all background noise they discovered a signal that despite their best efforts they could not get rid of. This signal stayed the same regardless towards which direction in the sky they oriented their instrument's microwave receiver. Eventually, after discussing this observation with some of their astrophysicist colleagues who were in the process of preparing experiments just for measuring such a radiation, they realized that they had discovered the faint afterglow of the *Big Bang*.

When the early universe was very small it was not transparent for light. Only after it had cooled sufficiently so that electrons could combine with protons to form stable atoms did light have a chance to escape. This is thought to have happened sometime between 240,000 to 300,000 years after the *Big Bang*. It was at that time that the first light emerged from the expanding universe. Penzias and Wilson detected the afterglow of this first light flash. That we do not detect this flash of light at the wavelength it was emitted back then is because of the continued expansion of the universe stretching the space in which the light was propagating and thereby shifting it to longer and longer wavelength until today we can detect this afterglow in the microwave region of the light spectrum. The equivalent temperature of this microwave radiation is now about 2.7 Kelvin, about -270 or -455 degrees on the Celsius or Fahrenheit scales, respectively; just what was predicted from models that assumed a very hot early universe originating in a *Big Bang* 13,800 million years ago. In a sense, this microwave afterglow gave us an equivalent to a baby picture of our universe.

For most of us, exposure to the universe may be limited to sunbathing, stargazing, or the occasional science fiction movie and it may be difficult to understand how scientists are able to speculate what happened during the very first moments of our universe. When physicists and cosmologists talk about the first moments of our universe, they are not just referring to the first minutes, or just the first seconds, but the first billionth of the very first second. They derive their insights into these very first moments of our universe from theoretical models. There are a number of competing models and all of them make of course assumptions, as our knowledge about the prevailing conditions in these first infinitely short moments of our universe is quite sparse or nonexistent. However, those model assumptions have consequences and that is what separates good theories from not so exciting ones. Good theories make measurable predictions, even if they are only indirect predictions. Any theory of the early universe has to account for how the universe looks today. There is a limited number of possibilities for any such theory constrained by the laws of physics as to how our universe could have arrived at its present shape. Therefore, it is well worth listening to scientists when they explain to us how they think our universe got started. One of the first such accounts written for a broad public readership, walking us through the first three minutes of our universe, came some thirty years ago from the pen of the American physicist Steven Weinberg (born 1933). Even though science has evolved since then, it still very much captures the sense of being able to understand one of the most fundamental things in nature, how nature came to be.

As exciting as all these new discoveries and the knowledge gleaned from them were for scientists and the broader public alike, cosmologists and physicists were in for a surprise. Just as it looked like our scientists were able to explain how much of our universe worked, even back to its very beginnings, it became clear that we are still far from understanding how our universe really works. Any scientist living at the beginning of the twentieth century would have been amazed if they could have known what their heirs had learned by the close of the century. It is astounding just how much we have learned over a few generations and it remains a tremendous achievement. However, as it happens frequently, the more we know about something the more it often becomes clear how much we actually still do not know. The signs had been there for some time, and it was a problem of matter, literally. Hubble estimated the number of galaxies to be around 100,000 million; today's estimate puts the number of galaxies in the trillions. An order of magnitude estimate for the average number of stars in a galaxy is around 100,000 million or 100 billion stars; our Milky Way Galaxy has anywhere between 200 and 400 billion stars. However, even with all this matter we find in our universe, including generous allowances for substantial errors in underestimating the matter we can detect, it was clear that the universe was missing a substantial amount of matter. How did scientists know that? Well, they simply had no other choice than to draw this conclusion.



Fritz Zwicky
1898 – 1974

Where we stand today, physics cannot explain much of our universe unless it contains a lot more matter than scientists can account for. Already in the late nineteenth century, Lord Kelvin (1824 – 1907), the Irish-born physicist and engineer of Scottish descent, speculated that our universe must contain more matter than the visible matter we can see.[99] Using the velocities at which stars move in our Milky Way Galaxy he estimated its mass and found a significant discrepancy with the mass of our galaxy estimated from its visible mass. The first to infer the existence of what we now call dark matter were in 1922 the Dutch astronomer Jacobus Kapteyn (1851 – 1922) followed by his countryman and fellow astronomer Jan Oort (1900 – 1992) in 1932 as well as Fritz Zwicky in 1933. Galaxies in a so-called galaxy cluster do not orbit around a central object such as our planets do around the Sun or how the stars in our galaxy swirl around the supermassive black hole at the center of the Milky Way. The enormous galaxy masses in such a cluster warp space in and around the cluster, thereby determining the trajectories through space that each of its galaxies follows. Galaxies in such a cluster are gravity-bound to each other and the speeds at which they move are indicators for the strength of the gravitational fields they are subject to. More importantly, in order for galaxies to stay in their observed trajectories they cannot move faster than a certain speed because otherwise the gravitational field would not be strong enough to keep them within the cluster. This is where Fritz Zwicky's observation of galaxy movements in the so-called Coma Cluster with about 1,000 identified galaxies came in. They allowed him to estimate the total mass of the Coma Galaxy Cluster, which he then could compare to his mass estimation from the light emitted by the trillions of stars in the galaxies of this cluster. He found that the mass calculated from emitted starlight could not account for a large amount of the Coma Galaxy Cluster mass he had derived from galaxy movements within the cluster. Fritz Zwicky called this missing matter *"Dunkle Materie"* which is German for dark matter; the name stuck. At the time, either astronomers did not take much notice of this observation, or they seemingly did not worry about it. However, even though it took a while, this was to change.

The next firm indication that put astronomers on notice that there was something seriously amiss came from the systematic study of galaxy rotation curves, pioneered by the American astronomer Vera C. Rubin in the 1960's. In such curves, astronomers plot the orbital velocities of stars rotating around the galaxy center against their radial distance from it. In our solar system, in accordance with Newtonian physics, the orbital velocities of planets decrease with their distance from the Sun. Of course, astronomers expected to find the same behavior for stars orbiting around their galactic centers. However, they did not. Instead of decreasing, orbital velocities of stars increase the farther out they orbit their galactic center. Not all galaxies show this behavior to the same degree and there seem to be a few for which star velocities match what one would expect from the estimated visual mass contained within the galaxy.

Vera C. Rubin
1928 – 2016

However, for many galaxies their galaxy rotation curves clearly indicate that they could not maintain a stable disk of stars if they only contained their visible mass. Without the presence of dark matter, such galaxies would disintegrate or at least lose a good many of their outermost stars depending on the specific galaxy rotation curves. How much matter is actually missing in our universe? As it turns out, from what we know today there must be roughly 4.6 times as much dark matter as there is visible matter in the universe. The next question is of course, what is the nature of this dark matter? Well, that question is quite difficult to answer because the only way dark matter manifests itself is through its gravitational interaction. Just like the mass of a star can bend space and therefore bends light so can dark matter. This is what astronomers observe in the night sky in what they call gravitational lensing. To use a simple analogy, consider a transparent glass of water giving us a distorted view of what is behind it when we look through the glass. In similar fashion, dark matter can act as a gravitational lens distorting space in such a way that we can detect objects even though other objects in front of them block the direct line of sight. Because of that, astronomers can observe the distorted images of galaxies hidden behind other galaxies. Systematically studying effects like gravitational lensing has given astronomers an idea about the distribution of dark matter in our universe. Seemingly, dark matter provides the invisible scaffolding of a superstructure along which galaxies align and cluster. While we start to understand the distribution of dark matter in our universe, we do not yet understand its origin and worse, we do not know what it actually is. For our normal matter, the stuff from which everything around us including ourselves is made of, we know what its constituents are. Namely, atoms having electrons and a nucleus, the latter containing protons and neutrons which in turn are made of smaller particles called quarks; in short, we know the members of the particle zoo of normal matter. For dark matter, the only fact we really know is that whatever it is made of, it follows the laws of gravity.

For some time scientists had speculated whether so-called primordial black holes could make up a substantial part of that dark matter. Primordial black holes, as the name insinuates supposedly created in the early universe, could also have been the seeds of the supermassive black holes that we find at the centers of galaxies. However, primordial black holes, if they ever existed, would have been much smaller and lighter, the most massive ones having much less mass than our Sun has today. Like all black holes, they would have been subject to evaporation through Hawking radiation as discussed in the previous section and these much lighter primordial black holes, if they then existed, would have evaporated long ago. Sure, they still could have seeded the much more massive black holes that we see today but we do not know this, it is pure speculation. Therefore, primordial black holes cannot help us explain what dark matter is. What we do know is that supermassive black holes are part of normal matter as they swallowed all those protons, neutrons, and electrons, which make up our regular or known matter.

Primordial black holes cannot help us explain where the dark matter we believe our universe contains comes from nor can they account for the large amount of it. As if not knowing what the bulk of matter in our universe actually is and where it comes from was not bad enough, worse was to come.

As was briefly mentioned earlier, in the late twentieth and early twenty-first century it became clear that the universe was not only expanding but also that it was doing so at an accelerating speed. That was a serious problem. Cosmological models could explain a universe where the expansion would slow, even were the universe would reverse expansion and start contracting but they held no explanation as to why the expansion of the universe was accelerating. When scientists say that the universe is expanding and that it does so at an accelerating rate what they really mean is not that the galaxies are somehow picking up speed and moving away from each other. What they are saying is much more profound and confounding. It is the space between galaxies, which seems to be expanding and it does so at an ever-increasing rate. We better let that sink in for a moment.

Maybe a simple analogy helps. Just consider two ships at sea within sight of each other, their engines shut down and not moving at all because we know that there are no currents. However, to their consternation, the crews on the two ships still observe that they move away from each other because seemingly the ocean itself between them expands. Or, to illustrate this for yourself, take a marker, and put two dots next to each other on a balloon inflated just enough so you can do so. Then blow the balloon up even more and watch those two dots moving away from each other. Something like that seemingly is happening to the space that separates galaxies from each other. To expand space requires of course some kind of force or energy and because physicists do not know what this energy really is, they call it dark energy. One way to correct cosmological models would be to reintroduce a version of Einstein's cosmological constant. Maybe, what Einstein thought to be his biggest blunder, may not have been a blunder after all as a modified version of his cosmological constant can reproduce the effect of dark energy pushing our universe apart. However, this is a big maybe because the honest answer today is that scientists do not know what causes the expansion of our universe to accelerate.

We seemingly do not yet understand fundamental properties of our universe. In a sense, this dark energy seems to act like negative gravity, but that perception is likely flawed as it implies massive objects being pushed apart while it is space itself that is being pushed apart. Is it not astounding that the vacuum of space where there should be seemingly nothing contains a force that is driving it apart and with it our universe? It certainly looks like as if we do not understand the nature of space in our universe. That brings us to this question: how big is our observable universe? Most would venture that it must have the equivalent radial extension of 13,800 million light-years as this is the farthest our universe could have expanded at the speed of light in each direction. But that is not so, our observable universe extends some 46,500 million light-years in each direction. Seemingly, the expansion of space itself has outrun light, expanding space by more than two times the distance that light can cover in 13,800 million years. This shows us that we really do not know what this dark energy is. We have to be careful in how we think about our expanding universe and how it relates to what we can observe.[100]

It is in our human nature to name things even if we do not know anything about them. We have done so since times immemorial, just think about the atom. The concept of it has been around for more than two millennia but it is only a mere hundred years ago that we started to understand what the atom really is. Therefore, the concepts of dark matter and dark energy may be around for quite some time before we really understand the physical realities behind them. It is disturbing just how much of our universe we do not know yet. Until a few decades ago, it looked like we knew a lot about our universe; certainly not everything but more than enough that we could hope to have understood much of how it works. Not so anymore. If we look at the total distribution of energy in the universe and transfer this data into a pie chart, then there would be three segments. The normal matter that we mostly understand accounting for about a 5% slice of our universe; the estimate for dark matter making up about a 23% slice of our universe; and finally, dark energy filling up the remaining 72% slice of our universe. In short, even with all the achievements of science over the past couple of centuries we know still nothing about much of the universe, about roughly 95% of it. That may be sobering for some but hopefully it will be exciting for many others as there is so much left to discover.

There seem to be two ways out of this conundrum. If our estimates of the amounts of dark matter and dark energy are correct, then the challenge is to discover physically what they are, just as we did with normal matter. Indeed, there are experiments underway today that try to detect dark matter particles and others that seek to create them. The first is extremely challenging. Dark matter does not interact with normal matter other than through gravity so the only way to detect it is in head-on collisions with atoms. Scientists have set-up such experiments hidden under mountains or in deep mines as otherwise, they never would be able to discover such weakly interacting particles in the experimental noise created by the high-energy radiation Earth's surface is exposed to. The other approach, to create them in extremely high-energy collisions, is not much easier and requires the infrastructure of particle accelerators like at CERN in Europe. Alternatively, if we cannot detect or create dark matter, not to speak of dark energy, we may have to acknowledge that some of our physics theories are incomplete. Much of our understanding of space and time that shapes current cosmological models we owe of course to Einstein's general theory. Einstein's theory has been correct in many instances, but physics theories are never really complete. Just consider Newton's law of gravity. It worked excellently for a long time and still does, but some of what we observe in nature it cannot explain. This is where Einstein's theory came in. Maybe there are improved or alternative physics theories that in a similar way can predict everything that Einstein's theory can but at the same time help us to understand much more. Maybe such theories will eventually enable us to explain our universe as it is without having to resort to assuming huge quantities of dark matter and dark energy. It could also be a combination of both with new technologies allowing us to detect, measure, and quantify dark matter and dark energy, as well as new physics theories that do not require such large amounts of them to describe the reality of our universe.

If our current physics theories hold and we got the amounts of normal matter, dark matter, and dark energy correct, there are consequences. One of them is that the observed accelerating expansion of our universe is for real. With that, the universe will one day be a much darker place. Earth will be long gone in some 100 billion years when in a manner of speaking, the lights will turn off. Thinking about this brings us back to an age-old question: why is the night sky dark to begin with? Many will say that the night side of Earth faces away from the Sun and therefore it is dark. However, that should not be the case in an infinite universe with infinitely many stars. As we saw earlier, Thomas Digges was first to envision an infinite universe with infinitely many stars and in essence, he was correct. However, in an infinite static universe with infinitely many stars, no matter which direction we look up in the night sky, we should always see a bright-lit star. Therefore, our night sky should not be dark at all. Why is that not the case? We do not know if Thomas Digges was already aware of this paradox but his fellow astronomers soon after him certainly were. In the eighteenth century, astronomers referred to it as the dark night sky paradox or Olbers paradox, named after the German astronomer Heinrich Wilhelm Matthias Olbers (1758 – 1840). It would take until modern times before science could resolve this paradox. As it turned out, the explanation for the dark night sky directly links to our discoveries of the *Big Bang* and the expanding universe. While the whole of the universe may be infinite, our observable universe is not. Since the *Big Bang* some 13,800 million years ago, the size of the observable universe has grown to the equivalent of a $\sim$46,500 million light-year radial extension; with about two-thirds of this growth due to the expansion of space itself. As the expanding space of our universe moves galaxies farther and farther apart, light from distant galaxies reaching us will become more and more shifted to the red of the spectrum and eventually completely move out of the visible spectrum. If our current model of the universe were correct, for the same reasons, long before our solar system formed, a virtual observer based in the position of Earth today would have witnessed a much brighter night sky. Playing this movie forward instead of backwards there will be a faraway future where the only bright objects in the night sky will be the stars of our own Milky Way Galaxy. There will not be a future human species on Earth to witness any of this. Long before the night sky will get darker our Sun will have died and Earth will be gone for many billions of years. However, in all of that we must not forget that when we look up into the night sky all the stars which we can see with the unaided eye are stars within our Milky Way Galaxy. The only objects outside of our galaxy that we can see in the night sky without optical aids are close-by galaxies such as the Andromeda Galaxy. Therefore, for observers finding themselves in the place where Earth's position will then be in our galaxy, the darkness descending on them will be a relative one as the stars in our galaxy will be alight a lot longer.

Before any of the above happens in the first place, our galaxy will actually change quite a bit. Even though everything in the night sky located outside of our galaxy is seemingly moving away from each other, within groups of galaxy clusters moving through space the gravitational force can be large enough to move galaxies closer to one another until they eventually collide and merge. This is just what is happening between the Andromeda Galaxy and our own Milky Way Galaxy. In 1913, the American astronomer Vesto Melvin Slipher (1875 – 1969) discovered that light from the then still Andromeda Nebula showed a shift towards blue wavelengths, meaning that with respect to our solar system the Andromeda Nebula was moving towards us. In Slipher's time, given the tools available, this was an extraordinary achievement, notably several years before Heber Curtis would firmly establish the Andromeda Nebula as a separate galaxy. Relative velocities of galaxies are difficult to measure exactly but astronomers expect that the Andromeda Galaxy and our Milky Way Galaxy will merge in about 4,500 million years. Some portray this future event as a collision but that is not a good analogy because there is so much space between stars in galaxies that collisions of celestial objects will be the rare exception. However, the merger of the two supermassive black holes into a new much larger supermassive black hole at the new galaxy center will be anything else but gentle. After the merger, the new galaxy should have many more stars adorning night skies as viewed from Earth. However, by then Earth will be lifeless. Any humans still admiring night skies then will do so from a different home in our galaxy.

In addition to the starlight coming from other galaxies dimming in the far distant future as they move farther and farther away from us, there is also another factor to consider that will eventually result in a dark universe. To us humans the resources of the universe seem unimaginably enormous and virtually inexhaustible but they are not. Our universe started out with a finite amount of hydrogen, the fusion of which continues to power stars. Hence, in the very far future, not in billions but trillions of years from now, our universe will run out of hydrogen to power new stars and eventually it will go completely dark. That may sound gloomy, but it is more of a philosophical than a practical relevance as a future in trillions of years is something way beyond what we humans can comprehend. The good thing is that all such events, the darkening of the night skies because of the expanding universe and our universe eventually running out of hydrogen, are so far out in the future that there is still a lot of time left for intelligent life such as us to discover all the wonders of this universe, much of which we do not understand yet.

We can count on the universe having many other surprises in store for us that will continue to confound science and likely in many such instances we can expect to find that our assumptions and thinking were not correct. Just consider the *Big Bang*. Calculating back towards this very moment at the beginning of our universe, scientists arrive at the fundamental limits of time and space. At these earliest points in time where the universe was infinitely small, the measures for time and space intervals become infinitely short. This earliest time in the history of our universe is referred to as the Planck Epoch, after the German physicist Max Planck (1854 – 1947), one of the founding fathers of quantum mechanics. The measures of the Planck Epoch are what physicists consider as the shortest possible intervals of time and space which they refer to as the Planck time and the Planck length. Physics as we know it today tells us that there can be no shorter time intervals than the Planck time or shorter distances than the Planck length. The Planck length is a measure derived from quantum mechanics considerations. It only includes three fundamental constants of nature: Newton's universal gravity constant $\boldsymbol{G}$, Planck's constant defining the smallest increment of electromagnetic action that can exist in nature, and the speed of light. Its value is $1.6 \cdot 10^{-36}$ meters which is 1.6 meters divided by one million million divided by one million million and once more divided by one million million. For comparison, the building blocks of atomic nuclei, protons and neutrons, have a diameter of roughly 10^{-15} meters. The diameter of our Milky Way Galaxy is on average about 170,000 light-years across.[64] The size of a proton or neutron as compared to the Planck length is similar to the diameter of our galaxy compared to an object of a meter in length. Planck time measures the time it takes light to travel the Planck length and it comes out as $5 \cdot 10^{-44}$ seconds, a five with forty-four zeros in front of it before we get to the decimal point. The Planck length and time are so infinitely small and short that our imagination cannot really make sense of them. Regardless of that, the problem lies not so much with these infinitely small measures but with the laws of physics that govern nature in the infinitesimal small, be that time or space.

With the very early universe, we are in the domain of the infinitesimally small, a world ruled by quantum mechanics. However, that very small measure of space and time that is the nascent universe, contained all the energy of the universe as we find it today. The latter is the domain of Einstein's general relativity. Physicists have not succeeded yet in reconciling these two theories. How can this quantum universe, so infinitely small but containing all the energy that this universe will ever have, evolve to become our universe ruled at large by Einstein's general relativity. Just think of this for a moment: when we talk about space and time in the very early universe, what meaning do these words have? Energy and mass are equivalent in Einstein's theory. When we refer to the bending of light, we think of it being bent by mass because mass distorts space-time and as light propagates in this space-time it traces these distortions; similar for time dilation due to gravity, where again the distortion of space-time produces the observed effects. Looked at it differently, distortions of space-time cost energy, bending space time requires energy and bent space-time in essence stores energy. Now, if we remove the mass that causes the space-time distortion than the distortion vanishes and space-time reverts to its lower energy state, which is flat space-time. Could there be a situation where space-time distorts so much that is will not revert to flat space-time by itself?

When we look at the early universe, what does it mean to talk about space and time given the unimaginably enormous energy densities at the first moments of the universe? Does the concept of a flat-space time even make sense there? What if space-time was not always flat and what if not all the expanding universe evolved into a more or less flat space-time, locally distorted by the distribution of mass which we find in the universe today? What if there are some creases left in space-time from an early universe, in which space-time was anything but flat? Creases in the space-time fabric that so to speak never flattened and could store enormous amounts of energy. Such extremely distorted areas of space could have the same effect on the flat space-time areas adjacent to them that enormous masses would have. Flat space-time has not only been a cherished notion for physicists but has been fundamental to us in how we understood the world around us for most of human history. We may not always have called it this way, but the understanding and the implied assumptions of a flat space-time were always there. Only a century ago, with Einstein's revolutionary insights, did physics finally have to concede that space-time is not always flat. Maybe we need to go further. Who knows, space-time may be even more complex than we think and just maybe some of what we observe in the universe and cannot understand today is due to space-time being much more weird than we ever thought.

There are limits to what we can know, and modern physics has made some of those very clear. While we know so much more than past generations and can hope this progress to continue, we also know now that there is a point we cannot go beyond. We have no way today to go beyond the point in time when the *Big Bang* happened. It has no meaning to talk about space and time before the *Big Bang*. Our time and our space start with the *Big Bang*, what we refer to as time or space did not exist before it. It is astounding that this revelation by modern physics of a beginning of time and space beyond which we cannot go, has been absorbed so calmly. Maybe we just have become so used to such discoveries or this is just too unimaginable for us to grapple with it. There was a time when philosophy and religion would have explored such new discoveries and would have dissected them in any possible way to understand the implications they held. Have philosophy and religion surrendered their roles to help us make sense of what frames our human condition which they held for thousands of years throughout human history; or do they today just focus on what is closer to home? There once was havoc, religious and intellectual turmoil, when natural philosophers removed Earth from the very center of our human perception of the cosmos. Now we understand ourselves as living on the outer edge of an average galaxy, one of trillions of galaxies in a universe so large we cannot even fathom it anymore; and physics tells us that all of that came out of nothing in what scientists call the *Big Bang*. What does that mean for humans, are there no consequences for what we believed to be true, no ethical or moral implications? Humanity has been thoroughly dethroned, but many seem not to realize it. We are not at the center of the universe anymore but just the inhabitants of a small planet in a rather unremarkable region of a galaxy that is one among trillions. With regard to how humanity thinks about itself for what we think about us, about our human condition, this vastly changed universe seems not to matter at all; except in science fiction and only rarely can one find deep thoughts there.

Quantum mechanics has shown us that space is never empty; there is never anything like "nothingness". At the quantum level, empty space is teeming with particles created in an instant and then disappearing again. This may sound strange but it all follows from what physicists have named the Heisenberg *Uncertainty Principle*, discovered by the German physicist Werner Heisenberg (1901 – 1976) in the 1920's. According to it, we can never know exactly where a particle is located in space and what precise speed it is traveling at the very moment at which we measure its exact coordinates.[101] What physicists found for the location and speed of a particle, applies also to other such pairs of measurements, energy and time being one of them. But this is not just about measurement as its implication is that particles can be generated out of nothing by borrowing energy from the vacuum of space without violating energy conservation as long as the energy is given back quickly enough. In this way, space is teeming with what are so-called quantum fluctuations, representing matter-antimatter particle pairs emerging from nothing only to quickly recombine into nothing again.

The electron and its antimatter particle the positron comprise one such well known matter-antimatter pair. The Austrian physicist Erwin Schrödinger (1887 – 1961) formulated the fundamental equation of quantum mechanics in 1925.[102] A few years later in 1928, the British physicist Paul Dirac (1902 – 1984) combined special relativity - not general relativity - with quantum mechanics. The resulting new equation had as one of its solution an anti-electron – the positron, which was experimentally confirmed in 1932.[89] Since then, the antimatter equivalents of many other elementary particles have been discovered. The laws of quantum mechanics governed the very early universe and physicists believe that the quantum fluctuations of this early universe imprinted on the expanding universe. In essence, what they are saying is that the large-scale structures of our universe such as the distribution of galaxies and clusters of galaxies throughout space, traces back to those quantum fluctuations.

If neither time nor space did exist before the *Big Bang*, in which realm did the *Big Bang* actually happen if there was no space and time before it. We have to realize that asking such question may not make sense because we are kind of bootstrapping ourselves here, like the fictional figure of the German Baron Münchhausen pulling himself out of a quagmire by his own hair. However, scientists are nothing less than curious and this has led them to formulate theories that seek to explain why we live in a universe with three spatial coordinates. You may have heard some of the names of such models, like of different versions of so-called string or superstring theories with something called M-theory seeking to unite several versions of self-consistent string theories. What it boils down to is that physicists speculate that we may live in a ten-dimensional universe of which there are only three spatial dimensions accessible to us. They give two possible reasons. First, it could be that these additional dimensions are so extra-ordinarily small, such that they are not part of a reality we could experience. Second, we could be living in a so-called three-dimensional manifold to which our reality is restricted. In a sense, the second possibility is similar to the situation the English theologian and schoolmaster Edwin A. Abbott (1838 – 1926) sketched in his 1884 published *Flatland*, a satire on Victorian life playing out in two dimensions. Make no mistake, life in the theoretical multidimensional worlds which physicists conceive is much more complicated than anything that takes place in *Flatland*, but certainly equally interesting. That we must mention multiple dimensions here at all is because at one point in time the universe must have been infinitely small and hence all space dimensions must have been of equal size, so-to-say. Being able to find answers as to why some of them remained very compact and why we may live in a three-dimensional sub-plane of a much larger-dimensional universe could help us connect the very beginning of our universe to the physical reality we see today.

Scientists are not just searching for explanations why we live in a three-dimension spatial world today, but also seek to understand what is behind the different relative strength of the four fundamental forces in nature, which in decreasing strength are: the strong nuclear force, the electromagnetic, the weak nuclear force, and the gravitational force. The gravitational force is the weakest but as we have seen in the discussion on star life cycles, it eventually can overwhelm all other forces. All matter is subject to this force and its effect is proportional to the respective masses involved and is inverse proportional to the square of their distance. The weak nuclear force, sometimes just called the weak force, is of a much shorter range, with a reach of only 0.1% of the size of a proton, it is the second weakest force and is responsible for radioactive decay. The electromagnetic force scales in the same way with distance as the gravitational force, however, it is not proportional to the masses of matter involved but to the respective electrical charges they carry. Different from gravity, it can be attractive as well as repulsive depending on the respective bodies carrying equal or opposite electric charges. Regardless of how large a body is, as long as it contains equal numbers of negative and positive charges, it will itself be electrically neutral and therefore in its movements will not be subject to the electromagnetic force itself. The strongest force in nature is the strong nuclear force, sometimes just called the nuclear force. It is 137 times stronger than the electromagnetic force and it is what holds the nucleus of an atom together. It is always attractive and acts only on a short range, roughly the diameter of a proton or neutron. Just like our three dimensions unfolded from a compacted higher dimensional space at the very beginning of our universe, so did the four fundamental forces of nature we know today in a way crystallize out of an original force that at a much higher energy density had them all united in one. Similar to the hidden additional dimensions invisible to us, maybe there were more forces crystallizing from the one original force, but they remained hidden from us as well. Such forces could for example govern the physics of dark matter and dark energy similar to how for example gravity governs the world we can see, or electromagnetic forces govern the physics of charged particles enabling our modern industrial societies that light up our planet at night.

Why, you may ask, is it worthwhile to mention any of this here at all? In our day-to-day lives, theoretical physics is seemingly something very remote. However, together with other sciences it has shaped much of our modern societies over the past century. It has given us the tools through which we know our world and the cosmos today. The above should just give you a flavor of how science grapples with the fundamental questions concerning the origin of our universe. Some may feel that quite a bit of what physics today speculates about multiple dimensions or forces we cannot detect, the latter seemingly governing much of a universe that we cannot see, is akin to the sages of antiquity building models of the universe without understanding the fundamental physics underlying the movements of planets and stars. This is very understandable. While sciences today subject themselves to a very different rigor than they did in the times of Aristotle or Ptolemy, one certainly has to wonder if not to some extent modern cosmology is in danger to develop its equivalents of epicycles or equants. Science faces a deep philosophical problem here. How to judge the validity of our physical models of something that we increasingly cannot test anymore? How to deal with the fact that our best physics theories have no explanations for dark matter or dark energy? How deeply do we really understand the universe we live in? What are we missing?

Today, we still have so many fundamental questions about our universe for which we cannot find any answers. There are clearly things we do not know yet but for some of them we may not even be in a position to formulate the right questions so we can meaningfully explore them at all. Maybe to some of our questions we may never find an answer unless as a species we have sufficient time to learn and grow. As a learning species, we have been taking our first baby steps, moving from myths and beliefs towards knowledge. We cannot expect to know in only a few generations everything there is to know about an enormous universe that has been thousands of millions of years in the making. We must make sure we continue on our path as a learning species and look for answers to those ultimate questions for as long as we can. There is no guarantee that we will succeed in this quest for the ultimate knowledge about our universe but try we must. Keep this in mind the next time someone tells you that we know everything there is to know. Famously, a physics professor purportedly told the young Max Planck in the 1870's when he inquired about studying physics that he should turn to another field as everything there was to know about physics was pretty much known. No such thing as complete knowledge of cosmology and astrophysics exists today, far from it. Yes, we have learned so much about our universe, but most of it remains a mystery to us providing more than ample opportunity for bright and energetic students to make their mark in advancing our knowledge of the cosmos we live in.

Our Home - Planet Earth

Colliding Worlds

Today it sure looks to many as if the human species owns planet Earth. However, that is an illusion. The justifications for this mistaken perception we do not know, but ignorance and hubris may be a good starting point. Anybody willing to take a closer look at Earth history will quickly discover that humans are a very recent addition to this planet's biological diversity. All life on Earth including humanity is the result of roughly 4,500 million years of natural geological, mineralogical, chemical, and biological experimentation in processes we today refer to in a broad sense as evolutionary. Over those seemingly endless eons, the reshuffling on the decks of our mothership Earth hurtling through space and time has been nothing but extraordinary. Sometimes these changes were severe enough to spell doom for many a species while opening up opportunities for others to gain prominence. There was never a preordained outcome and there is no pinnacle of creation as some would like to have it. If humanity will be more than a footnote in the records of Earth remains questionable. Planetary evolution and the evolution of life shape worlds on geological time scales, not human ones. Seemingly, everything on this rocky planet Earth and as some like to believe even the planet itself is at our command and mercy. We could not be more wrong. The human species could not survive for long without the delicately balanced life support put in place over the last roughly four thousand million years. Many of Earth's ecological systems are quite robust, and it takes gigantic forces to push them out of balance. In the past, such momentous changes affecting all life on Earth have been set in motion by natural events, including among others massive volcanic activity over long time spans, asteroids impacts, or the reorientation of Earth's axis. However, life has also been the cause for global changes on planet Earth, not through the brute forces of cataclysmic events such as the ones listed above, but through sheer persistence over hundreds and sometimes thousands of millions of years. Life has shaped Earth as much as geology and the geological evolution of our planet has in turn shaped life. A notable example is the production of an oxygen atmosphere that can support the aerobic respiration needs of large animals such as us. This has allowed us to evolve and at the same time changed Earth's surface. The oxygen-rich

Earth atmosphere sustains all higher life including ourselves and protects it from harmful ultraviolet radiation; at least as long as our atmosphere retains an intact ozone layer. Conditioning our planet to be suitable for lifeforms such as us played out on a time scale that counts in 1,000 million years. Then, towards the end of the last century, humankind almost managed to destroy substantial parts of the ozone layer in only a few decades.

Never before the arrival of the human species have any higher lifeforms been in a position where they could eradicate life as we know it today. Many feared throughout the cold war that this could happen even by accident. Fortunately, the cold war never became a hot one and the nuclear annihilation threat, while still present, never materialized. More recently, a new danger to life on Earth has manifested itself and that is the human induced change of Earth's climate. This time it is not any nefarious weapons of mass destruction that endanger life on Earth, but it is the way we live that puts our own species future at peril as well as that of all other life on Earth. Perversely, human greed threatens to destroy Earth's biodiversity through economic activities focused on cashing in short-term profits while offloading the long-term costs and consequences on future generations. The way we live today puts human greed ahead of life itself. Yes, we really are unique as in the end we may well succeed in destroying our own life support systems on this planet. Which other species can make such claims? Surely, if we really were to go down this path, we would take a large number of other species along for the ride into extinction. However, if that happens, most likely it will not be the end of life on Earth. It will just be different life starting over and after many hundreds of millions of years maybe there could be another sentient species evolving on Earth; or not.

It would seem that every human being should at least be somewhat familiar with our planet's history and how we, the human species, have come to play our seemingly dominant role on Earth. Regrettably, we teach our children little about the relevant facts. Sometimes we intentionally withhold the facts from them or even worse, teach them superstition and ask them to accept belief as a surrogate for knowledge. In today's age we all too easily allow our attention to be captured by fidgets of our imagination that increasingly create virtual worlds in which fiction trumps facts. In those realities it does not matter if and how much we understand anything. Instead, what seems to matter are the pleasures derivable from endless instant gratifications. No wonder that we often have little energy left to ask critical questions and invest the efforts and mental facilities required for learning and critical questioning. These fictional edifices, some call them virtual worlds, are raised on top of the remnants of political or religious beliefs that we are often still all too willing to take for absolutes. Despite knowing at heart that they are just relative truth at best. This believe and make-believe world we live in likes to have us focused on our here and now; both of which seem to be moving from the real to the imaginary. Through technology, we may become better at socializing across vast distances, which certainly should be welcome. However, this convenience may come at the expense of a reduced positive knowledge and diminished capabilities for sound reasoning when we compare ourselves to a few generations back when traveling to the next village was as far as most people got. At least back then people had to have a firm grip on reality outside of church matters because their very livelihood and survival depended on it. Not so anymore. Knowing how to handle one of those electronic gadgets is certainly a skill that we master today and that our grandparents could not even have dreamed of. Those marvels of technology very much appeal to our social instincts but networking our days away is in many cases likely an enormous waste of intellectual capital. Eventually, we may be isolating ourselves in big social bubbles where everything including our physical environments becomes virtual. There, we will be able to manipulate our worlds as we wish because laws of physics do not apply there, just as they never seem to do for superheroes or those who can imagine themselves as such. However, the physical world around us will still persist, governed by the laws of nature, while too many of us may waste away uploaded into our virtual worlds where we can all be superhuman and live our dream lives; until we cannot. It is only at the peril of our own extinction that much of humanity remains ignorant of where we come from and how life has developed on Earth. Every human being should know the broad outlines of how our planet came to be the blue marble that eventually would be able to host higher life on Earth such as us.

Climate change is an unfortunate reality, and the course is set for major changes to come. Maybe we can avert even more serious climate change than what we already have set in motion; the consequences of which we and future generations will have to face for sure. But this would require all human societies to stop immediately making things worse and intently focus our energies on developing mitigation strategies to lessen the impact of what has become unavoidable. Not everyone has understood this yet and while many politicians nowadays acknowledge climate change as a reality and understand the need to act, they do so only very slowly. The changes we must make will be very painful and there is nothing that politicians seeking reelection dislike more than to push for something potentially alienating their constituencies. Hope lies here with our younger generations including many not even yet of voting age. They increasingly realize that the older generations literally consume their future today without them having a say in it. That will not hold, and we should cheer for our young generations as they push for seminal changes in how we conduct our human affairs. Our societies must become much less motivated by greed and more by the responsibility we bear for the future of life on Earth, including our own future generations.

In the early 1970's we still had an almost balanced budget, meaning that we consumed all of Earth's annually renewable resources in about a year. Today, our industrialized world takes less than half a year to consume Earth's annually renewable resources with some countries consuming their proportional allotments in less than a few months. For millions of lifeforms including us, our planet Earth is a precious and finite resource. Where life comes from is still a mystery to us and maybe we will never fully understand it. We will look into this more deeply in the next chapter. Before we can go there, we need to look at Earth itself. There is of course highly specialized knowledge about Earth history and geology available, but it usually comes in a language only few outside the scientific community can appreciate. While much of that is far beyond the scope of this book anyway, the underlying story is something I believe everyone can understand. Therefore, in what follows the focus will be on what should be more or less common knowledge about our planet Earth, the kind of knowledge we could expect our youngsters to have when they graduate from high school. In that vein, what do we know about Earth and how did we learn about it?

Geology in History

Geology is a young science, relatively speaking. But geological thought in a broader sense is something that has been with us for much longer. We have to go back to the ancient Greeks to whom we owe the term geography. The ancient Greeks were excellent geographers and one of the last ones in the long line of Greek geographers of antiquity was Ptolemy. Ptolemy, the very same person we encountered earlier as the author of the *Almagest*, was also the author of a book titled *Geography*, dated to roughly 150 CE. Interestingly, the ancient Greeks did not coin the word geology. The modern use of the word geology originates in the eighteenth century but its first occurrences date to Medieval Latin sources. There the word *geologica* refers to the "study of the earthly things" and derives from Ancient Greek, combining the word for Earth, *gē*, and the ending *logía* for study or discourse.

Just because the ancient Greeks did not have a word for the study of Earth that does not mean that they did not entertain geological thoughts. Indeed, we find the ancient Greeks to be the first wondering about whether Earth's surface always used to look the way it did in their own time. We already met Thales of Milet as the first Greek astronomer in the previous chapter. Thales seems also to have been the first one we know to believe that Earth's surfaces once were completely covered by water. A conclusion he likely arrived at by discovering marine fossils on dry land, well away from the shores of the Mediterranean. That may have been fortuitous from his perspective as Thales based his view of the world on his belief that everything in nature originates from a single material substance, which he thought was water.

Anaximander (c. 610 – 546 BCE) was also a pre-Socratic philosopher from the city of Milet and a student of Thales. Most of his original work is lost but enough of his thoughts survive in what other Greek philosophers discussed and wrote about his work. Anaximander believed Earth to be a flat disk, the circular face of a short cylinder made out of a substance he called *apeiron*, an Ancient Greek word that identifies something unlimited or boundless. Here he differed from his teacher Thales who thought the original substance to be water. According to Anaximander the world had arisen from this cylinder filled with *apeiron* substance and Earth formed out of this *apeiron* covered by water with plants and animals arising from mud. Interestingly, he also believed humans came later and arose from fish. We do not know if Anaximander thought that Earth was initially covered by water because he also found marine fossils on dry land. We can assume that he encountered them as well and he must have been familiar with how Thales thought about this matter.

Xenophanes of Colophon (c. 570 – 475 BCE) was another pre-Socratic philosopher from Ionia.[1] He likely was well aware of what his earlier Ionian compatriots Anaximander and Thales had come to understand. Through observation of fossil fishes and shells, he too concluded that the lands where they were found must have been submerged under water at some time. Xenophanes is credited with having developed a theory of alternating periods of worldwide floods and droughts, inspired to some extent by the observation of marine fossils within Ionian coastal lands.

The ancient Greek historian Herodotus, whom we already met earlier, also encountered fossilized shells on land. He traveled widely, practically writing the most famous travel and culture guide of antiquity. Maybe he did not always get his facts right but his life's work, *The Histories*, is one of the best sources we have to understand how the ancient Greeks saw the world around them; and he is a good, though not always fully reliable source on the customs and cultures of other ancient people. The fossil shells he found in Egypt were certainly evidence to him that sometime in its distant past, Egypt must have been covered with water. The ancient Greeks took their geological thoughts no further than thinking about creation itself. They did not apply their methodical thinking in the same way to the understanding of Earth as they did in other areas of science. This is to some extent surprising as mining was one of the most important economic activities in antiquity. One would think that the Greeks might have been learning something about Earth along the lines of geology from their mining activities; but they did not. Maybe that was so because almost all of the backbreaking mining work was done by slaves and not by free Greek citizens.[2]

The Romans did not entertain more geological thoughts than the ancient Greeks did. The Roman we know as Pliny the Elder, born as Gaius Plinius Secundus (23 – 79), left us his encyclopedic thirty-seven-volume work *Naturalis Historia* (*Natural History*), aiming to describe the whole of the natural world. He drew heavily on Greek sources and to a much lesser extent on Roman scholarship to which Pliny added his own observations. From a geological perspective, there was not a lot to learn as the discussion of metals in various parts of the work and of mineralogy in the last two books remained descriptive only. However, this descriptive information made it the reference work with respect to metals and mining for almost 1,500 years.

Similar to what we have seen in the previous chapter for astronomy, Muslim scholars also served as conduits for Greek thoughts and learning in geography and geology. Just as with astronomy, Muslim scholar did not just pass knowledge from ancient Greece to early medieval scholars in Western Europe, they also added their own significant contributions. Not surprisingly, among those Muslim scholars we find some of the same names we already encountered earlier. The compartmentalization of science into specific disciplines is a modern phenomenon driven by the sheer amount of accumulated knowledge and the required specialized training to succeed as a scientist in a given field. However, for classical Greek as well as for Muslim scholars back then an astronomer who by necessity was already a mathematician and physicist would just as likely be interested in geography and geology, as we would call it today. The latter two often for practical reasons such as cartography and land partitioning, which could include assessing land values for its potential use such as for agriculture or mining.

In one of his two most famous works, *The Book of Healing*, Ibn Sīnā, known as Avicenna in the West, thought about the nature and causes of mountain building including the very long times such processes may take if they were not the result of enormous earthquakes. He also considered the potential effect of erosion on mountain ranges over very long stretches of time. He was not the only one thinking in terms of what may sound to us like modern geological concepts. His contemporary, Abu Arrayhan Muhammad ibn Ahmad al-Biruni (973 – 1048), frequently just referred to as al-Biruni, was like Avicenna a polymath, geographer, and interested in geology. His works on Indian culture, history, and geography remained influential sources well into medieval times. Al-Biruni also authored a manual of minerals and gems.

In the West, we know much less about how further in the East Chinese scholars occupied themselves with what we would call today geological thought. In a general sense, this is true also for other areas of sciences and has its roots in the fact that natural sciences were not included as part of the cannon of the imperial examinations that started with the Sui Dynasty in the late sixth century.[3] Geological thoughts of scholars such as Du Yu (222 – 285) or Ge Hong (284 – 364) were just not part of that. Men like Shen Kuo (1031 – 1095), no less a scholar than any of his early medieval Muslim colleagues, were exceptions to the system. A century before the Englishman Alexander Neckham (1157 – 1217) in the West, Shen described a magnetic needle compass in his *Dream Pool Essays* in 1088 and was first to understand the concept of true north. Shen also inferred from his marine fossils finds in China's interior that oceans must have covered such lands at some time in the past and he could see how sedimentary layers had been uplifted in the landscape.[4] Together with his observations of soil erosion and silt deposition, this led him to formulate a theory of land formation. Other Chinese scholars such as Zhu Xi (1130 – 1200) followed Shen's example but would not surpass him.

Only in the sixteenth century did Europeans begin to make the kind of geological observations at home which Shen Kuo had recorded much earlier in China. Georgius Agricola (1490 – 1555), a German mineralogist and metallurgist, considered the father of mineralogy and one of the founders of geology, spent much of his life in an area whose economy was dominated by mining, the Ore Mountain Range in central Europe.[5] His most famous work, *De re metallica* (*On the Nature of Metals*) published posthumously in 1556, summarizes the state of the art of mining, refining, and smelting metals in his time. Many consider Nicolas Steno (1638 – 1686), the Danish scholar, scientist, and eventually priest, as one of the founders of stratigraphy and modern geology.[6] When he dissected the head of a shark, he realized that shark's teeth looked very similar to the stony objects called *glossopetrae* or "tongue stones". Frequently embedded within rock formations, they are indeed fossilized shark teeth.[7] *Glossopetrae* had already been convincingly identified as such by the Italian naturalist Fabio Colonna (1567 – 1640) in his treatise *De glossopetris dissertatio* (*A Dissertation On Tongue Stones*) published in 1616. However, Nicolas Steno went further as he not only thought about the fossils but also about how they may have become embedded in rock layers and how these layers were deposited in the first place. He came to realize that through examining rock layers, or different strata as geologists refer to them, one might be able to infer which ones were deposited first and which ones came later. Nicolas Steno summarized his geologic studies in his *Dissertationis prodromus* (*Preliminary Dissertation*), published in 1669.

Edmond Halley was not only a distinguished astronomer, but he also sailed the seas to map isogonic lines which connect points on Earth's surface where the magnetic declination is the same.[8] Halley published these measurements in *General Chart of the Variation of the Compass* in 1701. While mapping isogonic lines he discovered magnetic anomalies through the seemingly erroneous compass readings they produced. To explain these, Halley proposed in 1692 a concentric hollow Earth shell model consisting of three shells and an inner solid core, the hollow shells themselves he seemingly envisioned as separated by atmospheres. When Halley died in 1740, his model was already in doubt and in 1778 the Scottish geologist and naturalist James Hutton disproved it for good.

James Hutton
1726 – 1797

James Hutton played also a key role in another controversy regarding Earth's geology and history, the dispute between *Neptunists* and *Plutonists*, which started in the middle of the eighteenth century.[9] *Neptunism* advocated the viewpoint that rocks formed by crystallization of minerals in the early Earth oceans. The German geologist Abraham Gottlob Werner (1749 – 1817) was first to propose this hypothesis and this is why *Neptunists* sometimes referred to themselves as *Wernerian's*. Opposed to it was the hypothesis of *Plutonism*, originally proposed by the Italian abbot Anton Lazzaro Moro (1687 – 1764), which advocated that those rocks formed in fire. Because of that, natural scientists supporting this point of view were sometimes also referred to as *Vulcanists*.[10] *Neptunists* maintained that basaltic rock was a sedimentary deposit as it seemingly often included fossils and so could not be of volcanic origin. This is where James Hutton entered the debate. He opposed the *Neptunist* view that all rocks were precipitated during a single enormous flood. His argument was that granite or basalt rock never contains fossils but penetrates sedimentary or metamorphic rock, which carries the fossil inclusions. Hutton understood that sedimentary and metamorphic rock had very different origins from granite and basalt rock, which he believed formed from molten rock deep inside Earth. As erosion broke up sedimentary rock it was replaced by new granite and basalt rock reaching Earth's surface. He found examples of basalt cutting through metamorphic schist rock that proved his point. In many ways, this came very close to the modern view. Consistent with this *Plutonist* view, Hutton proposed that the interior of Earth is hot, hot enough to melt rock. For good reasons James Hutton is referred to as the father of modern geology. He was the first to propose a fundamental principle of geology, the theory that came to be known as *Uniformitarianism*, or the *Doctrine of Uniformity*. Hutton put forward his views on geology in his *Theory of the Earth*, published in 1788 in the *Transactions of the Royal Society of Edinburgh*. However, the ideas summarized in this work had been circulating for several years before that.

The theory of *Uniformitarianism* essentially asserts that the present is the key to the past. The way we see Earth's surfaces being shaped by natural forces today is no different from how it happened in the past. Because the natural processes we see shaping Earth today operate on a geological time scale, the very nature of the theory of *Uniformitarianism* would quickly lead geologists to realize that Earth must be much older than what had been assumed until then. It was through the work of another remarkable Scottish geologist, Charles Lyell, that *Uniformitarianism* found broad acceptance. Charles Lyell's *Principles of Geology*, published in three volumes from 1830 to 1833 was not only written for geologists. Lyell clearly believed that the Earth was very much older than anyone had imagined until then, but he did not try to estimate how old it really was. However, he was the first to use the term evolution in a context that Charles Darwin would come to appreciate. In fact, Lyell became an early mentor and lifelong friend to Charles Darwin.

Charles Lyell
1797 – 1875

Darwin carried the first volume of *Principles of Geology* with him when he left on HMS Beagle from Plymouth in December 1831; the second volume reached him a little less than one year later in Montevideo.[11] *Uniformitarianism*, the theory of geology relying on the accumulation of natural changes over geological time spans, certainly had a profound influence on Darwin's thinking. The infinitesimal changes through which he believed the evolutionary processes worked required such a geological time scale. When Darwin would formulate his theory of biological evolution through natural selection, Lyell would come to support evolution but not wholesale, as he would question the mechanism of natural selection. At the time Lyell's *Principles of Geology* was published, *Catastrophism* was still the dominant geological theory and had a substantial following. *Catastrophism* advocated the viewpoint that for the most, Earth had been shaped in the past by sudden violent and short-lived events that could be worldwide in scope. What may have made *Catastrophism* still so attractive to many geologists in Lyell's time is that it provided room for religious accounts to be part of Earth history. For example, Noah's Flood could have been a real geological event. While today this may make many of us wonder, we should remember that even in our day and age there are still a few among us who interpret religious books quite literally. Seemingly, such people are not troubled at all that in order to do so they must discard reality to fit scripture, regardless of however contortionist this effort may become over time as science progresses. In the early nineteenth century, this worldview was still much easier to uphold than it is today. Unlike *Catastrophism*, *Uniformitarianism* could not provide the consolation of Earth's geological history obliging to biblical accounts.

However, there was also more serious opposition stemming from a perceived lack of evidence, seeing *Uniformitarianism* as a mere theory only, for which Lyell was pulling together evidence as it fitted him. Maybe that had to do with the very practical bend of geology at the time. Geology was no supposed to theorize but to explain the facts as they were found in nature. In the eyes of its most respected practitioners, geology was a strictly empirical science. Eminent geologists of the time held this view including for example the British Adam Sedgwick (1785 – 1873), one of the foremost geologists of his time and like many of his Cambridge or Oxford professorial colleagues an ordained Anglican minister. *Catastrophism* supposedly had all the evidence given that fossil bones of seemingly extinct animals, an increasing number of which were dug up in the nineteenth century, clearly represented the remnants of lifeforms perished in past biblical floods. For quite some time, the lack of a proper understanding of the fossil record made *Catastrophism* look like the explanation that was in better alignment with the geological and fossil record. But this would change once Charles Darwin had entered the fray. As natural explanations for the fossil record became more accepted most people took reality for what it was and eventually *Uniformitarianism* prevailed to become and remain the foundation of today's geological thought.

An Introduction to Earth

What we know or hypothesize about Earth's interiors comes from mainly four sources: the study of earthquakes, laboratory experiments, theoretical modeling, and the geological record. The oldest experimental method and still the only one that can provide us data generated from probing the deep inner Earth directly is seismology.[12] Seismology as a scientific discipline to explore Earth's interior is a little more than one hundred years old. It mostly studies propagation of seismic waves within the body of the Earth and on the surface of the Earth. Earthquakes are major sources of seismic waves being used in this effort. By measuring seismic wave reflection, transmission, and travel time through various parts of Earth, scientists can make conclusion about the material composition of Earth's interior. Shortly after seismology got its start, laboratory measurements that seek to replicate pressures and temperatures as we believe them present in various parts of inner Earth began to contribute to our understanding of how materials may behave under such extreme conditions. These are technically very demanding experiments, designed to extract data from various materials subjected to extremely high pressures and temperatures. More recently, with powerful computers becoming available, theoretical models have become a major tool to study the complex dynamics of inner Earth. Building self-consistent models that replicate experimental observations of inner Earth properties has become an important research field. Given the limited access we have to study Earth's interior directly, such models allow scientists to validate their hypothesis of certain aspects of Earth's dynamics and to identify more potential ways to test their assumptions directly or indirectly. In addition to seismology, laboratory experiments, and theoretical modeling, there is of course also Earth history as it reveals itself through the geological record. The forces of geology and the natural world constantly rework Earth on a geological time scale. Because of that, a detailed mineral record of a good part of Earth's interior lies exposed on Earth's surface today. One must only be able to read it. Over roughly the past two centuries, aided by increasingly sophisticated technology, geologists, mineralogists, paleontologists, seismologists, and their colleagues from a number of other scientific disciplines have been able to do just that.[13]

A brief qualitative description of the planetary system we call Earth, as we understand it today, will help set us up for the narrative of Earth's evolution to follow. Frequently, Earth is compared to an egg where eggshell, egg white, and yolk correspond to Earth's crust, mantle, and core, respectively. It is a neat analogy and just like an egg, Earth is not a perfect sphere. Because of its rotation, Earth bulges at the equator and therefore the Earth radius as measured to the equator is ∼6,378 kilometers while it is ∼6,356 kilometers as measured to one of Earth's poles. The thickness of Earth's crust varies from three to fifty kilometers, with the crust below Earth's oceans being much thinner than the crust making up Earth's continents. The former, referred to as the oceanic crust, ranges from about three to ten kilometers while the latter, referred to as the continental crust, ranges from about thirty to fifty kilometers. Below Earth's crust lies the mantle, which is on average about 2,900 kilometers thick, and below the mantle lies the core with an average radial thickness of about 3,470 kilometers.Fig. 2.1 gives a graphical schematic of the Earth shell model, and tab. 2.1 summarizes the characteristic properties of the three main layers, their subdivisions, and thickness ranges.

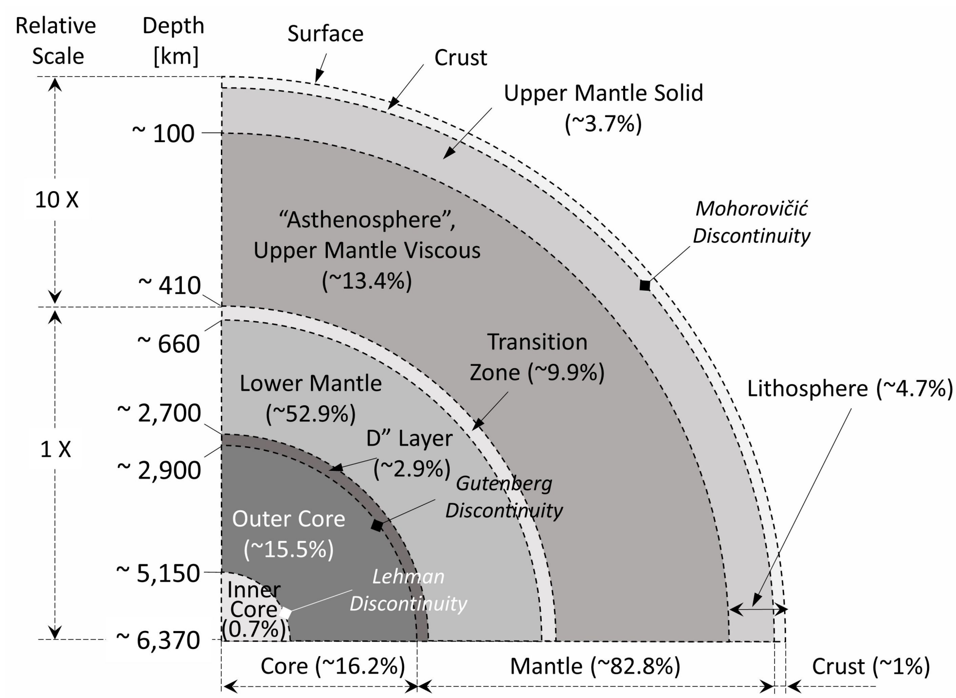


Figure 2.1: A simplified model of Earth's interior. The percentage numbers next to layers indicate the volume percentage a given layer makes up of Earth's total volume (not counting the atmosphere). Please note that for the layers starting from the upper mantle upwards the scale is enlarged by ten times relative to the layers below.

For the thicker layers in fig. 2.1, the thickness variation of individual layers is small relative to the thickness of the layers so showing them as shells is not too far a stretch. However, for some of the thinner layers the thickness variation can be a large fraction of their total thickness so showing them as shells of uniform thickness is not giving the right picture. Specifically the thickness of the D" layer can vary quite a bit as do the thicknesses of crust and lithosphere. The scale in fig. 2.1 has been blown up ten times from the upper mantle layers upwards as compared to the layers below because otherwise some of the thinner layers would be hard to discern on a uniform scale drawing.
Layer thicknesses are derived through the observation of seismic waves that are recorded all over the surface of the Earth as they are created by distant earthquakes. This technique takes advantage of the fact that seismic waves reflect at layer boundaries designated by density differences and refract when they enter layers of different density. The mantle is thought to be composed of four sublayers: the upper mantle, a transition zone, the lower mantle, and the boundary layer on top of the Earth core referred to as the D" layer. Earth's mantle makes up about 83% of our planets volume vs a combined roughly 16% for its core and ∼1% for its crust. The upper mantle layer is composed of a solid uppermost part on top of a viscous layer below. This viscous layer, where due to temperature and pressure rock becomes ductile, is the asthenosphere, the Greek name for a "weak sphere". The solid part of the upper mantle on top of the asthenosphere, together with the crust layer, is referred to as the lithosphere.

Table 2.1: Material composition, density, pressure, and temperature variations through Earth layer structure. Pressure ranges correspond to pressures estimates at the top/bottom of a layer while pressure and temperature estimates are given for layer boundaries. Between some layers there are significant density, pressure, or temperature discontinuities which have been named after the scientists first identifying and describing them.

Earth layers and layer subdivisions			Thickness [km] Range	Avg.	Materials & Chemical Compositions in Weight % Rocks	Minerals	SiO_2	Al_2O_3	FeO	MgO	CaO	Other	Density [g/cm³]	Pressure [GPa]	Temperature [°C]
Crust	Lithosphere	Continental Crust	~30 - 50	~3 - 50	Granite	Feldspar Quartz	~61	~17	~6	~4	~6	~6	~2.7		
Crust	Lithosphere	Oceanic Crust	~3 - 10		Basalt	Feldspar Pyroxene	~51	~15	~10	~8	~11	~5	~3.0		
	Lithosphere	MOHOROVIČIĆ DISCONTINUITY													~500
Mantle	Upper M.	Upper M., solid	~50 - 100	~2,900	Peridotite	Olivine Pyroxene	~44	~4	~8	~37	~3	~3	~3.3 ⋮		
													⋮		~1,300
Mantle	Upper M.	Asthenosphere (Upper M., viscous)	~310												
													~4.4	~14	~1,900
Mantle		Transition zone	~250			Wadsleyite / Ringwoodite		Majorite garnet					⋮		
													⋮	~24	
Mantle	Lower M.	Lower mantle	~2,040			Bridgmanite Ferropericlase Silicate Perovskite							⋮		
													~5.6	~127	
Mantle	Lower M.	D" layer	~200			Silicate post-Perovskite									
		GUTENBERG DISCONTINUITY												~136	~5,000
Core		Outer Core	~2,250	~3,470	Liquid iron and nickel								~9.9 – ~12.1		
		LEHMAN DISCONTINUITY												~392	~5,500
Core		Inner Core	~1,220		Mostly iron								~12.7 – ~13.0		
														~364	> 6,000

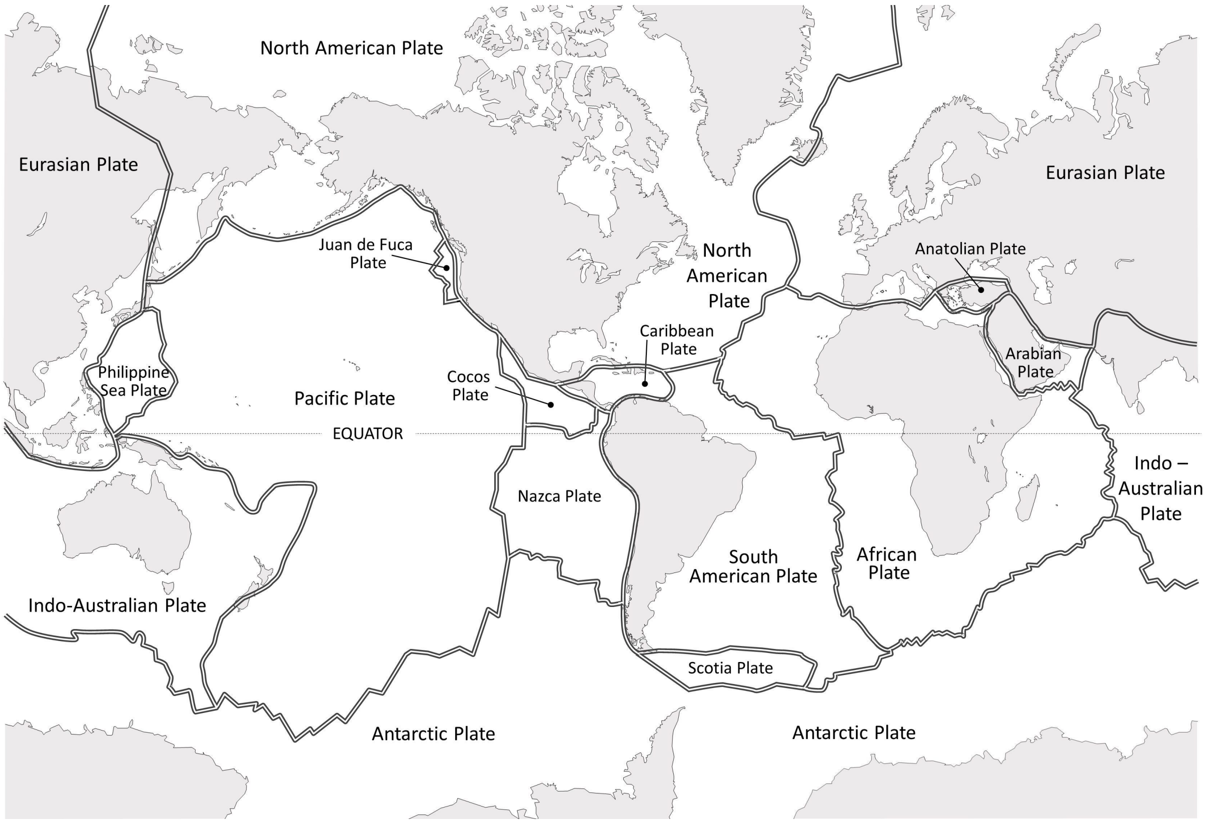


Figure 2.2: The seven large and eight small lithospheric plates of Earth's crust with the continents overlaid. Lithospheric plate boundaries are indicated by double lines.[15]

The lithosphere, which is Greek for "rocky sphere", constitutes the rocky outermost shell of Earth. Its radial thickness varies with that of the uppermost solid part of the mantle and with the thickness differences between oceanic and continental crust. The lithosphere layer is not a continuous shell but fractured into fifteen major so-called tectonic plates.[14] In alphabetic order, the seven larger plates are the African, Antarctic, Eurasian, Indo-Australian, North American, Pacific, and South American plates; whereas the eight smaller plates include the Anatolian, Arabian, Caribbean, Cocos, Juan de Fuca, Nazca, Philippine Sea, and Scotia plates.[15] Fig. 2.2 shows a map of these plates with the continents overlaid. As can be seen, the lithospheric plates do not usually coincide with boundaries of continents and oceans.

Modern geology has its roots in the nineteenth century. Before geology came geography and only a few hundred years ago, our knowledge of Earth's geography was still quite limited. When in the late fifteenth century, the age of European seafaring discovery began, the initial exploration and commerce with the Americas, Africa, and Asia provided the information needed for much improved maps. Commerce quickly turned into outright conquest and cutthroat competition with frequent armed conflict among the colonial powers in the making. This made accurate maps even more valuable, and the science and craft of cartography entered a golden age. Not surprisingly, cartographers started to realize how the coastlines of South America and Africa seemingly fit together.

Eduard Suess
1831 – 1914

Abraham Ortelius (1527 – 1598), the great Dutch cartographer credited with the creation of the first modern atlas, *Theatrum Orbis Terrarum* (*Theater of the World*) printed in 1570, certainly noticed it. It is not clear who first spoke of a jigsaw fit between the coasts of South America and Africa, but it was not the English scientist and philosopher Francis Bacon (1561 – 1626) as is often asserted. Francis Bacon's *Novum Organon Scientiarum* (*New Instrument of Science*) printed in 1620 is a milestone for modern science in many ways but it only refers to the coasts similarities and not to a jigsaw puzzle. While we cannot be exactly sure who first saw the coast lines of South America and Africa as jigsaw puzzle pieces, we can be sure that in the sixteenth century the similarities of the two coastlines were obvious to learned women and men alike. In the footsteps of the age of exploration followed the age of exploitation. During this historical phase of empire building by various, mostly western nations, it became evident that geologies on different continents were too similar for this to be an accident. Likewise, fossil plant and animal discoveries indicated a much greater similarity than one would expect to have flourished on continents separated by vast oceans. Certainly, the German naturalist and explorer Alexander von Humboldt (1769 – 1859) believed so. Not only did he travel extensively but he was also a prolific writer whose works were widely read and acclaimed; Charles Darwin, for one, admired him greatly. Humboldt was the first to explore and describe Latin America from a scientific point of view and proposed that South America and Africa were once joined together in a single landmass. To explain the accumulation of all those facts pointing towards a different arrangement of continental landmasses in the past, the Austrian geologist Eduard Suess proposed in 1861 the existence of a giant supercontinent in Earth's distant past. He called this unified continental mass Gondwana or Gondwana-Land and believed that sometime in Earth history it joined together the continents of South America, Africa, India, Australia, and Antarctica.[16] Suess was a famously productive geologist, summarizing much of the geological knowledge of his time in his four-volume treatise *Das Antlitz der Erde* (*The Face of the Earth*) published between 1883 and 1909. In it, he set out to describe and explain Earth's geologic structures and offered theories as to their evolution. In the process he coined many terms that we still use today even without knowing it. This notably included the first use of the term biosphere to refer to the domains of Earth supporting life. Today, Eduard Suess is mostly remembered for his proposal that the occurrence of glossopteris fern fossils documented in India, South Africa, South America and Australia are due to these continents being united in a past supercontinent, his Gondwana. He not only proposed and named this ancient supercontinent but he also coined the name for the Tethys Ocean, an ancient Mediterranean, separating the southern continent of Gondwana from the remaining continental landmasses to its north. These northern landmasses were united in their own supercontinent called Laurasia, comprising much of North America and Eurasia.

In his detailed studies of the Alps, Suess could see and understand that there had been lateral movement of continents. However, he was still too much a child of his time and remained beholden to the orthodoxy that continents do not move much laterally, at least not to the extent that they could merge into a supercontinent such as Gondwana-Land. So according to Suess, Gondwana came about by sea levels falling or lands rising thereby establishing land bridges that could foster the spread of animal and plant live to produce the fossil record we find today. However, this hypothesis could neither explain the neat jigsaw puzzle fit of the South American and West African coasts nor the extent of mountain building in the present or in the past as indicated by vast eroded mountain ranges; the remnants of which are still visible today, depending on one's preferred view, as enormous scars or stitches crossing whole continents. New discoveries would even be harder to explain in the static picture of nineteenth-century geology that for the most part only allowed vertical continental movements. It would require a paradigm shift for geology, similar to what physics had experienced during the age of Newton, to break the mold and find a new approach that could eventually evolve into the modern understanding of geology that we have today. This moment came in 1912 when Alfred Wegener, a German geophysicist and meteorologist, proposed continental drift being responsible for continents repeatedly joining together and breaking apart throughout Earth history. According to Wegener, this resulted in the formation of the supercontinent Pangea several hundred million years ago, uniting all continents on Earth into one giant landmass.[17] He first proposed this at a meeting of the local branch of the forerunner of the German Geological Society followed by the publication of *Die Entstehung der Kontinente* (*The Origin of Continents*) in the preeminent German geographical journal of the time, *Petermanns Geographische Mitteilungen.* In 1929, when Wegener was a professor at the University of Graz, the capital of the Austrian province of Styria where Kepler had taught a few hundred years earlier, he expounded his continental drift theory with his book *Die Entstehung der Kontinente und Ozeane* (*The Origins of Continents and Oceans*).

Alfred Wegener
1880 – 1930

In addition to explaining the geological and fossil similarities found on different continents and the remarkable jigsaw puzzle patterns of some continental boundaries, continental drift could also help explain a number of other phenomena where no good explanations were available at the time. It had for example been known for some time that very similar glacial deposits can be found on different continents, including South America, Africa, the Arabian Peninsula, India and Australia. It is hard to reconcile the existence of these glacial deposits located today on separate continents and at very different latitudes. Unless one would assume multiple glaciations at these very different geographies around the same time which happens to be some 250 million years ago, towards the end of the geological period called the Permian (we will come to geological periods in a few pages when we will look into the "The Concept of Geological Time").

However, such an assumption does not reconcile with other evidence for glaciations. Therefore, the only plausible explanation left was that all of those geographies were at one time in the past joined together at the same latitude, somewhere close to the South Pole. The South African geologist Alexander du Toit (1878 – 1948) made this very suggestion in his 1927 publication *A Geological Comparison of South America with South Africa*. It provided another strong support for Eduard Suess' hypothesis of a Gondwana-Land once uniting the continental masses of South America, Africa, India, Australia, and Antarctica as well as a few smaller pieces into a single landmass whose southern portion covered the South Pole about 250 million years ago and back then experienced a severe glaciation event.

Du Toit was one of the strongest supporters of Alfred Wegener's ideas and in 1937 he published his *Our Wandering Continents; An Hypothesis of Continental Drifting*, where he famously made his case for continental drift. In it, he proposed the existence of an ancient supercontinent he named Laurasia, which included most of the landmasses we find in the continents of the northern hemisphere today. Essentially, Laurasia formed the northern shores of the equatorial Tethys Ocean and was the counterpart of the southern supercontinent Gondwana that Eduard Suess had proposed. Laurasia combined the landmasses called Laurentia and Eurasia; Laurentia being the name given to the land that is at the core of the North American continent while Eurasia included substantial continental pieces that form large parts of today's Europe and Asia.

In the late 1930's, du Toit was quite literally a lonely voice among his Anglo-Saxon colleagues as many of his peers in the geological science establishment in the United Kingdom but specifically in the United States were not yet sharing his opinion on continental drift and remained adamantly opposed to it. That was even though Alfred Wegener's hypothesis also held the promise for a much better understanding of what drives major changes in Earth's climate system, including volcanism, mountain building, or how earthquakes come about. Continental drift had a major problem. There was no known mechanism to explain what exactly caused continental drift in the first place. Continents plowing through the bottoms of the sea was not something geologists were willing to accept, specifically if it came from an outsider to the profession; after all, Alfred Wegener was a geophysicist and meteorologist, not a geologist. True, in principle continental drift could explain how landmasses had moved in the past. However, it remained a mystery which underlying physical processes could make continents move.

After the Second World War, new technology developed during the conflict became available for civil use. This turned out to be a game changer for a number of scientific fields but in particular for marine exploration and there specifically for mapping the world's deep ocean sea floors. It quickly became clear that a mountain ridge, the mid-Atlantic ridge, divides the Atlantic Ocean with the sea floor falling from there towards Africa and Europe on one side and to the Americas on the other side. Similar topographic information discovered around the world quickly established the existence of connected sea floor ridges on all ocean floor basins. In addition, it became clear that along the center of these oceanic mountain ridges ran deep trenches and on both sides of these trenches magnetic stripe patterns were found revealing parallel to the trenches materials of higher and lower magnetic strength in an alternating sequence. At this point, geologists benefited from their paleontologist colleagues who had been accurately dating magnetic rocks for some time.

As magma reaches Earth's surface and solidifies, Earth's magnetic field aligns magnetic materials, for example magnetite contained in the lava, along its direction from the magnetic North Pole to the magnetic South Pole. Paleontologists had known this fact for some time and understood that this in principle allowed them to understand geological movements including the movement of continents across Earth's surface. Earth's magnetic field is essentially a dipole field, similar to the magnetic field of the bar-shaped magnet we have all seen demonstrated during our school days, with the magnetic field lines being parallel to Earth's surface at the equator and perpendicular to Earth's surface at the magnetic poles. Because of that, lava pouring out at latitudes close to the equator will have its magnetic moment oriented parallel to Earth's surface, whereas lava solidified at higher or lower latitudes will have its magnetic moment tilted towards Earth's surface. The latter the more so the closer this happens to the magnetic poles where it eventually would be oriented perpendicular to Earth's surface. For magnetic rock inclusions at high latitudes with their magnetization parallel to Earth's surface the conclusion then is that the rock formation where this magnetic layer is embedded was created at the equator and not at the higher latitude it was found. Provided of course, other geological events such as layer folding did not produce this reorientation.
However, paleontologists observed something else. Seemingly, the orientation of Earth's magnetic field powered by Earth's outer liquid core layer must have switched many times during Earth history by 180 degrees, essentially switching the positions of magnetic North Pole and magnetic South Pole. A magnetic orientation that is parallel to Earth's surface at the equator can be either parallel or anti-parallel to today's Earth magnetic field. Similar, as one moves towards the magnetic poles, the magnetic orientation of lava material pouring out will be tilted towards the surface with its magnetic direction either pointing into or out of Earth's surface. This direction depends on Earth's magnetic field being oriented as it is today or switched by 180 degrees when the lava containing the magnetic material hardened; and also, on whether this happens in the northern or southern hemisphere. Lava embedded between continental rock formations can be dated and therefore it provides a record of the orientation of Earth's magnetic field over time. The timetable paleontologists established this way shows Earth's magnetic field over many millions of years alternating between regular periods where Earth's magnetic field orientation is the same as today, and reversed periods with its poles rotated by 180 degrees. This was exactly the information required to understand the measurement results of alternating magnetic field strength perpendicular to the oceanic ridges. As it turned out, the alternating weaker and stronger magnetic signatures measured parallel to the sea floor ridges actually represented recordings of the orientation of Earth's magnetic field in the past. When lava from the trench at the center of an oceanic ridge pours down left and right to the ridge, magnetic lava materials align with Earth's magnetic field as the lava solidifies. Then, once Earth's magnetic field changes direction by 180 degrees, magnetic materials in new lava pouring out from the ridge align in the opposite direction. A measurement of the magnetic field strength in a perpendicular direction away from the ocean ridge across these alternating stripes will register a sequence of higher and lower field strength. This is because a higher magnetic field will be measured across stripes which have their magnetization parallel to today's Earth magnetic field as the fields add up; and a lower magnetic field will be measured across stripes whose magnetic

orientation is anti-parallel to today's Earth magnetic field as the two fields partially cancel each other out in the measurement. The result is a record of alternating higher and lower magnetic fields as one moves away in a perpendicular direction from the oceanic ridge. This is exactly what scientists discovered when they first made such measurements in the 1960's. The alternating magnetic stripes were clear and unambiguous evidence for sea floor spreading, and that in turn provided the much-needed understanding of how continents could move. By the 1960's, mid-ocean ridges and continental rifts, transform faults, fracture zones, subduction zones, active volcanoes and hot spots as well as submarine volcanic chains had been mapped. In the 1970's, the broader scientific community came to fully accept the new theory of plate tectonics that described the interaction of the moving lithosphere plates.

In its most simple description, plate tectonics is the process whereby heat convection processes in Earth's mantle drive movements of the lithosphere plates. The respective transfer of heat and material in the mantle essentially landscapes Earth's surface. It is the driving force behind the observed sea floor spreading at mid-ocean ridges, where magma rising from the asthenosphere creates new crust. Heat convection processes in Earth's mantle lead to destruction and in the long term to recirculation of Earth's oceanic crust through plate subduction, where one plate dives below another into the asthenosphere and deeper into Earth's mantle. Subduction zones are mostly created where two tectonic plates carrying oceanic crust meet with the leading edge of the denser of the two plates bending downward creating ocean trenches and chains of underwater volcanoes. Eventually, in geological time, this can lead to the formation of islands. Ocean trenches also form by subduction between a plate carrying continental crust and a plate carrying oceanic crust. Because continental crust is always less dense and therefore more buoyant than oceanic crust, oceanic crust usually subducts below the continental crust. This process creates the deep ocean trenches that we find right at the edges of continental shelves. Stretching from the coast where this happens towards inland, this results in mountain building as the continental crust buckles up. When two continental plates run into each other, the inexorable outcome is mountain building. The Himalayas are a prime example for that. For about fifty million years now, the Indo-Australian plate is colliding with the Eurasian plate creating the Himalayan mountain range, which continues to rise. Oceanic crust is continuously created at mid-ocean ridges leading to sea floor spreading and oceanic crust is continuously recycled back into the mantle in subduction zones. Continental crust gets seldom recycled this way. Rather, it weathers and erodes with the sediment washed into the sea where eventually it will be recycled as the oceanic conveyor belt takes it down with it into the subduction zone. Oceanic crust is therefore much younger and does not preserve Earth's geological record as well as continental crust. For example, dating of the oceanic crust that has been spreading from the mid-Atlantic ridge shows that the basaltic material found at the western coast of Africa and the eastern coast of North America is only about some 150 million years old. Compared with the oldest rocks found in continental crust dating back some 4,000 million years, oceanic crust is clearly much younger.

Mountain building, most volcanism, and earthquakes are thus the results of tectonic plates colliding. When we look at an Earth map showing the major mountain ranges, volcanoes, and earthquake zones we get a good idea where plates collide. Because the tectonic plates are moving on the surface of a sphere and not on a flat unbounded surface there are specific mathematical constraints that limit how the plates can move. This knowledge combined with the understanding of the forces behind the plate movements has resulted in a detailed picture of how the tectonic plates have moved in the past and how we can expect them to move in the future. In turn, this has enabled scientists to reconstruct the formation and separation of several supercontinents or more loose continent agglomerations, allowing them to predict when we can expect the next formation of such a supercontinent. It will take a while though until this next supercontinent referred to as Pangea Proxima, literally *the next Pangea*, will have taken shape in about 250 million years. The separation of the human and chimpanzee line happened only sometime between five and seven million years ago and modern humans have only been around for some 200,000 to 300,000 years. So, 250 million years is an evolutionary stretch for the human species. If humans should still be around to witness the slow-motion changes occurring over millions of years in this process, it will be interesting to see how they deal with the vanishing of their countries borders or seas. If we do not learn how to live together as a peaceful species and stop bickering about who owns which piece of planet Earth, the likelihood of our species still being around then will be close to zero. Hence, nothing to worry about vanishing or merging countries, it will not be a problem either way.

On our individual human time scales, we can only observe the consequences of the gigantic geological forces at work in such places where earthquakes and volcanic eruptions release stresses building up at plate boundaries. As a species, we are just not around long enough to have any inherited memory of mountain building processes. The formulation of plate tectonics represented as fundamental a paradigm shift for geophysics as the theory of evolution had been for biology. Plate tectonics reshaped our understanding of Earth as a dynamic system on a geological time scale. Similar to the theory of biological evolution, the theory of plate tectonics continues to grow and adapt as knew technologies and creative scientists help reveal new facts and perspectives. Importantly, it also provided a tool set for a better understanding of Earth's past as well as of its potential future. Since the discovery of plate tectonics, our understanding of Earth's past has improved in many aspects. That is no coincidence, as it was plate tectonics, which revealed the dynamic processes shaping much of Earth's geological history. The tectonic plates of Earth's lithosphere rest on the mantle in what is referred to as isostatic equilibrium or isostasy.[18] In isostasy the depth to which a plate or block of crust sinks is a function of its weight and varies as the weight changes. The crust underneath a mountain range is thicker and penetrates deeper into the mantle, the asthenosphere part of it. This is not unlike floating icebergs where the larger an iceberg, the larger its volume is below water. Crust material has a lower density than mantle material, similar to ice having a lower density than water. Therefore, mountain ranges are supported by a larger mass of crust beneath providing the "buoyancy" to keep everything in isostatic equilibrium and preventing mountains from sinking into the mantle. We should be careful not to refer to tectonic plates as floating because the asthenosphere is not a liquid layer but a solid, and a ductile and viscous solid at that. Tectonic plates do not float, instead they are being pushed apart by spreading sea floors and pulled beneath adjacent plates in subduction zones. Temperature, pressure, and material largely determine the dynamics between the layers of Earth outlined in tab. 2.1. If a rock in a given layer remains solid or not depends on the combinations of temperature and pressure prevailing in this layer. These two parameters also govern material densities throughout layers. The temperature differentials between layers, specifically between the mantle and the core, are essential to drive the heat convection that provides the energy needed for plate tectonics. Earth so-called rock cycle continuously reworks our planets materials with denser rocks and mineral sinking and lighter ones rising.

The temperature in the lithosphere varies from the ambient temperature on the surface of the crust to about 1,000 °C at its boundary with the mantle. The continental crust is mostly made of granite, a rock type many of us have likely held in hand or seen used in buildings. The granite rock in the continental crust has an average density of 2.7 g/cm^3 and the dominant minerals are feldspar and quartz. The resulting chemical composition (see tab. 2.1) is ∼61% silicon dioxide (SiO_2), ∼17% aluminum oxide (Al_2O_3), ∼6% iron oxide (FeO) and calcium oxide (CaO) each, ∼4% magnesium oxide (MgO), and ∼6% of other materials including all the elements that allowed humans to build their modern-day world. The oceanic crust is mostly made of basalt rock with an average density of 3.0 g/cm^3 and the dominant minerals are feldspar and pyroxene.[19] The chemical composition of the oceanic crust is ∼51% SiO_2, ∼15% Al_2O_3, ∼11% CaO, ∼10% FeO, ∼8% MgO, with the remaining ∼5% composed of other elements (see tab. 2.1). A mixture of oceanic and continental crust can be found where two tectonic plates meet and a subduction zone forms. Depending on where this happens, at continental margins where the oceanic crust is subducted under the continental crust, or at places where one oceanic crust is subducted under another and archipelagos are formed, it is referred to as a continental arc or an island arc, respectively. The crust density in those places lies in-between that of continental and oceanic crust.

The boundary between Earth's crust and mantle is characterized by the Mohorovičić discontinuity, often simply referred to as the Moho. It is named after the Croatian seismologist Andrija Mohorovičić (1857 – 1936), considered to be one of the founders of seismology, who first identified it in 1909. Immediately above the Moho, the velocities of so-called primary seismic waves are consistent with those waves moving through basalt while below the Moho the observed wave speeds reflect their movement through peridotite rock.[20] The Moho discontinuity can be up to 500 meters thick and in some areas, ancient Moho zones are exposed above-ground where sections of the Earth's oceanic crust and the underlying upper mantle have been uplifted.[21]

The mantle is divided into a number of layers. As we have already seen, the first mantle layer, the upper mantle, is split into a solid part of the upper mantle which belongs to the lithosphere and a viscous part of the upper mantle below it that is referred to as the asthenosphere. While for a good part of the lithosphere crust we can sample the material, for example by drilling boreholes, this is not possible any more for the upper mantle. Fortunately, there are a few places on Earth where the upper mantle has been thrust on top of the crust and lies exposed.[21] The dominant rock in the upper mantle material is peridotite rock made up mostly of the minerals olivine and pyroxene.[22] The chemical composition of the upper mantle is ∼44% SiO_2, ∼37% MgO, ∼8% FeO, ∼4% Al_2O_3, ∼3% CaO, and ∼3% other elements (see tab. 2.1). The average density of the upper mantle ranges from ∼3.3 g/cm^3 at the boundary between the crust and the upper mantle in the lithosphere, to $\sim$4.4 g/cm^3 at the bottom of the upper mantle layer where the asthenosphere borders on the transition zone that lies beneath the upper mantle. On average, the asthenosphere layer of the upper mantle is around 200 kilometers thick. It has a higher density than the lithospheric part of the upper mantle and temperature and pressure increase with increasing depth as well, leading to a condition in the asthenosphere where rocks soften and partly melt. Because of that, the material in the asthenosphere is ductile and viscous, resembling flowing solid rock. This enables the tectonic plates of the lithosphere with its brittle materials to be moved in a conveyor belt fashion on top of the asthenosphere, driven by convection processes which are fed by heat transfer from deep inside the mantle. At weak spots in the lithosphere, material from the asthenosphere can reach the surface as lava erupting from volcanic fissures.

The next layer down below in the mantle stack is the transition zone followed by the lower mantle layer, and finally the core-mantle boundary layer which is also referred to as the D" (D double-prime) layer.[23] We currently still know quite a bit less about the mantle layers below the upper mantle. The general composition of rocks in these layers is thought to be similar to what we find in the upper mantle, but with increasing temperature and pressure, conditions certainly change. Rocks in the transition zone do not melt or disintegrate but become much denser. This is thought to largely prevent exchanges of material between the upper and lower mantle, possibly precluding the transport of subducted parts of the lithosphere into the lower mantle layer. While this is still speculative, if correct, subducted pieces of tectonic plates could likely sit above the transition zone for many million years. Over time this plate material would be mixed with other mantle rock before eventually the lighter materials separate and return to the upper mantle as part of the asthenosphere to eventually become once more part of the lithosphere through erupting lava or the creation of new oceanic crust at sites of sea floor spreading.

The dominant minerals that make up the transition zone are silicate spinels and garnets. Temperatures and pressures in the transition zone are sufficiently high for the mineral olivine, which is the α-phase of $(Mg,Fe)_2SiO_4$, to undergo phase transitions into denser crystal structures.[24] The respective rocks containing these high pressure, high temperature spinel phases of olivine are wadsleyite, known as β-$(Mg,Fe)_2SiO_4$ or β-spinel, and ringwoodite, known as γ-$(Mg,Fe)_2SiO_4$ or γ-spinel.[25,26] The transition from olivine to wadsleyite starts somewhere around the lower boundary of the upper mantle, at about a 410-kilometer depth and a pressure in the 13-14 GPa range, whereas the transition from wadsleyite to ringwoodite happens at a depth of around 520 kilometers and a pressure of roughly 18 GPa.[27] Remarkably, both of those minerals, wadsleyite and ringwoodite, can store water in their crystal lattices; albeit not in the usual form of H_2O but in the form of the hydroxide anion OH^-. The amount of water that scientists believe is stored in the transition zone in hydroxide form is surprisingly large and could be as much as all the water contained in Earth's oceans. How did the water get there in the first place? While we do not know for certain, the current hypothesis is that it got into the transition zone through the subduction process of oceanic plates. As temperature and pressure increase further towards the bottom of the transition zone, these "water-containing" crystal structures are broken up again. The thus freed hydroxide moves upwards through the transition zone where it is recaptured by crystals that can hold water, thereby enabling the transition zone layer to maintain a reservoir of water. The third dominant mineral in the transition zone is majorite garnet which has the chemical structure $Mg_3(MgSi)(SiO_4)_3$. Majorite garnet is thought to be formed largely through transformation of pyroxene minerals into a garnet type structure. It is a transformation rather than a phase transition as the chemical composition changes in the process. In addition to majorite garnet, a number of other garnets that have inclusions different from Mg such as Al, Fe, or Ca are also thought to be present in the transition zone.

Below the transition zone lies the lower mantle with an average thickness of a little more than 2,000 kilometers. The temperature increases further from the transition zone boundary throughout the lower mantle. Because the pressure also rises, the lower mantle is solid and has a high density. We know little about the structure of the lower mantle. At the boundary of the transition zone and the lower mantle at ∼660 kilometer depth, the pressure reaches about 24 GPa. Under these conditions further phase transitions and dissociation processes occur. One of them transforms ringwoodite into $(Mg,Fe)SiO_3$ perovskite which since 2014 is officially known as bridgmanite, and ferropericlase also known as magnesiowüstite with the general composition (Mg,Fe)O.[28,29,30] Bridgmanite, when joined with Ca in the $CaSiO_3$ perovskite crystal structure becomes silicate perovskite, the third major mineral of Earth's lower mantle. Bridgmanite is the most abundant mineral in Earth's interior making up ∼38% of its volume. Earth's lower mantle is estimated to contain about ∼73% bridgmanite, ∼17% ferropericlase, and ∼10% silicate perovskite. Bridgmanite can also incorporate a small amount of water in hydroxyl form.

The last mantle layer is the D" layer, the boundary layer sitting on top of the outer core. At the interface of the D" layer with the core the temperature has risen to about 3,700 °C. The average thickness of the D" layer is around 200 kilometers but it can vary substantially from a very thin boundary layer in some areas to having thick accumulations of iron and silicates in other areas. The dominant mineral in the D" layer is thought to be a very high-pressure phase of magnesium silicate ($MgSiO_3$) called post-perovskite. At the temperatures prevailing in the D" layer this mineral is only stable at the high pressures present in this layer, which are well in excess of 100 GPa, almost a million times the atmospheric pressure we experience on the surface. The D" layer can be molten in large areas. Parts of the D" layer seem to serve as heat and material transfer conduits to the lower mantle and transition zone and may even be erupting as upwellings of very hot rock within the Earth mantle, so-called mantle plumes. Over time, those mantle plumes may reach shallower depths and cause volcanic hotspots.

The core layer at the center of Earth is about 3,470 kilometers thick and divided into the outer and inner core layers with radial thicknesses of ∼2,250 and ∼1,220 kilometers, respectively. The outer core layer is composed mostly of iron which is thought to account for ∼80% of the outer core's mass with nickel contributing another ∼5%. Less heavy elements such as oxygen, silicon, and carbon are thought to make up most of the remaining mass together with smaller amounts of nitrogen or hydrogen. Temperatures in the outer core range from about 4,500 to 5,500 °C and densities are between 10 g/cm^3 at the top and 12.3 g/cm^3 at the bottom of the outer core layer. Under these conditions, the outer core is a highly viscous fluid. Liquid iron and nickel that moves around driven by convection processes in the outer core is the source of Earth's magnetic field but understanding exactly how this magnetic dynamo works is a theory that scientists are still refining.

However, that much seems to be clear, it is the heat transferred from the inner to the outer core which provides the main energy source for the convection currents driving Earth's magnetic dynamo in the outer core. The inner core is made mostly of iron which accounts for ~96% of its mass with a smaller amount of nickel and maybe small traces of other elements. Temperatures in the inner core reach 6,000 °C and densities average about 13 g/cm^3 with pressures exceeding 300 GPa. At Earth's center, the pressure reaches ~360 GPa which is roughly 3.5 million times the pressure we experience at Earth's surface. Under these conditions, the inner core is a solid iron crystal ball with a roughly 1,220-kilometer radius. The Moon has an equatorial radius of just a little more than 1,738 kilometers. Earth's inner core size is only about 35% of our Moon's volume but it has about 1.4 times its mass.

Our planet Earth is a much more complex, intricate, and dynamic system than what can be described in a few pages. It is only a couple of centuries ago that we started to systematically build a better understanding of our planet. There is no way to travel down through Earth's interior and study conditions close up. Unfortunately, in this case the futuristic visions the French novelist Jules Verne (1828 – 1905) published in his 1864 novel *Voyage au centre de la Terre* (*A Journey to the Center of the Earth*) remain just that. However, Jules Verne, whom many consider the father of science fiction literature, was right about humanity eventually exploring the deep oceans and traveling to the Moon. Both of which he envisioned in his novels *De la Terre à la Lune* (*From the Earth to the Moon*) and *Vingt Mille Lieues sous les mers: Tour du monde sous-marin* (*Twenty Thousand Leagues Under the Sea: A Tour of the Underwater World*), published in 1865 and 1870, respectively. Within less than a century science fiction has become reality and our increasing capability to explore the deep oceans has opened up a new window to study the geology and life of our deep seas. We can appreciate how difficult the challenge of exploring the deep seas is when we consider that it was seemingly easier for humankind to first travel to the Moon before venturing to the very bottom of our oceans. As we will see in the next section, traveling to the Moon has taught us at least as much about Earth early history than about the geology of the Moon itself.

Science and technology have progressed enormously over the past centuries enabling us to study materials in our laboratories materials under the very high-temperature, high-pressure conditions prevailing in Earth's interior. At the same time, our tools for probing Earth's interior, mostly seismology, have become much more precise and accurate. Compared to just one century ago we do have a much-improved understanding of our planets interior and the processes driving the mechanics of plate tectonics. In parallel, our knowledge of Earth's geology, present and past, has vastly improved, enabling scientists studying Earth rocks and minerals to read Earth's geological record and through that our planets evolution. Only recently has it become clear that our planet and life on it have essentially co-evolved. Geological evolution and the evolution of life are not two different stories but intricately woven into each other. Today, climate change is educating us about just how close the link between the dynamic system Earth and life on Earth is. We have not touched on this topic yet as we have left out Earth's atmosphere in our introduction to Earth. We will get to it in time, once we have gained a better overview of our planet's evolution.

Planetary Evolution

Blue Marble Earth

One of the great scientific paradigm shifts in understanding life on Earth has largely gone unnoticed, and it is not the concept of biological evolution as pioneered by Charles Darwin more than 150 years ago. It is the much more recent understanding and appreciation that the canvas on which we have been depicting the evolution of life has itself played a major part in enabling life. Biological evolution on Earth is only the last stage in a very long process that started out around 13,800 million years ago. The early universe knew only two chemical elements, mostly hydrogen and a smaller amount of helium. It took the life cycles of countless stars, the first ones presumably formed some 1.6 million years after the *Big Bang*, to produce all the other chemical elements we recognize from the periodic table.[31] As we saw in the previous chapter, the universe forges elements heavier than hydrogen up to the metal iron in stellar nuclear furnaces through stellar nucleosynthesis. Once stars have depleted their nuclear fuel, they seed the universe with their chemical produce through processes such as the formation of planetary nebulae and supernova explosions; the latter adding elements heavier than iron through supernova nuclesynthesis. In this way, each generation of stars enriches stellar clouds with materials heavier than hydrogen and helium. Such chemically enriched stellar nurseries in our galaxy provided the conditions for the formation of later generation stars such as our Sun and the planets around it, which we call our solar system.

Only over the last two generations or so have we learned how critical geological and mineral evolution as well as chemical evolution have been for enabling biological evolution on planet Earth. In essence, what we are learning is that those different facets of evolution are inseparable. Without stellar evolution, there would be no heavy elements to build life from such materials as carbon; and there would be no solar systems formed with rocky planets orbiting a star such as our Sun. Without planetary evolution, there would be no biological evolution and without biological evolution, our planet would not look like the blue marble Earth that we first glimpsed in 1972 when the Apollo-17 astronauts photographed Earth from a 29,000-kilometer distance.

Before there could be any thoughts about Earth's planetary history, Earth had to be recognized a one of the planets of our solar system. From the records we have, the idea of a spherical Earth seems to have originated with Pythagoras of Samos (c. 570 – c. 495 BCE) in late sixth-century ancient Greece.[32] Before that, even the ancient Greek sages believed in some versions of a flat Earth model as we already saw earlier with Anaximander, the student of Thales of Milet. Likewise, other ancient people had their own versions of a flat Earth model. This includes the ancient cultures of Sumer, Akkad, Babylonia, or Egypt in the Near East as well as the ancient peoples of India and China. Envisioning

Earth as a flat slab floating in water must have seemed natural. In China, the belief in a square Earth under the cupola of a celestial half-sphere would last well until after the *Copernican Revolution* had completely altered the perspective on Earth and our solar system in Western Europe. Eventually, it became clear to many that the Earth is a spherical body, and some found early proof for that, as did Aristotle.

However, understanding that Earth is a spherical body and, as we learned with Copernicus, that it is just one among several planets orbiting our Sun, is only the first step. Understanding planetary evolution requires a different concept of time. The concepts of time we can find in most mythological accounts, whether those of ancient Greece or those of other ancient cultures around the world, speak for themselves. In many such accounts, Earth as a planet is created in what amounts to humanly comprehensible time spans or even shorter than that, sometimes even in days. The *Book of Genesis* for example would use one of those accounts to present believers with a divine creation accomplished in some six days. Such accounts should of course not be taken literally as they carry symbolic meaning. For those seeking to understand our world through religious texts, the latter do have the advantage of being much easier to comprehend as they turn what seemingly cannot be explained into a matter of faith. For others not believing in creation stories, quite on the contrary, Earth often seemingly was eternal, just as Aristotle thought. While closer to the truth as compared to human lifespans, it was not any more helpful towards gaining an understanding of Earth's planetary evolution.

Given those concepts of time, it should be no surprise that regarding geological history the ancient Greeks had not much to say. Similar to other peoples of the ancient world their sense of history derived from how far they could look back into their own ancestry. Like how many kings they could trace back in their people's history, be they real or mythical kings. As for the age of Earth itself, that seems not to have come to mind as what mattered much more in ancient creation stories was how to legitimize the rule of kings or pharaohs. That frequently tied the origin of Earth and usually also the creation of the larger cosmos to a divine ancestry that also was responsible for establishing the cosmic order. That must have put the ancient Greeks at some disadvantage, as for example the pharaohs of ancient Egypt could count their ancestors much farther back. The ancestry of the ancient Greeks was of a more recent origin. They anchored their history in Mycenaen Greece, a time familiar to most of us through Homer's songs of the Trojan wars, a time when gods still fought alongside men in battles.[33,34]

While their history did not lead the ancient Greeks to a different conclusion with respect to Earth's past than other ancient peoples, their philosophical speculation did. Again, we find Aristotle foremost among them as he argued that the cosmos and Earth with it were eternal and remained unchanged throughout time. Considering our short human lifespans during which little of geological significance changes, if anything at all, this perception is understandable; and it also reflects the true cosmological time scales more than any creation mythologies. The discussion about the potential eternity of our world would continue long after Aristotle. In the thirteenth century, in one of the few examples where Christian theologians firmly disagreed with Aristotle, the Catholic Church finally declared the teaching of an eternal world a heresy as it contradicted the Bible. With that, the age of Earth became a matter of interpreting the Bible and that would not change for the next roughly four hundred years.

Until a few hundred years ago, European scholars believed Earth to be only several thousand years old; and they likely thought that their answers gained from Bible interpretations to be no less precise than we believe ours to be. The danger with being precise is that for the chance of being precisely right one risks being precisely wrong. That can easily happen if one for example gets the facts wrong or confuses religion with science. The famous and often cited case in point with regard to the age of Earth is James Ussher (1581 – 1656), Archbishop of Armaghif and Primate of all Ireland. In his famous or infamous work, depending on your point of view, the *Annales veteris testamenti, a prima mundi origine deducti* (*Annals of the Old Testament, deduced from the first origins of the world*) published in 1650, he calculated the date of creation to have been nightfall on October 22, 4004 BCE. Now that is being precise, or better, being precisely wrong. We should not hold this error, which did not harm anyone, against him. He was a son of his times and a child of his church and there was at the time nothing wrong for a catholic archbishop to engage in this kind of calculation. However, the sad part is that more than three-and-one-half centuries later there are still too many who believe our planet Earth to be only a few thousand years old. There is no such excuse for them, and we will not give them space here but just say that it reflects a sorry state of humanity if religious beliefs when confronted with the facts of reality still can win out and leave people ignorant and susceptible to exploitation by those who benefit from keeping them in this state.

A century after Ussher's pronouncement it became clear to some that Earth must be much older because they could see the very processes forming Earth in their own time working on time scales much longer than a few thousand years. In his *Les Époques de la Nature* (*The Epochs of Nature*), published in 1778, the French naturalist and encyclopedic Georges-Louis Leclerc, Comte de Buffon (1707 – 1788), clearly disagreed with the few-thousand-year history that the Bible allowed for Earth. He assumed that the collision of a comet with the Sun created Earth in the process. In his view, there were several periods in Earth history. First came a period of mountain forming; next a period where oceans completely covered the Earth followed by volcanism once the oceans receded; then came the emergence of large elephants; followed by a time where enormous landmasses broke up to form today's continents; and finally came the emergence of higher animals such as humans. While he was wrong on several accounts, some of his views definitely sound almost modern. For all of this to happen, Comte de Buffon assumed that Earth must be about 75,000 years old. While this is much longer than the few-thousand-year age that Bible scholars such as Ussher believed in, it still underestimated the true age of Earth by a factor of 60,000, illustrating just how difficult it is to grapple with the vast expanse of geological time.

We already encountered James Hutton, the Scotsman considered the father of modern geology and the originator of the theory of *Uniformitarianism*. As we saw earlier, *Uniformitarianism* essentially states that the same forces and natural laws that Hutton could see at work in geology in his time did not change throughout Earth history. This meant that it must have taken a very long time for today's geology to form. As we enter the eighteenth century, most geologists had become convinced that Earth must be many millions of years old. Lord Kelvin whom we already met earlier, came up with another way to estimate the age of Earth. He assumed that Earth was originally molten and then calculated when Earth could last have been in a molten state. To achieve this, he conducted field measurements of Earth's temperature in mine shafts, studied the melting of rocks, and made assumptions about heat transport in the Earth through conduction and radiation. His initial calculations in the early 1860's gave a rather large range estimate from ∼20 to 400 million years which by 1892 he corrected to about 100 million years to finally arrive in 1897 at the conclusion that Earth could not be much older than ∼20 million years.

Scholars used a number of other approaches to estimate the age of Earth. Among them are the time it took the oceans to become as salty as they are today, or the time required to produce the fossil record of evolution as preserved in the various geological layers. Some of these came up with similar ranges for the age of Earth while others indicated our planet to be somewhat older but none of the estimates for the age of Earth went beyond 100 million years, expect for one related to sedimentation. Measuring the thickness of sediments and using current sedimentation rates to calculate how long it took to form the thick sediments known to geologists, raised the number of Earth's age into the 500-million-year range. However, estimates based on sedimentation rates came with large uncertainties because it was not clear how one had to account for the erosion that was taking place in parallel to sedimentation.

The first accurate geological dating methods became only available after the 1896 discovery of radioactivity by the French physicist Henri Becquerel (1852 – 1908). In 1904, the New Zealand-born British physicist Ernest Rutherford (1871 – 1937) was the first to understand that radioactive decay represented a heat source that had continued to heat Earth throughout its history. Lord Kelvin's earlier assumption that Earth had no such heat source was therefore incorrect and so was his calculation of Earth's age. Importantly, Rutherford realized he could use radioactive decay for very precise dating. When rocks form, radioactive minerals inside them essentially become clocks that allow scientists to establish their respective ages. One such clock is the decay of uranium into lead. The American chemist Bertram Boltwood (1870 – 1927) built on Rutherford's work and in 1907 published the first dating of rock samples based on radioactive decay. This famously included rocks from the Cambrian, a geological time period modern dating methods bracket between 541 and 485 million years ago, but also rocks found to be older than 2,000 million years. From then on, it only became a question of refining radioactive dating methods and finding older and older rocks to determine the true age of the Earth. In 2017, a 4,280-million-year-old rock from Canada, discovered in an area known as the Canadian Shield, took the title of oldest known rock on Earth.[35] The age of this 4,280-million-year-old rock still falls short of Earth's true age of 4,567 million years. The problem is that the farther one goes back in Earth history the more difficult it becomes to find rocks from that time. This is because Earth continuously recycles its crust. Therefore, we never may find any Earth rocks that are as old as Earth.

Our solar system took shape about 4,600 million years ago and the birth of our Earth dates to 4,567 million years ago. How do we know that Earth is that old and how can we pinpoint the age of the Earth so precisely? As we will see shortly, scientists discovered the answer to this question not in the rocks of Earth but in meteors falling onto Earth. But first, we need to look at the concept of geological time and how it maps out the 4,567 million years of Earth's geological history.

The Concept of Geological Time

Geological time is the canvas on which Earth scientists have been painting the picture of Earth's evolution starting in the late eighteenth and early nineteenth century, roughly some two hundred years ago. It is a time scale that has been built on the insight, first formulated by Nicolas Steno, that rocks are laid down in successions of layers referred to as "strata", each representing a slice of time.[36] Since then, a number of geological time periods have been identified based on what scientists agreed as their sufficiently differentiated stratigraphy. This could be the composition of the layers themselves or events marking major geological or paleontological changes differentiating such layers. An example for the latter would be a mass extinction, like the one that killed the dinosaurs some sixty-six million years ago. This event is marked by a thin deposition of the chemical element iridium, laid down when the asteroid causing the extinction of dinosaurs hit Earth. As with all classifications schemes, the first versions of geological time scales were quite simple but as scientists learned to read the geological record, a significantly more differentiated picture of Earth's geological history emerged. Some of the names given to geological time intervals remained the same while others were changed, and all of them were subdivided further. Methods for dating rock samples and fossils continuously improve, and geological time stamps are adjusted accordingly. The community of Earth scientists keeps refining the geological time scale as they agree or disagree on the significance of new discoveries in paleontology or geology. Sometimes those changes happen quickly; sometimes debates go on for many years until an issue is finally settled. Plate I through plate VII show a simplified version of the geological time scale, which will be sufficient for the purpose at hand. The longest time interval on this geological time scale is the geological *Eon*, represented by the bars on the left most side of the plates. To the right of the *Eons* follow the subdivisions into geological *Eras*, geological *Periods*, and geological *Epochs*. Not all geological *Eons* are equally subdivided, the earliest one is for example not subdivided at all. Similarly, not all geological *Eras* and *Periods* are equally subdivided. The finest level of detail in the geological time scale shown here is the geological *Epoch*. *Epochs* are further subdivided into geological *Ages* but that is a level of detail too much for what is of interest here. In some instances the naming of geological *Epochs* used here also stays with a simpler older form that subdivides a given geological *Period* into an *Early-*, *Middle-*, or *Late Epoch*. This makes the geological time scale easier to read without losing too much information and in addition avoids having to consider still ongoing naming discussions which can be controversial, and which are really for the experts. However, for a number of geological epochs specific naming schemes have long been established and they are used here accordingly. Some geological time scales consider even longer time intervals than the geological *Eon* of which there is currently only one named example, the Precambrian *super-Eon*. For paleontologists the Cambrian period holds a special significance as the so-called Cambrian explosion of life marked for a long time the emergence of life on Earth as documented by the large number of small shelly fossils found in Cambrian rock layers and virtually none before that. Because of that, the Precambrian was practically seen as devoid of life and therefore identified with the time on Earth when there was no life at all. However, for quite some time now, we know that this is not correct.

There are four eons recognized in Earth's geological time scale: the Hadean, the Archean, the Proterozoic, and the Phanerozoic eon. The Hadean eon spans the time from Earth's formation 4,567 million years ago to 4,000 million years ago. The naming of this eon was likely inspired by what Earth must have looked like back then, as *Hades* is the name ancient Greeks gave to the god of their underworld. The end of the Hadean eon and the beginning of the Archean eon was initially assumed to mark the formation of Earth's first solid crust. We know today that this already happened earlier during the Hadean eon. The Archean eon, which simply translates from Ancient Greek into the *"ancient eon"*, lasted for 1,500 million years, ending 2,500 million years ago. Next comes the Proterozoic eon, which is the longest of the four eons lasting from 2,500 million years ago to 541 million years ago. The name of this eon, Proterozoic, is derived from a combination of the Ancient Greek words for former or earlier, *protero*, and for life, *zoe*. It was set to span the time from 2,500 million years to 541 million years ago, from the appearance of oxygen in Earth's atmosphere to just before the time scientists believed back then that complex life had started to proliferate on Earth. Today we know that complex life was present on Earth significantly earlier than the end of the Proterozoic eon. However, the start of this eon still correlates with the beginning of the so-called Great Oxygenation Event or GOE, which by about three hundred million years into the Proterozoic eon had given Earth an atmosphere with a roughly 1% oxygen content. As indicated above, the Hadean, Archean, and Proerozoic eons are sometimes referred to as one geological super-eon, the Precambrian, presumably comprising the time on Earth - as it was believed before we knew better - preceding the emergence of complex life. Finally, the fourth eon in Earth's geological time is the one we still live in today, the Phanerozoic eon. It started 541 million years ago, and its name is a combination of the Ancient Greek words for visible, *phaneros*, and for life, *zoe*, literally translating into the eon of visible life. This naming really sets it into contrast to the Precambrian super-eon preceding it for which no traces of life were visible for a long time.

The Hadean eon has no subdivisions, a wise admission that we just know too little to break this roughly 567-million-year time interval down into meaningful smaller time periods. The Archean eon is subdivided into four eras starting with the Eoarchean era, followed by the Paleoarchean, the Mesoarchean and the Neoarchean eras. It almost feels as if there was a perceived need to subdivide the Archean eon, this vast space of 1,500 million years, somehow. However, as the names for the eras indicate, there are no real markers that we have to differentiate them so these eras simply create categories for an *"early-ancient"* or Edoarchean era, an *"old-ancient"* or Paleoarchean era, a *"middle-ancient"* or Mesoarchean era, and a *"young-ancient"* or Neoarchean era. While initially, the beginning of the Edoarchean was thought to correlate with Earth's first solid crust we now know that this actually happened already earlier in the Hadean eon. Currently, the four Archean eras are not subdivided further into periods.

Following the Archean eon, the Proterozoic eon is broken down into three eras, starting from the *"old-proterozoic"* or Paleoproterozoic era, through the *"middle-proterozoic"* or Mesoproterozoic era to the *"young-proterozoic"* or Neoproterozoic era. The Paleoproterozoic era lasted from 2,500 million years to 1,600 million years ago. It is subdivided into four periods: the Siderian, the Rhyacian, the Orosirian, and the Statherian. The Siderian lasted for 200 million years, deriving its name from the Ancient Greek word for iron, referencing the deposition of so-called banded iron formations at the time. The next period, the Rhyacian lasted for 250 million years and it literally translates from Ancient Greek to "lava flow" indicating that the Rhyacian was a period of high volcanic activity. The third and fourth periods in the Paleoproterozoic era, the Orosirian and the Statherian, lasted for 250 and 200 million years, respectively. These period names derive as well from Ancient Greek words with Orosirian reflecting the extensive mountain building during this time and Statherian referring to a stable period following the geologically more active periods of mountain building and of increased volcanism preceding it.

The Mesoproterozoic era stretches from 1,600 to 1,000 million years ago. It is subdivided into three periods of equal length of 200 million years each: the Calymmian, the Edasian, and the Stenian periods. The Calymmian period derives its name from the Ancient Greek word for "cover". It is a geological period, during which layers of igneous or sedimentary rock accumulated over the oldest parts of the continental crust, the so-called cratons, in formations referred to as platform covers.[37] The Ectasian period following it derives its name also from an Ancient Greek word, in this case for "extension", indicating an extension of continental landmasses during this period. The last period in the Mesoproterozoic era, the Stenian, meaning "narrow" in Ancient Greek, carries its name because of the narrow belts of rock formations dating from this period.

The final era in the Proterozoic eon is the Neoproterozoic era, which began 1,000 million years ago. It lasted until 541 million years ago, marking the end of the Proterozoic eon and of the Precambrian super-eon. The somewhat more unequal division of the Neoproterozoic era into the three periods of the Tonian, Cryogenian, and Ediacaran reflects our more detailed knowledge for these geological time periods. The Tonian lasted for 280 million years. Its name derives from the Ancient Greek word for "stretch" and it is a reference to the break-up of the supercontinent Rodinia during this period.[38] Most of us would guess that the name for the Cryogenian period must have something to do with cold and indeed, it derives from the respective Ancient Greek word for "cold". The Cryogenian lasted for 85 million years and was characterized by a nearly worldwide glaciation of Earth. The Ediacaran, the final period marking the end of the Proterozoic eon lasted for 94 million years. It is the first period deriving its name not from an Ancient Greek word. Instead, it is named for the Ediacara Hills of South Australia. It was the fossils found in these hills, which finally provided ample evidence for the existence of complex life in the Precambrian.

With the end of the Precambrian super-eon 541 million years ago, we enter the Phanerozoic eon, the eon that scientists once believed to witness the first forms of visible life and the eon that lasts to this day. The Phanerozoic eon is subdivided into three eras, the *"old-life"* or Paleozoic era, the *"middle-life"* or Mesozoic era, and the *"new-life"* or Cenozoic era; the Cenozoic being the area we currently still live in. The Paleozoic era began 541 million years ago and lasted for 289 million years up to 252 million years ago. This geological era is itself subdivided into six periods. These are the Cambrian, the Ordovician, the Silurian, the Devonian, the Carboniferous, and the Permian periods. The names of these periods are testament to the leading role Victorian scientists and scholars of geology, botany, and biology played in the nineteenth century in the evolving fields of geology and paleontology. Going out with a rock pick to explore the local geology or search for fossils was very much seen as a worthy occupation for a gentleman as it was seen equally suitable for Ladies to indulge in botanical exploration. Naming geological periods was in a sense almost the culmination of a geologist's career and frequently contentious. Adam Sedgwick named the Cambrian after Wales, using the Latin form of its Welsh name, Cymru. The English geologist Charles Lapworth (1842 – 1920) named the Ordovician period after the Celtic tribe of the Ordovices. Similarly, the Silurian period was named for another Celtic tribe, the Silures, this time by the Scottish geologist Roderick Murchison (1792 – 1871); he did so in the very same publication in which Adam Sedgwick named the Cambrian period. Naming of geological periods was hard fought over and if Murchinson could have had his way, there would not have been an Ordovician but only a Silurian with some more subdivision. However, he compromised as in the end he was beholden to geological facts. Compromise is also how the naming of the Devonian period came about, which Sedgwick and Murchison jointly proposed after their arguments finally won out in a long-standing debate regarding the geology of the county Devon. The English geologists William Conybeare (1787 – 1857) and William Phillips (1775 – 1828) named the Carboniferous period. Its name derives from a combination of the Latin word for coal, "carbo", and "fero", with the meaning of "to bear" or "to carry", referring to the coal-bearing layers typical for this period. Given the importance of coal to the Victorian age, it is no wonder that scientists named the Carboniferous period for the coal deposits laid down during this geological period. The final period of the Paleozoic, the Permian, was again named by Roderick Murchison, this time after the city of Perm near the Ural Mountains in reference to the rock formations overlaying the Carboniferous geology in this eastern part of Russia.

The Cambrian period lasted for 56 million years and it is characterized by an abundance of life preserved in the fossil record, which seemingly came out of nowhere Only more recently has it become clear that the Cambrian does not mark a boundary before which there was no complex life. In the Ordovician, which lasted for 41 million years, we see first plant life moving from the sea onto land. This marks the slow start of a transition that eventually, after many millions of years, would transform weathered barren rock landscapes into the green planet we know today. During the Ordovician, the oceanic ecosystems started to look like what would seem to be familiar today but populated by different lifeforms occupying the respective ecological niches. The Ordovician ended in a glaciation period that extended into the early Silurian. Although the 25 million year long period of the Silurian started out cold, this quickly changed, and we find the oldest fossilized land plants. Animal life such as arthropods, the distant ancestors of spiders and scorpions, made the move onto land as well. The Cambrian and the Ordovician are subdivided here into three geological epochs referencing the early, middle, and late phases in these geological periods: the Early-, the Middle-, and the Late Cambrian epoch; and the Early-, the Middle-,and the Late Ordovician epoch. The Silurian is subdivided into four epochs with distinct names: the Llandovery, the Wenlock, the Ludlow, and the Pridoli epoch. These names are tied to town or place names in Wales (Llandovery), England (Wenlock and Ludlow), and the Czech Republic (Pridoli). The Devonian lasted for 60 million years during which land plants and invertebrates continued to diversify and the first amphibians appeared. Here, it is broken down into an Early-, a Middle-, and a Late Devonian epoch, referencing the early, middle and late phases in the Devonian period respectively. The Carboniferous lasted also for 60 million years. Its coal forests witnessed the emergence of winged insects and reptiles as the supercontinent Pangea began to form. The Carboniferous period is broken down into the Early- and the Late Carboniferous epochs, which are also referred to as the Mississippian- and the Pennsylvanian epochs. The 47-million-year-long Permian period sees the fully formed supercontinent Pangea and the end of it is marked by the mass extinction event known as the *"The Great Dying"*. The Permian period is here subdivided into the Early-, the Middle-, and the Late Permian epoch, which identify the early, middle and late phases of this period. These three epochs are also known as the Cisuralian-, the Guadalupian-, and the Lopingian epochs. The name Cisuralian derives from a Latin phrase meaning *"on this side of the Ural mountains"*, the latter being formed during the final assembly of Pangea. The Guadalupian references the Guadalupe Mountains of West Texas. This mountain range incorporates the fossilized remains of a large reef built up during the Permian, which in much later times was uplifted to become part of what are today the Guadalupe Mountains. As for the Lopingian, this epoch was named by the American Amadeus W. Grabau (1870 – 1946) who is considered the father of Chinese geology and derives from the name of the city of Leping in the Jiangxi province of China.

The Mesozoic era of the Phanerozoic eon lasted for 187 million years, from 252 to 65 million years ago. Its three periods are the Triassic, the Jurassic, and the Cretaceous. The name of the 51-million-year-long Triassic period is a translation of the Latin period name Trias, meaning triad or three, which was introduced by the German geologist Friedrich August von Alberti (1795 – 1878) to describe three characteristic strata of rock types he found in Germany. The Triassic period is subdivided here into the Early-, the Middle-, and the Late Triassic epoch and it famously witnessed the first emergence of dinosaurs. The Jurassic lasted for 56 million years and derives its name from the Jura Mountains along the French-Swiss border, which are part of the European Alps. Though the limestone rocks of the Jura Mountains formed during the Jurassic, the mountain range itself along with the rest of the Alps only rose much later. During this period, about 175 million years ago, the supercontinent Pangea started to break up. Like the Triassic, the Jurassic period is also split here into an Early-, a Middle-, and a Late Jurassic epoch. The Cretaceous period lasted for 80 million years and its name derives from the Latin word for chalk, "creta", the type of rock formation typical for this period. In North America it witnesses the first uplift of the modern Rocky Mountains and the rise of the ancestral Sierra Nevada. This last period of the Mesozoic era is subdivided here into the Early- and the Late Cretaceous epoch. It is during the Cretaceous period that the first flowering plants and the first mammals appear. The end of the Cretaceous period and with it the end of the Mesozoic era is marked by the asteroid impact, which caused the extinction of the non-avian dinosaurs.

The Cenozoic era of the Phanerozoic eon is our current era, beginning some 65 million years ago. The Cenozoic is divided into three periods, the Paleogene, the Neogene, and the Quaternary, our current period. The Paleogene period lasted for about 42.5 million years. Its name, deriving from Ancient Greek, refers to the *"old-born"* or Paleogene period. This is in recognition of a similarity of marine fossils found in this period, which contrast with fossils from the following *"young-born"* or Neogene period. The Paleogene is divided into three epochs, indicating the *"old-recent"* or Paleocene epoch, the *"early-recent"* or Eocene epoch, and the *"slightly-recent"* or Oligocene epoch.

The Paleogene period saw a diversification of flowering plants and mammals, including the rise of large mammals. The following period, the Neogene, lasted for just a little longer than 20 million years. It is divided into two epochs, the *"less-recent"* or Miocene epoch and the *"more-recent"* or Pliocene epoch. With the Neogene period, we enter the age of modern plants and large mammals, the latter also including our very first human ancestors. The final period of our current Cenozoic era, the Quaternary, has now lasted for about 2.59 million years. The name of this period is a legacy left by the Italian geologist Giovanni Arduino (1714 − 1795) who devised a stratigraphic system that divided Earth history into four periods: the Primary, the Secondary, the Tertiary, and the Quaternary. These periods where associated with different rock formations he identified in the Alps. Because some of them are unique to the Alps this quickly led to his first two categories being discarded. However, the Tertiary period survived until recently, comprising what today are referred to as the Paleogene and the Neogene periods. Sometimes these two are still referred to as the Lower or Early Tertiary and the Upper or Late Tertiary. What remains of Giovanni Arduino's scheme is the naming of the Quaternary period. The Quaternary period is divided into two epochs, the *"most-recent"* or Pleistocene epoch, and the *"wholly-recent"* or Holocene epoch, which we still live in. The Pleistocene saw the spread and rise of humans and repeated glaciation events. By the end of the Pleistocene, humans had established themselves on all continents and almost everywhere, except Africa, they had hunted most large animals to extinction. With the Holocene period, we see the beginning of a new warm time and the slow transition of humanity from hunter-gatherers to herders and agriculturalists followed by the first human civilizations built around city-states or along riverbanks.

Schematics are always somewhat dry and the classification of geological time is no different in that respect. However, some familiarity with naming conventions and some of its background will be useful for taking our actual walk through geological time itself. Before we do so, a few comments are in order regarding the time stamps shown in plate I through plate VII. For the respective geological time intervals themselves and some of the major events that are listed next to them, radiometric rock dating has become an exact science. Because of that, errors in dating, even of the oldest rocks, are quite small. Geochronologists are chemists specializing in the accurate dating of rocks and they have achieved a very high accuracy for dating the oldest known rocks on Earth. These rocks are 4,567-million-year-old meteorites and the error margin for this number is only plus or minus two million years or less than a 0.01% error. Therefore, the numbers shown in plate I through plate VII bracketing the geological time intervals are quite precise when based on rock dating, which practically all of them are nowadays. For the selected events highlighted in this chronology, the dates reflect our current best understanding. Many of them derive from rock dating but some of them do not and their respective dates will continue to be refined. If the picture of the fossil record changes in the future, it may also affect the classification of the geological time intervals. It is best to look at geological time and the major events shown in these plates as an accurate calendar where we have a good idea of some of the events recorded but are sometimes not quite sure if we entered them on the right day. However, overall, as we look through the calendar in time, it still gives us a good sense of what happened in the past as long as we have been diligently working to correctly enter the events and update them when necessary.

A Brief Walk Through Geological Time

The solar system was born out of the gravitational collapse of parts of a giant interstellar cloud of dust and gas. The nebular disk formed in the collapse contained mostly hydrogen, helium, and dust. The question is: what was that dust made of? This dust must have contained all the materials, or better minerals, which the initial solar disk contained, and which eventually agglomerated into spherical grains of mineral dust. Scientists refer to those grains as chondrules.[39] When first baked together into small pebbles some 4,567 million years ago, they became the oldest known rocks in our solar system, the chondrites. This sets the beginning of the first geological eon on Earth, the Hadean eon, which ended 4,000 million years ago. Most meteorites, over 80% of them, are of the chondrite type and represent the oldest rocks in our solar system. How Chondrules and chondrites formed we do not know yet. Neither are we sure which early solar system heat source forged chondrites into the meteorite materials we find today. The inner solar system is thought to have been very hot during the formation of the planets. Within today's Earth orbit, temperatures would have been in the range of roughly 800 °C to 1400 °C. This would have made it impossible for agglomerated dust particles to survive unless they possessed high melting temperatures. The materials that fall in that class are precisely the silicate minerals that make up today's Earth.

Compared to Earth, the inner rocky planets in the making, Mercury and Venus, must have seen higher temperatures and the outer rocky planet Mars somewhat lower temperatures during their own planetary growth processes. In contrast, further away from the Sun, low temperatures allowed volatile gases to survive and be stable in condensed solid form such as for example methane. The initial blast from the solar nuclear furnace helped differentiate the material distribution in the collapsed nebular disk further, essentially cleaning out the inner solar system of volatile gases. The result was that dust made of silicate minerals dominated in the inner solar regions and provided the starting materials from which eventually the rocky planets Mercury, Venus, Earth, and Mars would form. In the outer regions of our solar system volatile gases dominated, mostly hydrogen with helium coming in second but also a number of other gases in smaller amounts. There, they aggregated to form the gas giants Jupiter and Saturn as well as the ice planets Uranus and Neptune.

Just how the agglomeration of minerals into chondrules got started that would eventually result in rocky planetary bodies remained a mystery for a long time. It was not clear what could make small particles stick to each other in the turbulent environment of the early solar system. Gravity between such small particles was just too weak to achieve that. Modern science has taught us that tiny grains of minerals tend to stick together due to charge exchanges and the resulting electrostatic forces between them. The hypothesis is that the protoplanetary dust cloud in which they were moving around was, at least locally, not electrically neutral. Once chondrules had formed, the next step was for them to aggregate into chondrites, the meteorite material that would be the earliest rocks in all rocky planets, including Earth. Therefore, by dating meteorites of the chondrite type, the rocky material that provided the seeds for the growth of planet Earth, scientists are able to calculate the birth date of Earth. Our planet is still subject to substantial bombardment from space by mostly smaller objects but sometimes also larger ones. Material from the Allende meteorite, which came down in the state of Chihuahua in Northwestern Mexico in 1969, contained so-called carbonaceous chondrites.[40] The meteorite was dated to be 4,567 million years old and this is how we know the age of our planet Earth. Of course, from this birth date it took some time for Earth to grow to the size of proto-Earth, a name often used to refer to Earth before it acquired the Moon. Initially, the growth of proto-Earth is thought to have been driven by accretive collisions that added to the size of the growing rock instead of fracturing it. Exactly how proto-Earth grew to the size of a so-called protoplanet, objects typically on the order of a thousand kilometers wide, is not fully understood yet.

One theory is that protoplanets form just outside of what astronomers call the snow line, sometimes also referred to as the frost line or ice line, of the protoplanetary disk. This snow line marks the distance from a star where the star's radiation density or to put it differently, its heat flux, has fallen sufficiently such that liquid water turns into ice. Outside this snow line perimeter dust grains start to be covered by ice. First, this ice will be water ice but as one moves farther away form the Sun other volatile gas compounds such ammonia, methane, carbon dioxide, or carbon monoxide also condense on dust grain surfaces. Hence, there is not just one snow line but a sequence of snow lines as with decreasing temperature the heavier gases start to condense. These ice-covered dust grains can agglomerate into large icy bodies much more quickly than their counterparts just inside the snow line perimeter in the liquid zone. However, the snow line of our solar system where the protoplanet of Earth formed was in a different position than today's snow line where liquid water turns into ice.

The early Sun was burning less bright and the solar nebula surrounding it was an opaque cloud, resulting in a different radial temperature profile as one moves away from the Sun than what we can observe today. Today's snow line at $\sim$5 AU from the Sun is further out than the snow line at the time when protoplanets first formed. Scientists believe that the snow line where the one protoplanet formed that eventually would become Earth was located $\sim$2.7 AU from the Sun within today's Asteroid belt, somewhere between the orbits of Mars and Jupiter. From there, this protoplanet and the protoplanets to become the seeds of the other rocky planets must have been pulled in by the Sun to eventually occupy the orbits they follow around the Sun today. But how did this protoplanetary seed of Earth grow into its current size?

For quite some time, scientists thought that once a protoplanet's mass had grown sufficiently, gravity would have kicked in to accelerate its growth. Instead of relying on chance encounters to drive the growth of the nascent planet, gravity would have started to pull in rocky objects in its neighborhood. In such a scenario, the protoplanet would eventually clean out its orbit of anything else in its path it could either hit directly or add in clumps to its growing size through its increasing gravitational pull. Mostly this was thought to have happened by adding other small planetary bodies called planetesimals, objects whose size starts to exceed one kilometer in diameter. Planetesimals themselves are large enough to grow through gravitational accretion. Therefore it was assumed that in a process where the bigger ones grow at the expense of the smaller ones, protoplanets and planetesimals grew until there were no more small rocks or dust left; and eventually protoplanets must also have gobbled up planetesimals, leaving us with the four rocky planets we know today.

However, there was a problem. It would have taken a much longer time for this hypothetical planetary growth mechanism to produce the results we see today than was available in the time window within which we know that the planets actually formed. A planetary growth model was needed that could achieve this result much faster. The model that scientists now believe is closest to the correct one is planetary growth by pebble accretion [10,11]. Instead of large clumps being added as protoplanets sweep their orbits, this model has planets grow by adding pebbles to their mass, many of them and very quickly. The challenge with fast accretion is to find a model that reduces the speed differential between a protoplanet and an object passing it close by, such that the gravitational pull of the protoplanet will be sufficient to add the slowed-down object to the protoplanets mass. The solution to this problem are the gas clouds surrounding protoplanets. The diameter of these clouds is much larger than the diameter of the respective protoplanets, effectively increasing the volume from which a protoplanet can pull in more material to add to its mass.

As faster moving small pebbles pass through this gas cloud, they are slowed down by friction towards the speed of the protoplanet and this very much increases the chance for such pebbles to be gravitationally captured by the protoplanet. Larger objects such as planetesimals, of which there are much fewer than pebbles, are little affected by the gas surrounding a protoplanet and their likelihood of capture therefore does not change. Because most of the mass in the accretion disk in which a protoplanet orbits its star is made up by small pebbles, protoplanets can grow very quickly. As a protoplanet quickly grows in size it will be able to capture ever-larger pebbles being slowed down by its surrounding gas cloud or larger rocky bodies passing close by that its previously smaller mass would not have been able to capture. Growth by pebble accretion is still a model of the big eating the small but with a twist. First, the big eat the vast mass of many small objects and this helps them grow very fast; then they can increasingly attract also much bigger objects and add them to their mass. Protoplanets essentially sweep their orbits clean of any rocky debris. Protoplanetary Earth was likely sharing its orbit around the Sun with other protoplanets or large planetesimals circling the Sun at similar distances. However, Earth must have been growing faster than any other protoplanet in its orbital plane, so fast that it eventually captured all of them before they had grown too large.

With protoplanets, size is everything and the larger ones grow at the expense of the smaller ones; and small initial differences in protoplanetary mass may have resulted in Earth growing much faster than other protoplanets in a similar orbit. Eventually, by probably not much more than ten to twenty million years after rock accretion from chondrites got started, nascent Earth reached proto-Earth status, the size it had before acquiring the Moon. What did Earth look like at this stage called proto-Earth, which it finally had reached around 4,540 million years ago? Once a planetary body assembled in the way described above reaches a certain size, it starts to melt and gravitational differentiation between metallic and silicate materials takes place. This is how proto-Earth would have started to produce its layered structure. Three potential heat sources could have contributed to this melting of proto-Earth. We already considered the higher overall temperatures in the inner region of the forming solar system which likely resulted in temperatures in the range of somewhere between 800 °C to 1400 °C within Earth's orbit.

Collisions are another source of heat. Initially, there were smaller differences in the velocities of objects orbiting the Sun in the accretion disk and most collisions likely happened at small velocity differences. As planetesimals grew and collisions became less frequent, the Sun's gravitational pull would gradually increase these small differences in velocities. Conversion of kinetic energy into heat energy in higher velocity collisions certainly would have contributed to increasing Earth's temperature. However, that by itself would not have been sufficient to produce a molten Earth. Scientists now believe that the major culprit responsible for the heating that enabled gravitational differentiation between metallic and silicate materials was heating of the Earth from the inside out. The energy for this process came from the heat Earth had retained from its accretion process and from the decay of radioactive elements heating up Earth's rocky material to the point of melting. This radioactive heating, added to the retained heat Earth already possessed, was sufficient to melt Earth from the inside out. Thereby setting in motion the layer-differentiation of proto-Earth as the heavier elements sank down towards Earth's core while the lighter ones remained close to the surface starting to form a thin crust. This is how Earth acquired its layer structure with a metal core at its center and so did other rocky planets that had managed to grow sufficiently large.

Scientists have been trying to understand the formation of the Moon for a long time. Before humans set foot on the Moon all explanations were more or less educated speculation. The Apollo missions brought back Moon rocks that enabled us to understand its geology and chemical composition. As a result, it became clear that proto-Earth must have acquired its Moon through a gigantic collision with another planet believed to have been about the size of Mars. This former planet, now referred to as Theia, was in a planetary orbit close to proto-Earth.[41] Theia likely had been cleaning up its orbit in the same way proto-Earth did. However, once a gravitational nudge by another planet moved Theia and Earth into the same orbit, a collision became inevitable as according to physics no two planets can share the same orbit. It looks like the planet Jupiter largely determined the size to which the inner rocky planets of our solar system could grow. In the early solar system Jupiter was drawn towards the Sun before being pulled back by its gravitational interaction with Saturn. The latter happened only after Jupiter had come close to Mars, likely stunting its growth but maybe also the growth of the other inner planets by diverting and gobbling up material from the accretion disk. Maybe Jupiter had also a role in the collision of Theia and Earth; but we do not know that.

Proto-Earth was the larger planet and absorbed most of the mass of Theia when the two planets collided in a glancing impact. If the collision would have been head-on, maybe proto-Earth could have fragmented with the mass of proto-Earth and Theia divided among a number of smaller planets, none of them capable to support the planetary evolution that later enabled life on Earth. However, this is not what happened. In the glancing collision between Theia and proto-Earth, the latter absorbed most of Theia while a large but much smaller mass than Theia was ejected into space. As a result, Earth's mass increased to about 90% of what it is today and our planet acquired a Moon much smaller than Theia. This is because only about half of the material ejected into space in this planetary collision, mostly part of proto-Earth's mantle, eventually aggregated through gravitation to form the Moon. Any remaining ejected debris from this collision that did not help to form the Moon fell back to Earth or drifted off into space.

This collision between the two planets creating the Earth-Moon system took place about fifty million years after the solar system formed, when proto-Earth was a few ten million years old. It looks as if it took a surprisingly short time, less than one hundred years, which is nothing on a geological time scale, for the Moon to coalesce from the debris circling Earth after the collision. Today, the Earth-Moon distance is about 384,000 kilometers. However, when the Moon formed it was much closer to Earth. Estimates of the distance at which it formed are bounded by the maximum distance the debris from the collision could have been ejected into space and by how close to Earth the Moon could have formed without being pulled apart again by Earth's tidal forces. Because of those constraints, the Moon is believed to have formed at a distance from the Earth surface of about four times the Earth's radius, or roughly 24,000 kilometers.

At this distance, the Moon would have loomed over Earth with roughly sixteen times the size of the Sun's disk. Ever since its formation, the Moon has moved away from Earth and it is not finished doing so. With the help of mirrors set up by Apollo astronauts we can bounce laser light from a source on Earth off the Moon surface. By measuring the time it takes the laser light to travel to the Moon and back to Earth we can precisely know the Moon's distance from Earth. That distance has increased by an average of 3.82 centimeters every year since the first measurements were taken in the early 1970's. The moving away of the Moon also had other consequences and these came about because of a fundamental law of physics, known as the law of conservation of angular momentum. Most of us have seen it in action or experienced it ourselves in some way. Like for example when someone sitting on a rotating chair with barbells in her or his hands can change their rotational speed by moving the weights closer to their bodies (speeding up) or farther away from their bodies (slowing down). Another example of course are ice skaters when they do their pirouettes. Earth and Moon do something similar. Because the Moon rotates very slowly around its own axis, the angular momentum in the Earth-Moon systems is mostly contained in only two degrees of freedom, the rotation of the Moon around Earth and Earth's rotation around its own axis.[42] Today, Earth rotates around its own axis every twenty-four hours. However, this was not always so.

The glancing collision between Theia and Earth likely contributed substantially to the combined angular momentum of the resulting Earth-Moon system; it must have converted some of Theia's kinetic energy into rotational energy. We do not know by how much that changed Earth's rotational speed, but we do know that at the time the Moon formed, Earth was rotating much faster and a day was only a few hours long. Over time, this compared to today much larger angular momentum of Earth's rotation around its own axis was slowly transferred to the Moon's rotation around Earth. The result was a slowing down of Earth's rotation while the Moon moved ever farther away from Earth. Figuratively speaking, Earth is the ice skater doing the pirouette and the Moon is the equivalent of a barbell tied to Earth by the flexible hands of gravity. This process has left us with a Moon now at a 384,000-kilometer distance from Earth and a twenty-four-hour Earth day.

The Moon continues to move away but it looks like it does so with a decreasing speed. Assuming for the sake of calculation that the Moon formed 4,500 million years ago and multiplying this number by the average annual 3.82 centimeters observed for the Earth-Moon distance increase since the 1970's, one would expect the Moon to have moved away only about 172,000 kilometers and not the roughly 360,000 kilometers it actually has receded from Earth since then. Therefore, the rate at which the Moon moves away from Earth must have been different in the past and on average must have been higher than it is today.

Shorter Earth days in the distant past are a logical consequence of how the Moon formed. As Earth's new satellite started to move away from it, the laws of physics ensured through the conservation of angular momentum of the combined Earth-Moon system that Earth's rotation would slow down. We currently cannot say exactly if an Earth day at the time of the Moon's formation was three, four, or five hours; but we do know it was much shorter than today.[43] There are fossil records that provide clear evidence that Earth had shorter days in the past. Comparisons of growth lines of fossil corrals that grew 350 million years ago with modern corals show that back then an Earth year had around 385 days. Hence, some twenty million years before the supercontinent Pangea starts to form, an Earth day was about twenty-three hours long. And from tidal cycles recorded in sedimentary rocks in places such as southern Australia 620 million years ago and in Utah 900 million years ago, scientists calculated a day's length back then to be 21.9 and 18.9 hours, respectively. So, the trend is clear, the farther back we look in time the shorter Earth days were, the faster Earth rotated, and the closer the Moon was to Earth.

Scientists believe the collision with Theia had another lasting impact on Earth that would help shape the evolution of life on our planet. Earth's rotational axis is not perpendicular to its orbital plane around the Sun but tilted twenty-three degrees away from the perpendicular position. A collision as massive as that between Theia and proto-Earth certainly can be expected to change the orientation of Earth's rotational axis from whatever it was before the collision. Earth is not the only planetary body in our solar system with a large tilt of its rotational axis, so that by itself is not remarkable. However, independent of Theia's contribution to increasing Earth's axis tilt, the formation of the Moon helped stabilize it. Earth's axis in the combined Earth-Moon system is less susceptible to disturbances due to the torque exerted by the Sun on Earth's equatorial bulge. For this it is also helpful that in relation to its planet our Earth trabant is by far larger than is the case for any other planet with moons in our solar system. This stability must have mattered for the evolution of life on Earth. An unstable rotational axis producing frequent and unpredictable changes of climate patterns across Earth would likely not have been a positive for the evolution of life on our planet.

When did Earth's crust solidify? As we have just discussed, the Moon formed around 4,500 million years ago, or some fifty million years into Earth history, give and take ten million years on either side. The oldest rock we know of today dates back to 4,280 million years ago, although there is still some dispute over its dating, and it may be somewhat younger. There seem to be multiple indications that water was present as early as 4,400 million years ago. It may well be that as soon as 150 to 200 million years after the formation of Earth, our planet already had landmasses and oceans. When Earth formed it had no atmosphere, leaving its surface directly exposed to the chilling cold of outer space. Because of that, molten rock on Earth's surface would not have staid molten for long and Earth may have had an initial crust of rock not long after it had recovered from the impact with Theia. Scientists believe that this first crust was made of peridotite rock which as we have seen (tab. 2.1), is the dominant rock material in the upper mantle.

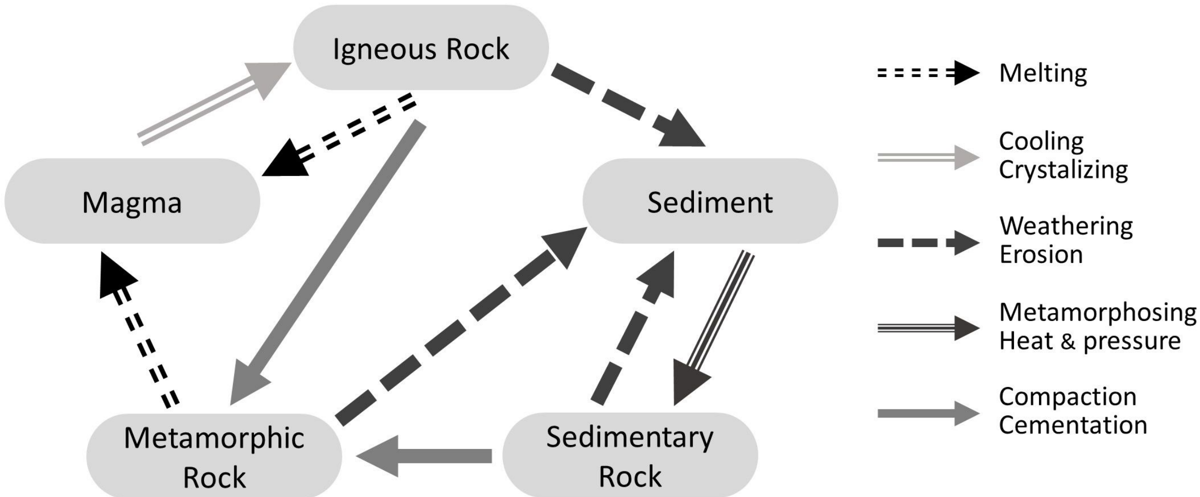


Figure 2.3: A simplified schematic of Earth's rock cycle.

However, for a solid crust to become more permanent it needs to be of a lighter material than the rocks below it; cooled and solidified peridotite on molten peridotite rock just does not provide much in terms of buoyancy. Because of that, this early crust did not last long, and it certainly did not allow for magma spewing volcanoes to rise much above Earth's surface. This is where the process of density differentiation comes into play. Illustrated in fig. 2.3, scientists call it Earth's rock cycle, it would eventually produce a more permanent floating crust of lower density.

Geologists differentiate between three main types of rock, igneous or magmatic rocks, metamorphic rocks, and sedimentary rocks. Igneous rock forms through the cooling and solidification of lava transported to Earth's surface from magma chambers in Earth's mantle; that is why it is also called magmatic rock. Metamorphic rock is, as the name indicates, a reformation of rock that sinks from the surface into the depth of Earth and undergoes a metamorphosis under the increased pressure and temperature including partial or complete melting depending on how deep a rock sinks. The third type of rock only enters the stage once we have weathering and erosion at work, which later on the evolution of life will help greatly accelerate. This sedimentary rock forms through the deposition and subsequent compaction of mineral and organic materials. In Earth's rock cycle, rock material is continuously being reworked and recycled in Earth's mantle under increasing temperature and pressure to return in a new form to Earth's surface. Density differentiation of rock material is an integral part of this process. As the descending rocks partially melt while sinking, the lighter materials in the melting rocks separate and rise to the surface again, now as magma with a somewhat lower density.[44] With each repeating cycle, the magma that spewed out of volcanoes or found its way up to the surface through fissures was lighter than the original peridotite of the first crust. This is how Earth acquired its first basalt crust.

Basalt is about 10% less dense than peridotite (see tab. 2.1) and therefore this new crust was a thicker and much more stable rock float than the thinner and short-lived peridotite crust. Its increased buoyancy provided for a more interesting Earth topography with volcanoes built of basalt now able to rise higher above the surface. Scientists believe that Earth's basalt crust was already in place 4,400 million years ago, roughly one hundred million years after the Moon had formed. The Moon continued to be a major force testing the solidity of Earth's crust. Its tidal forces tearing on Earth's thin basalt crust were much stronger than they are today because the Moon was so much closer. It is therefore unlikely that the first basalt crust was continuous like an eggshell enveloping all of Earth. More likely, the first basalt crust resembled a cracked eggshell with molten rocks continuing to seep through the cracks and large volcanoes spewing out obnoxious gas mixtures along with the yellow-white magma at their core feeding red lava streams covering vast surface areas.

The time Earth was covered by this shiny crust of basalt sometimes is referred to as the *Black Earth* period and it would not last for long. Early Earth was going through millions of years of large-scale volcanic activity churning out enormous amounts of material including much of the volatile compounds still part of early Earth's interior. The resulting atmosphere was very different than what we enjoy today, more likely a thick soup of noxious gases such as sulfuric gases, carbon dioxide, nitric oxides, and maybe methane, but most interesting for us, it also included water. To this day, there is a continued and sometimes contentious debate when and how Earth acquired its water and its oceans. What we know is that Earth must originally have had a much larger fraction of volatile gases than it has today. This follows simply by inference from what we learn from the composition of chondrite meteors, which we know are the type of material which initially formed Earth. Whatever water proto-Earth held, it likely lost a large fraction of its volatile components when Theia ripped through Earth's mantle in the glancing impact that was to give Earth its Moon.

There are two ways, or a combination of both, by which Earth could have obtained its surface water. First, there is the water still buried deep inside Earth's interior. As we have seen, the rocks of the mantle undergo phase transformations as pressure and temperature increase that enables them to store water in the form of hydroxyl in their crystal lattice. The amount of water stored not just in Earth's transition zone but also in other layers, even including hydrogen in Earth's core, could be many times more than all water contained in today's Earth oceans. How could the water trapped in Earth's interior have made it to the surface? Well, maybe not much different from how it does so today. Volcanic eruptions are accompanied by huge amounts of water vapor being released into the atmosphere. If Earth had undergone a prolonged period of massive volcanic activity early on, then the water vapor put into the atmosphere in this way could have condensed and rained down on Earth's surface. Those rainfalls must have lasted for a very long time to fill Earth's early shallow oceans. A problem with this explanation is that early volcanic activity may have been different from what we observe today given that volcanoes on a thin crust could not have risen much above Earth's surface. Maybe, it was more like hot water vapors oozing out through the many cracks in Earth's early crust. That certainly would better fit with Earth acquiring oceans much earlier than previously thought possible, maybe only 150 to 200 million years into Earth history.

This mechanism of bringing water from Earth's interior to its surface is still at work today, even though there is quite a bit less volcanic activity than there was in these early Earth days. In addition, since plate tectonics started, water re-enters Earth's interior in subduction zones as tectonic plates dive into the upper mantle in conveyor belt fashion. Therefore, there must be a balance of water being moved from Earth's interior to the surface via volcanic activity and water being transported into Earth's interior through subduction, but we do not know what this balance looks like. Does the addition of new water from Earth's interior to the surface outweigh the loss of water to Earth's interior in subduction zones or the other way round, or are the water volumes similar? For now, we do not have the answer. Making things more difficult, any such answer would also have to account for other sources or drains of water such as when early in its history, before it had an oxygen atmosphere, Earth lost a large amount of water to space.

If the water that we now find in Earth's oceans has always been there from the beginning, just not on the surface but trapped in Earth's interior, this also raises other interesting question. High amounts of water trapped in planetary bodies could be a standard feature of planets forming in the way they did in our solar system. If that is so, there may still be enormous amounts of water trapped in a planet like Mars. It just cannot make it to the surface anymore as the Martian engine that could drive plate tectonics and with that continued volcanic activity, has stuttered sometime in its distant past. We still know very little about the geology of Mars but it seems to have been geologically speaking dormant for a long time. Once we know more about it maybe we can understand what tectonic plate activity looked like on Mars and why it seems to have stopped. Maybe there is a critical mass a planet must have to maintain the conditions necessary for plate tectonics to work over many thousands of millions of years, as is the case for Earth.

A second way Earth could have acquired its ocean waters is through asteroids or meteorites. Miniscule amounts of water have been found in meteorites and it would have taken countless billions of them to bring the amount of water to Earth that we find in today's oceans. However, asteroids can contain larger amounts of water. Throughout its history, Earth has been subject to asteroid and meteorite bombardments alike, but much more so in its early history than today. Specifically so between 4,100 – 3,800 million years ago in what is called the Late Heavy Bombardment or LHB, when Earth was not only subject to meteorite bombardment but also hit by very large asteroids with tens of kilometers in diameter. After this roughly 300 million year-long LHB episode, during which Earth grew to its current mass and size, scientists believe the frequency of meteorite or asteroid impacts on Earth not to have been much different from what it is today. Therefore, any substantial amounts of water delivered through meteorite or asteroid rocks must have arrived before 3,800 million years ago. How much meteorite and asteroid impact could have contributed to today's water on Earth we do not know. However, while the late heavy bombardment may have added water to Earth it cannot explain how Earth could have had oceans only 150 to 200 million years after it formed.

Sometimes it is the little things that help us understand something much grander. Zirconium silicate crystals ($ZrSiO_4$), often just called zircon crystals, represent a good example for one such little thing as they have helped us understand when Earth may have had first liquid water. Zircons are like time capsules as they are extremely durable. Importantly, zircons can have as one of their metal inclusions the most common uranium isotope (U-238), which radioactively decays into lead with a half-life of about 4,500 million years. Therefore, if scientists find as many lead atoms as U-238 atoms in such a zircon they know that the crystal is around 4,500 million years old.

A second characteristic makes those zircons even more valuable. They can provide a historical record of the ratio of the regular oxygen isotope (O-16) to its heavier isotope brother (O-18) which has two more neutrons. Why is this significant? Well, early Earth had no oxygen atmosphere and because of that there is only one way a zircon predating the arrival of an oxygen atmosphere on Earth could have acquired a high relative amount of the heavy oxygen isotope O-18: it must have formed from the melt of an older rock that had been exposed on Earth's surface to a wet environment, to water. In addition, the variation in the relative abundance of O-18 to O-16 is an indicator of the temperature at which the respective rock was exposed to a wet environment. As we will see later, this ratio is also used in extracting Earth's temperature records from cores drilled from the deep seafloor or from polar ice sheets.

A challenge scientists face with extracting such data from zircons is that not just any zircon will do. Evidently, zircons can grow and add new layers on the outside, which would result in mixing the oxygen isotope records of younger ages with that of the older core of the zircon. Because of that, zircons that have not had the chance to grow further layers because they were included inside rock formations are easier to analyze and understandably much preferred. The oldest zircons discovered so far date back to 4,400 million years ago.

Even though there is still debate about the specific age range the oxygen isotope records of some of the analyzed zircons represent, there seems to be consensus that Earth had shallow seas as early as some 150 to 200 million years after it had formed. This is quite a change from the much more hostile environment previously presumed prevalent for this early time in Earth history, aptly named the Hadean eon. Earth's interior certainly was quite a bit hotter than it is today but with liquid water on its surface only a couple hundred million years after Earth was formed. Therefore, whatever conditions may have been on early Earth otherwise, they must have included surface temperatures suitable for this to happen.

Liquid water on Earth's surface can only exists within a narrow temperature range we are all familiar with, above its freezing point and below its boiling point, covering the range of 0 to 100 degrees on the Celsius scale. Of course, here we are also assuming that the atmospheric pressure was the same as it is today. If it would have been lower, then water would have boiled at temperatures below 100 degrees and it would have been liquid at temperatures above 100 degrees if the pressure had been higher. Unless the atmospheric pressure back then differed dramatically from what it is today, the temperature range over which water could remain liquid would have been very similar. However, when Earth formed, overall conditions in our solar system were quite different in other respects and that needs some explaining. As we saw in the previous chapter, the Sun's luminosity back then was only about 70% of what it is today. Therefore, surface temperatures on Earth should have been quite a bit less, unless there was something else keeping Earth's surface temperature in a range that would allow for liquid water to slush around. That other heat source must likely have been the much hotter interior that Earth in the Hadean eon certainly had. Maybe an added heat source was provided by an atmosphere rich in methane and carbon dioxide acting as a thermal blanket. However, this hypothesis is increasingly in question, as scientists understand how much atmospheric hydrogen early Earth has lost to space.

For at least 2.7 to 3.2 billion years anaerobic microbes aptly named methanogens have been relying on a metabolism that uses carbon dioxide and water and has methane as its end product. In this process scientists call methanogenesis, they essentially end up splitting water and thereby release hydrogen as well as oxygen. Because hydrogen is significantly lighter than water or oxygen, it rises through the atmosphere where it either can react with oxygen to return as water to Earth or is eventually lost to space if there is little to no oxygen in the atmosphere to start with. Because Earth's early atmosphere had no oxygen, Earth's oceans seem to have lost about a quarter of their original volume in this way. By knowing how much hydrogen was lost through this process to space, scientists can also infer how much methane existed in Earth's early atmosphere. While the amount of methane predicted by this analysis is many times more than we find in today's Earth atmosphere, it is not enough to account for the additional warming effect such a thermal blanket of methane would have had to produce.

A potentially greater amount of carbon dioxide in Earth's early atmosphere was also not sufficient to account for the missing additional heating that Earth must have had. Methane is a much stronger greenhouse gas than carbon dioxide so it is unlikely that carbon dioxide by itself may have been sufficient to close the gap. For some time, it looked like the atmosphere of early Earth was denser than it is today which would have amplified any greenhouse effects, but findings that are more recent indicate that the contrary may have been the case; early Earth's atmosphere seems to have been lighter. While greenhouse gases may have helped to warm Earth, it seems not even the combination of greenhouse gases with a hotter Earth interior was sufficient to help warm early Earth enough to compensate for the lower luminosity Sun.

The most probable current hypothesis is that the much larger amount of surface water itself may have been critical in keeping the planet warm. Continents reflect more sunlight than ocean surfaces and therefore larger ocean surface areas result in a warmer planet. Earth's early oceans were much shallower than our oceans today, but they covered a much larger surface area than the roughly 70% of Earth's surfaces oceans cover today. If large amounts of surface water were indeed required to ensure Earth's surface temperatures would be in the right range to support liquid water on its surface in the first place, then this begs the question as to how it all got started. Therefore, it looks like the so-called young Sun paradox of how early Earth could have had vast shallow oceans even though there was much less heat coming from the Sun is not quite solved yet. Nevertheless, it is amazing to find a number of factors including the early solar luminosity, the Earth system with its interior heat source, an atmosphere, and a surface water to land area ratio, all balanced just right to result in the conditions needed to support liquid water on Earth. Much of Earth history and the evolution of life on our planet over the past 4,000 million years would be a delicate dance around this detailed balance.

The picture that emerges towards the end of the Hadean eon is of an early Earth with a basalt crust and oceans of liquid water. Basalt is still the material that Earth's crust is made off today beneath the sediments of our oceans sea floors; it is just basalt that has been recycled many times over. However, to build mountains and large continents required a still lighter, less dense material. As the earlier discussion of Earth's layer structure showed, to build up mountains a material with higher buoyance is needed and that material is granite. Granite is about 10% less dense than basalt and all continental crust on our planet is made out of granite materials. Similar to the rationale provided to explain the emergence of the basalt crust, granite is a product of Earth's rock cycle where basalt crust is being reworked and recycled to emerge with a lower silica content. However, the way granite building initially got started must have been different from how it works today where basalt crust is subducted into the mantle and partially remolten for some of it to reemerge as granite. Back then, there were no tectonic plates and no plate tectonics to recycle Earth's rocks.

So how then was the first granite rock produced? The process at work that scientists consider most likely was similar to how peridotite had been reworked into basalt earlier, only that now the lower part of the basalt crust was partially remolten and reworked into granite. How granite started to grow more in certain areas than in others and started to form small granite islands is still debated. It may be that continued large asteroid impact is responsible, as it could have confined magma plumes driving granite growth to the areas underneath the unbroken basalt islands that were not shattered and turned into seas of molten magma. The granite building process could have started as early as 200 million years after Earth formed and there are even some indications that continental crust building may have started as early as 4,400 million years ago. It took Earth granite making factory several hundred million years to produce large granite islands which by 500 million years into Earth history, towards the end of the Hadean eon, were covering a small part of Earth's surface. These granite islands merged to become the cratons from which our continents then formed as they continued to grow in a process geologists call continental accretion.

All of today's continents are built around cratons of which there are roughly three dozens.[45] The first ones formed as early as 3,800 million years ago and they come in various sizes with several as large as continents and other more like large islands. Cratons just like continents have names, some of them ringing more familiar to us than others, but most of us have never heard of the majority of them. In a few areas, cratons are exposed to the surface and where they are, they are known as shields, such as the Canadian Shield. It is no surprise that some of the oldest rock has been found in such places. By the end of the Archean eon, about 2,500 million years ago, the continental crust area had grown from a few small granite islands to maybe as much as 70% of today's continental area. Much of this late Archean continental crust material is still part of our modern continents but only a small percentage of it is anywhere close to Earth's surface or exposed on top of it. Throughout the subsequent Proterozoic eon, continental masses were reworked as new land was added through continental accretion and at the beginning of the Phanerozoic eon 541 million years ago most of the continental mass that Earth has today was in place.

During our lifetimes, the continental map remains seemingly static as plate tectonics moves continents at maximum only a few centimeters per year; North America and Europe are currently drifting apart by roughly three centimeters each year. Over a human lifespan this does not amount to much but over many millions of years it adds up and reshapes Earth's surface map. It is now thought that the heat convection processes in the mantle that drive plate tectonics were in place by around 3,200 million years ago, putting the ignition of Earth's plate tectonic engine right around the beginning of the Mesoarchean period. The majority of geologists seems to agree on this but there is also some dissent, placing the start of plate tectonics much earlier, around 4,200 billion years ago, or much later, only 1,000 million years ago. We will stay here with the majority consensus of roughly 3,200 million years. This is almost 1,000 million years after Earth's rock cycle produced the first granite rocks and about 800 million years after granite islands appeared on Earth's surface. Plate tectonics is choreographing this dance of the cratons, because that is exactly what it is, into ever-different arrangements. Movements in this dance are constrained by the forces that drive it, by the arrangement of cratons into continents and with respect to each other as well as by the restrictions imposed on this dance by the fact that it takes place on a spherical surface and not on a flat stage.

Geologists have identified three forces that drive plate tectonics. First, there is heat convection, the mantle convection currents that drive and carry plates of lithosphere in conveyor-belt-like fashion as discussed earlier. Second, there is the process scientists call ridge pull. Newly created crust at ocean ridges has a higher elevation and its gravity pushes the neighboring lithospheric plate materials on either side of the ridge away from it. The third force is what scientists call slab pull, the process where in subduction zones older and colder oceanic plates sink into the mantle. Driving this slab pull subduction of oceanic plates are temperature differences between the descending slab and the lithospheric mantle the slab dives into. Initially, the higher density of the colder slab of the subducting oceanic crust drives the process because its density is higher than the surrounding mantle material. Then, as it sinks, the basalt rock in the slab of subducting oceanic crust converts into a higher density metamorphic rock dragging the slab further down into the mantle.

For a long time, it looked as if heat convection of the mantle was the most important process driving the plate tectonic movements that we see today. However, more recently, it has become clear that the forces exerted by the slab pull itself are dominating what drives plate tectonics; and therein lies the problem. With slab pull being the most important force powering plate tectonics how did plate subduction get started in the first place? Seemingly, geologists face a chicken and egg problem here. Much depends on what assumptions we make about what Earth's early crust really looked like. If Earth's early lithosphere was anything as thick as it is today and if it was practically a continuous unbroken shell it is hard to see how any part of Earth crust could break and how a slab of its crust could force its way into the mantle's asthenosphere. In one way, it would have helped if Earth early mantle was warmer as we know it was because that would have made it easier for plates to brake and sink into the mantle. However, therein also lies a problem as the slab pull only works if the oceanic plate subducted into the mantle does not break off. The conveyor belt moving Earth's plates only works if the belt is not broken. Subduction must bring in new parts of a plate's slab for the process of subduction to continue, and this does not work if the slab breaks.

While we do not know exactly yet how Earth plate tectonics engine started around 3,200 million years ago we do know that over the next some 700 million years it would produce as much as 70% of the continental crust that is still in place today; that is the high estimate. Lower estimates bracket the fraction of Earth's continental crust in place by about 2,500 years ago at no less than around 50% of today's continental crust. From what scientists understand, the process of plate tectonics may have evolved over time as it seemingly has changed around 2,500 million years ago. Continental crust was also not added in a continuous process since plate tectonics started, but rather there seem to have been more and less productive continental crust production phases; the last such productive phase occurring as "recently" as some 500 to 300 million years ago.

Since plate tectonics began to choreograph the dance of Earth's tectonic plates, several times throughout Earth history this has led to most or even all of its landmass coming together to form supercontinents. Geoscientists believe this has happened six or seven times so far and it is expected to happen again in about another 250 million years. Fig. 2.4 shows the continental maps of Earth for 800, 450, 230, and 80 million years ago, as well as the present continental map and the continental arrangement projected to be in place in about 250 million years from today. The first three of the maps showing continental configurations in geological time coincide with supercontinents being in place, namely Rodinia, Gondwana, and Pangea; the fourth map of the past reproduces the continental arrangement in place about 80 million years ago when dinosaurs still ruled the world. Continental maps of the past are calculated by running plate tectonics, the mechanism that is behind the movement of the continents, in reverse. The mechanisms of plate tectonics is today quite well understood and most such calculation show similar results. Computers can certainly run plate tectonics also in forward mode, allowing predictions of when a new supercontinent may form.

The maps shown in fig. 2.4, except for the map showing the present-day configuration, can of course not be completely accurate. First, when looking far back into the past or looking forward into the future for the same number of years, we can trust simulations of past continental configurations more than future projections. This is because we have a geological record that helps to test and verify our models of past continental maps based on plate tectonics. Of course, for the future we do not have this information. Also, the further we look back in time or the further we try to project into the future the more prone to errors these calculations become.

In addition to the actual movements of continents, sea levels also change over time and expose more or less of the continental shelves resulting in larger or smaller landmasses. This can result in landmasses staying connected or becoming separated by bodies of water; something to keep in mind when comparing continent maps as shown in fig. 2.4, which mostly trace only the movements of continental masses, as we know them today. Those maps will look somewhat different if sea levels were higher or lower but not so different that we could not clearly see how much the continental map has changed throughout geological time. Continents will continue to wander across Earth's surface through geological time as long as Earth's plate tectonics engine keeps running. From what we know today, long before its convection engine will show any signs of fatigue our planet Earth will have become inhabitable for humans, a topic we will return to in the last chapter.

The first supercontinent may have possibly formed as early as 3,600 - 3,300 million years ago during the Paleoarchean era. It is referred to as Vaalbara but scientists are mostly skeptical about this one.[46] However, there is consensus that a supercontinent likely formed around 3,100 million years ago, just shortly after plate tectonics may have started, in the early Mesoarchean era and it has been given the name Ur.[47] Because Earth's total landmass at the time was still significantly less than it is today, Ur is likely to have been smaller than the modern continent of Australia.

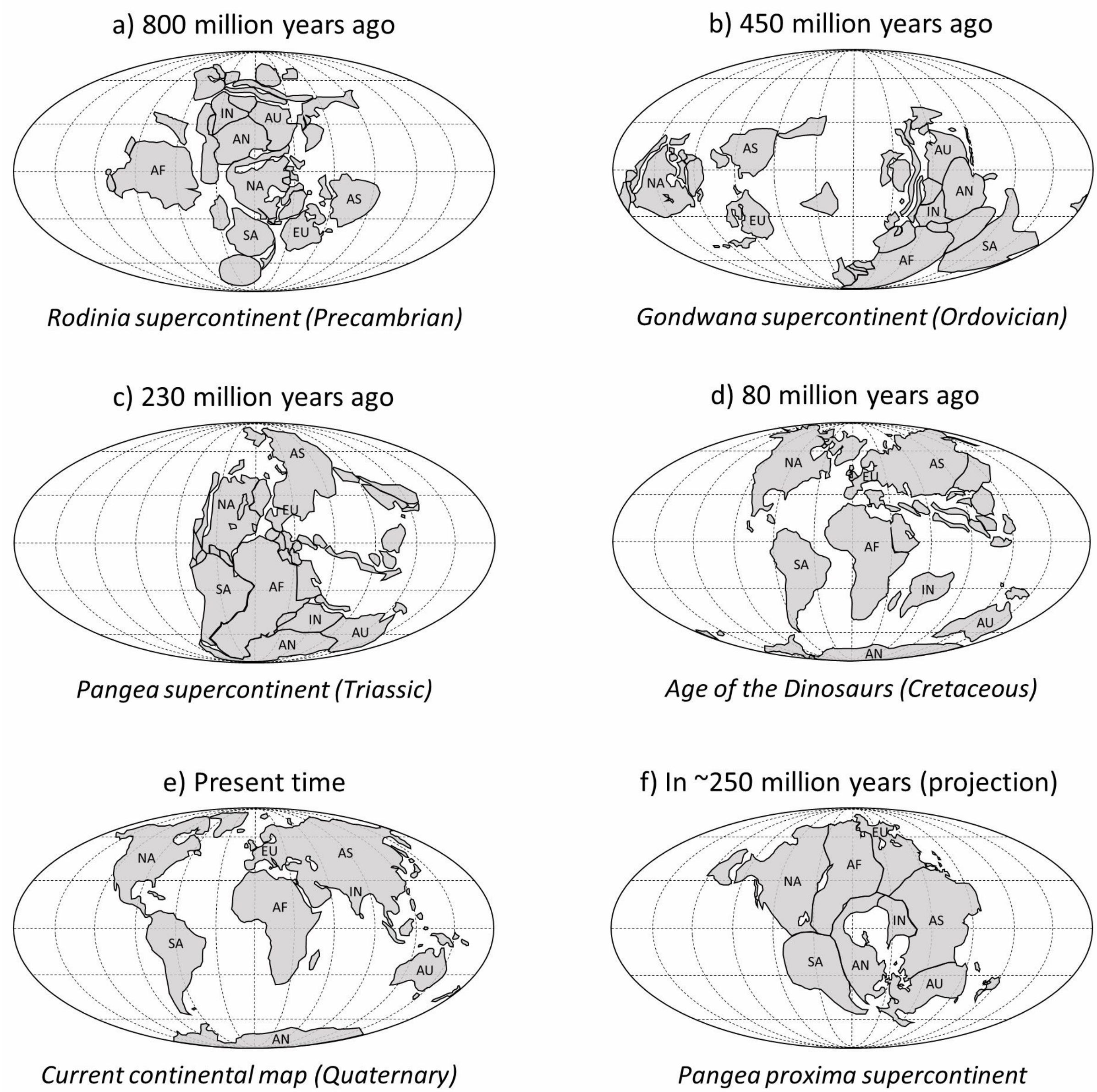


Figure 2.4: Earth's continental map as it looked like at four different times in the past (a-d), its present configuration (e), and a projected configuration (f) of the next supercontinent expected to have formed by 250 million years from now. The labels trace movement of continents / continent kernels: Antarctica (AN), Asia (AS), Australia (AU), Europe (EU), India (IN), North America (NA), and South America (SA). Maps (a)-(b) modified and adapted from Ref. [12] with (f) sketching a combination of several potential projections.

The supercontinent Ur is believed to have fragmented by 2,800 million years ago, towards the end of the Mesoarchean era, so it may have been around for only 300 million years or less. The next supercontinent, referred to as Kenorland or Superia, is thought to have formed around 2,700 million years ago in the early Neoarchean era and it is believed to have fragmented by 2,100 million years ago, 400 million years after the Archean eon ended.[48] Then comes the Columbia supercontinent that existed sometime between 1,900 and 1,600 million years ago. These first supercontinents were formed far back in geological time so there is quite some uncertainty as to what they may have looked like or the configurations of the landmasses in the respective supercontinent arrangements. This is different for the next supercontinent, Rodinia, which is thought to have formed about 500 million years after the break-up of Columbia, around 1,100 million years ago. Rodina is believed to have started fragmenting around 850 million years ago and to have been broken up by 750 million years ago. Fig. 2.4a shows a snapshot of Rodina's likely shape about fifty million years before it started to fragment.

For a short time, at least on a geological time scale, another supercontinent called Pannotia assembled from the fragments of Rodinia starting around 650 million years ago. It is thought to have been in place between 620 and 580 million years ago and to have fragmented again by 560 million years ago. In the late Precambrian the supercontinent Gondwana formed. It was a very large continent but unlike Rodina it did not include most landmasses as much of what would eventually become North America or Europe, the Laurentia craton and the Baltica craton, respectively, as well as a chunk of Asia, the Siberia craton, were not part of it. Fig. 2.4b shows Gondwana 450 million years ago, which is a little less than 100 million years after it likely started to form and a little more than 200 million years before it would become part of the next supercontinent Pangea. Pangea, the most recent supercontinent, took shape around 330 million years ago and joined together much of the southern supercontinent Gondwana with the northern supercontinent Laurasia which had formed out of the Laurentia, Baltica, and Siberia cratons, with the North China and South China cratons joining as they moved away from Gondwana into the northern hemisphere. In fig. 2.4c we see what Pangea looked like 230 million years ago, or roughly about 100 million years after it formed and about fifty-five million years before it started fragmenting some 175 million years ago. As shown in fig. 2.4d, by 80 million years ago, at a time when dinosaurs still thrived, the Atlantic Ocean had already opened. It would still have to open much more to reach its current size, but the continental map starts to look familiar. Major changes of the continental arrangement were still underway. India, after detaching from Africa and Madagascar, still had a long way north towards its current position. Much of what is today South- and Southeast Asia was still being assembled at the time; Australia had detached from Antarctica but still had a long journey ahead towards getting closer to the equator; the Anatolian plate was still to move into its current position; and finally, North- and South America had not joined yet by a land bridge.

A comparison of fig. 2.4e, the current arrangement of landmasses, with fig. 2.4f, a likely projection of what the next supercontinent will look like, gives a good idea of the dramatic changes ahead over the next 250 million years. Such enormous changes over geological time seem almost incomprehensible to us given our short lifespans and the brief history of our human species on planet Earth. However, these changes are real, and we can measure them. As pointed out above, each year North America and Europe drift apart by about three centimeters. This accumulates to a little more than two meters in a typical human lifespan; about three hundred meters since humans started agriculture; six to nine kilometers since our first modern human ancestor appeared; or some two hundred kilometers since human and chimpanzee lines split from a common ancestor. The message is clear, continental drift matters for species only if they are around for a long time. There are some species that lasted for a much longer time than humans are around, dinosaurs are one such example. During their more than one hundred-million-year reign, we are talking continental movements over large distances, on the order of three thousand kilometers if we assume a three-centimeter annual drift over that time. Other species have been around even longer and must have apparently succeeded in adapting to the changing climate conditions that typically accompany such large shifts in continental masses; a feat that humankind is still very far from accomplishing. As it looks now, human adaptability to changing climates will not have to wait to be tested by climate changes initiated by continental movements over geological time. We have managed to do something that only the planet itself could do in the past, changing Earth's climate, and we have been able to do it in what amounts to less than a blink of an eye on a geological time scale. But more about this later, for now, let us go back to our planet's story before we come to ours.

Since plate tectonics started, the material carried on top of the tectonic plates has been reworked many times and we likely will never find granite material that dates back to the first granite rocks being produced maybe only 200 million years after Earth formed. The oldest granite rocks we know today are dated to around 4,000 million years ago, forged roughly 500 million years after Earth formed. Maybe we will eventually discover rock material that once constituted granite islands floating on Earth's surface 200 million years into its history; or maybe not. But there is at least a chance that one day we will find such granite material.

At the end of the Hadean eon, Earth had granite islands interspersed in a sea of basalt crust forming ocean floors that were still shallow but started to get deeper. Earth's surface water back then, the primordial oceans, covered a larger fraction of Earth's surface area than the roughly 70% they cover today. In those early days of primordial oceans, tides must have been a lot higher than what we can experience today. At its current distance from Earth, our Moon accounts for about half of the forces producing our tides. We can tell that quite easily, because during full or new moons, which is when Earth, Sun, and Moon are near alignment, our tidal swings are larger. We refer to these twice-a-month occurring higher tidal swings as spring tides. If the Moon produces roughly half of our tidal swings today, we can imagine how much more powerful Earth's tides would have been back in early Earth history when the Moon was much closer to Earth. When the Moon formed it was about sixteen times closer to Earth than it is today which would mean its gravity pull on Earth would have been some two hundred and fifty times stronger than today. The Moon certainly had removed somewhat from Earth by the time Earth had its first oceans but the pull of Moon's gravity on Earth's water was likely still by a factor of one hundred stronger than it is today. Earth oceans must have seen gigantic tides, somewhat moderated by shallower seas, sweeping around our planet every few hours as it was rotating much faster back then with Earth days as short as three to five hours. It would take a long time before Earth tides would become tamer and settle into a magnitude range and periodicity that would help life to prosper.

Earth must have been a very different environment back then. Our Moon's gravity some one-hundred times larger pulling on Earth would not only have produced enormous tidal swings, but it must also have had an impact on the stability of Earth early crust. Wherever the Moon happened to be passing over Earth's surface, it essentially made everything beneath it lighter. Given how fast Earth rotated this would have happened every few hours. Just imagine yourself on this early Earth and every time the Moon looms over you in the sky, which would be much more frequently than today, you would weigh significantly less. These gravity tidal waves do still exist today but as the Moon is much farther away, they will change your weight now by only an imperceptibly small amount. However, back then, when the Moon was so close to Earth, tidal gravity waves would have had much more dramatic consequences as they are the source of what scientists call tidal heating. Tidal heating can be a powerful source of energy as we still can see for example with several of Jupiter's moons where it sustains massive geological activity. More recently, scientists have discovered that our Moon is not completely solidified yet but deep down seemingly still has a soft layer because of tidal heating by Earth gravity. The more we learn about early Earth history the more amazing it is to see evidence for liquid water on Earth so early. Seemingly, while still inhospitable to life, Earth was nevertheless putting in place all the ingredients for life to evolve. One of these vital ingredients certainly was protection from the high-energy particle radiation coming from the Sun provided by Earth's magnetic field. We have touched on Earth's magnetic field already several times in discussing Earth geology but not really said anything more about it. Many of us likely know Earth's magnetic field from the beautiful spectacles of polar lights, the *Aurora Borealis* we also refer to as the Northern Lights and the *Aurora Australis* that we sometimes call the Southern Lights. If we have not experienced their beauty in nature with our own eyes, most of us have seen videos or images of these great nature displays, the visual manifestations of Earth's magnetic field. With today's technologies, we are also able to measure and characterize many properties of our Earth's magnetic field. However, as it turns out we still know little about the origin of our Earth's magnetic field and its history in deep geological time. Clearly, with deflecting high-energy particle radiation and stopping it from reaching Earth's surface, processes manifesting themselves in the auroras, Earth's magnetic field is a vital precondition for the evolution of life. Without it life could still have evolved in the oceans as water has a much higher stopping power for high-energy particle radiation than our atmosphere. But plant and animal life on land could not have evolved. So, for how long does Earth have its magnetic field?

Magnetic rocks practically as old as the oldest Earth rocks we have indicate that our planet must have had a magnetic field very early on, shortly after it had formed. However, the magnetic field back then was likely quite different from the one anchored by a solid iron inner core that Earth has today. Since the time Earth has a solid inner core, its geodynamo works by heat-transfer from the inner core to the adjacent outer core causing iron and nickel to rise in the outer core. Once the material reaches the D" layer separating the outer core from the lower mantle it cools and sinks back towards the inner core. These heat convection currents produce the material flows responsible for Earth's magnetic field. In principle, we should be able to determine when Earth may have acquired its solid inner core. However, we just do not know enough about how materials like iron behave under the conditions prevailing in Earth's core, which are temperatures in excess of 5,000 °C and more than a million times the pressure at sea level. Experiments seeking to determine the properties of iron under such conditions are notoriously difficult, but scientists have tried nonetheless [13,14]. The first such measurement results still deviate quite a bit from each other giving potential dates for Earth's inner core to have been solid as early as about 3,000 million years ago and as late as 565 million years ago. Before Earth had a solid inner core, it must have had different convection currents in its metal core creating a magnetic field. We can find the evidence for this in very old magnetized rocks. What these core convection currents without a solid core may have looked like we do not know, but they could for example have been weaker and they most likely were much less stable.

The oldest magnetic rocks we find on Earth do not indicate noticeably weaker Earth magnetic fields before Earth had a solid inner core. However, there are indications that without a solid inner core, Earth's magnetic field seems to have been more erratic in recent geological time, like during the Ediacaran period. The field evidence pointing to this are fluctuations in the Earth magnetic field dating back some 600 to 700 million years ago which produced a weaker and erratic Earth magnetic field record. Scientists do not attribute this to an actually reduced field strength of Earth's magnetic field but to an averaging effect preserved in the respective rocks of an Earth magnetic field fluctuating at what scientists call hyper-reversal frequencies [15]. Computer simulations using such data indicate an Earth magnetic field transitioning from weak and erratic multipolar to a more stable dipole configuration around 560 million years ago [16]. As it looks like, the evidence seems to mostly point towards a solid Earth inner core formed by the end of the Ediacaran, which from then on would anchor Earth's magnetic dipole field. Earth's magnetic field would still experience polarity reversals but much less frequent. It is not interesting that Earth's inner core seemingly solidified just before the Cambrian explosion of life? Thereby providing a stable magnetic field to protect life. We do not know if this is a coincidence or not but a stable Earth magnetic field quite likely must have been one of the conditions favoring the evolution of higher lifeforms.

Earth as we know it today would look very differently if life had not developed on it. Without life on Earth, there would be no oxygen atmosphere. Without the remnants of life long gone by and countless skeletons and shells deposited over geological time there would be no limestone, no coal or other fossil fuel deposits. From its earliest beginnings, life has been enabled and shaped by Earth's mineralogical richness and in turn, life has shaped Earth geology to such an extent that Earth has become a living planet where the existence of Earth systems such as our climate and life itself are mutually interdependent. The first major impact life started to have on Earth itself is through the photosynthesizing bacteria that we call blue-green algae. Scientists call them cyanobacteria, cyan being the blue-greenish color on an artist's palette. As a by-product of their photosynthesis process, these cyanobacteria released oxygen into their environments. Until cyanobacteria started to produce oxygen, life on Earth as it existed then was relying on sulfur- or methane-based processes to extract energy from the environment to sustain itself. By then, such lifeforms relying on sulfur or methane for their energy needs had been around for a long time. There are some indications they may go back as far as only a couple hundred million years after Earth formed. A more conservative estimate is around 4,000 million years ago, give and take a couple of hundred million years. Maybe we will eventually know exactly when life first emerged on planet Earth but as of today, we do not.

When cyanobacteria started to produce oxygen, it was toxic to likely most other life on Earth. Lifeforms had either to adapt or perish. Because the events that we recount here played out on a geological time scale many likely were able to adapt, of many others that could not we will likely never know. Some may have just moved away from the surface into the underground. Earth is still home to countless lifeforms depending on methane and sulfur; we just usually do not see them. However, not all of them are tiny anaerobic life hidden in our soils as we have learned from the recently discovered so-called black smokers, the hydrothermal vents found on deep ocean floors that sustain very different islands of life around them. In a way, remnants of ancient life may have retreated into niches some of which we are discovering only now. What at first may have looked like curiosities has opened a small window for us to see how some of life may have looked like a long time ago. We should not be surprised to find that Earth has more such secluded niches of life, places we do not know of today where time practically may have been standing still. But as Earth continually recycles ocean floors every couple hundred million years or so, what we will find down there will be more like analogies to what life could have looked like in early Earth days rather than early life itself.

It takes uncountable numbers of blue-green algae and many hundreds of millions of years until the level of oxygen in the atmosphere of a planet like Earth increases to a point where it can support larger lifeforms. When cyanobacteria first started to release oxygen into Earth's oceans it was immediately used up. This is because Earth's early oceans contained vast amounts of dissolved iron. Oxygen will readily combine with almost any other element to gain a couple more electrons and thereby oxidize them. Metals such as iron would use up all the oxygen that cyanobacteria could produce for a long time. When oxygen reacted with iron, it formed insoluble iron oxide or rust, which sank to the bottom of the sea. Iron oxide containing colloidal nanoparticles suspended in the oceans, a byproduct of this process, seemingly gave them a reddish color marking an Earth period sometimes referred to as *Red Earth*. Rusting on land later on, once free oxygen became available in the atmosphere, likely added to this reddish taint. The iron oxide sinking to the sea floor produced the sedimentary rocks called banded iron formations or BIF, alternating layers of iron oxide and silica, that make up more than sixty percent of Earth's known iron deposits. It is not clear yet why there are actually bands and not just one giant deposition of iron oxide. Seemingly, there must have been phases of higher oxygen release into the oceans alternating with phases of lower oxygen release. Picturing a living and breathing ocean is certainly a nice poetic description, but it does not provide an explanation for the banded structure of iron deposits. All of the oxygen was readily used up in the oceans as long as they contained large amounts of dissolved iron. However, once all dissolved iron was oxidized, there was only one way for oxygen to go: bubble up and start enriching the atmosphere. That happened around 2,450 million years ago, some 950 million years after cyanobacteria started to produce oxygen, marking the start of the GOE. The oxygenation of Earth's atmosphere continued to be a gradual process. Only by 250 million years after the GOE began, around 2,200 million years ago, did the oxygen content of Earth's atmosphere reach a level of one percent. Today, the air we breathe contains close to twenty-one percent of oxygen. Cyanobacteria are still very much the source of that oxygen but now much more in the form of chloroplasts, the symbiotic cyanobacteria that are part of all plant cells.

Earth Climates and Ice Ages

Earth's climate and how it changed over the roughly past two million years, had a large impact on the path that human evolution took. Fortunately, we are today in a much better position to understand Earth's climates during this, at least on a geological time scale, rather short period that witnessed the evolution of the human species. But before we look into how we did and continue to learn about Earth's climate, a few words are needed about what climate actually is. Many people still confuse climate with weather, nobody seems to be immune from that as even some notable modern-day solons of otherwise seemingly advanced nations do not grasp the difference. Or in their case, as they are politicians, they actually may understand the difference but choose to ignore it to serve whatever interests pay for their publicly uttered opinions. The National Aeronautics and Space Administration, NASA, is responsible for the United States civilian space program, as well as aeronautics and aerospace research. NASA makes young girls and boys dream of futures unimaginable but then, maybe achievable after all; and many aspiring young engineers and scientists would love to work there one day. Everyone, even one of those solons mentioned above, can go to NASA's website, and read about the difference between weather and climate. It is all common sense and NASA says it as well as it can be said:[49]

> *"The difference between weather and climate is a measure of time. Weather is what conditions of the atmosphere are over a short period of time, and climate is how the atmosphere 'behaves' over relatively long periods of time."*

It does not take a genius to understand the difference, it rather takes a willful effort to confuse it. Maybe in our time, this is what politics for some among us is about. Earth's climate has been changing throughout its history and Earth can live with practically any climate but life cannot. Severe climate changes in the past have been responsible for some of the large extinction events we know of. Different forms of life have different tolerances for climate extremes. Some lifeforms made it through *Snowball Earth* in the Cryogenian, other lifeforms survived through long periods of arid and hot climates. Complex human societies such as the ones that have evolved over the last few thousand years will likely find it very difficult to adapt to severe climate changes like they have been experienced by some of our human ancestors; the ice ages certainly come to mind. It was not difficult for small human family groups in pre-historic times to pack up and leave for warmer latitudes. Moving billions of people with no "vacant" areas available to go to because we already have occupied much of the habitable zones of planet Earth, this becomes a different story. But then, we have the ability to understand Earth's climate and a quite impressive capability to develop solutions, if we put our minds to it. This is certainly something our ancestors could only have dreamed of. Therefore, the real question is not if we can do something about it but more likely: will we be ignoring the greatest challenge that our modern societies have ever faced, or will we address it in time? And if we do so, will we be doing it in such a way that the interests of all members of our human family scattered around the globe are considered equally? Only time will tell.

Paleoclimatology is the name given to the modern science committed to the study of ancient climates It has enabled us to reconstruct the climate conditions prevailing on

Earth over the last roughly two-and-a-half million years ago, a time period geologists refer to as the Quaternary (see plate VII); and it also has given us an understanding of climate conditions throughout the 541 million years of the Phanerozoic eon (see plate IV through plate VII). Finding the evidence and retrieving the samples that hold the secrets of past climates, makes paleoclimatology very much an outdoor science. Brought back into laboratories, these samples only reveal their secrets using the latest physical, chemical, and biological analysis methods. There are broadly speaking three sources of data to reconstruct past climates: geological, glaciological, and biological.

Geological data sources can come from marine or terrestrial environments. Marine data are obtained by drilling boreholes into the sea floor a few centimeters wide. The core samples obtained this way provide chronological records from the top to the bottom of the drill core. How far back in time these records can go is limited by the recycling of the oceanic crust, so for the Atlantic Ocean this would be about 150 million years at best. Today's deep sea drilling technology allows the exploration of deep ocean floor sedimentary layers with core samples extracted from a couple of kilometers below the sea floor. Marine sedimentary layers preserve Earth's record much better than similar samples drilled from continental sedimentary layers can. This is because marine sedimentary layers are not exposed to the same kind of weathering that can quickly erode continental sediments. Core samples from deep-sea beds can reveal all kinds of information about Earth history in their countless number of sedimentary layers. This can for example be through their elemental composition or the inclusion of trapped gases, or from the buried fossil records they contain. We have already seen earlier that the relative abundance of different oxygen isotopes can tell us something about seawater temperatures at any given geological time. There are different ways to reconstruct ancient ocean oxygen isotope ratios. But in the context here, one fossil type is of particular interest to paleoclimatologists, the so-called foraminifera, sometimes informally just referred to as forams.[50] Foraminifera are single-celled organisms, an example of a class called protists, which we will discuss more in the next chapter. Importantly, forams have shells and they have been around since the beginning of the Cambrian, some 541 million years ago. They exist in a wide variety of forms, an estimated 4,000 different species populate today's oceans, and can range in size from one-tenth of a millimeter to some twenty centimeters, though most of them are on the small side.

The large majority of foraminifera species live on or beneath the sea floor. When foraminifera die, their hard shells are incorporated into the sediments they have been living in. In this way, generations of them are buried over geological time in uniform layers. Foraminifera are very particular about the environments they prefer to live in. Some foram species can only be found in cold water, these are mostly the deep ocean dwellers. Then there are species that prefer warm and shallow waters, while others can only be found in brackish waters. Based on a detailed knowledge of the specific environmental preferences of these various species of foraminifera, paleoclimatologists can make inferences with respect to prevailing sea levels or ocean and climate conditions in the period when the respective sediments were deposited. Isotope analysis of oxygen and carbon atoms from foraminifera shells provide information about the amount of water trapped in ice sheets and the abundance of algae in the oceans. Plants prefer to use the lighter isotope of carbon, C-12, for photosynthesis and this shifts the atmospheric and oceanic balance between the heavier carbon isotope C-13 and C-12 in favor of C-13. Because of that, foraminifera will incorporate higher amounts of C-13 into their shells during times when plants and algae thrive. Evaporation and condensation of water molecules helps fractionate the heavier oxygen isotope O-18 from the lighter and more common oxygen isotope O-16. The way this fractionation works is that at a given temperature a lesser relative percentage of heavier water containing O-18 and a higher relative percentage of light water made of O-16 evaporates. This process runs in reverse for condensation where at a given temperature a higher relative percentage of O-18 water, and a lower relative percentage of O-16 water condenses. The net result is that a build-up of O-16 enriched ice sheets goes hand in hand with O-18 enriched oceans.[51] When ice sheets melt, oceans are replenished with O-16 rich water and the isotope ratio of oceans waters changes in favor of O-16. The oxygen isotope ratios of foraminifera shells can therefore tell us how much water was trapped in ice sheets through geological time.

Glaciological data to reconstruct past climates comes, as the name already implies, from glaciers, specifically from drilling ice cores. Ice sheets and glaciers accumulate snow layer by layer. Layers of freshly fallen snow are not compact but have air pockets inside. Those of us living at higher latitudes experience this every winter when shoveling freshly fallen snow or when making snowballs for a snowball fight or building a snowman for fun. As fresh snow layers accumulate on top of older ones, an increasing weight compacts deeper layers in the stack, thereby trapping some of that air. Eventually, the weight of newly added snow layers compresses the earlier snow layers into solid ice that contains gas bubbles retaining the atmosphere composition from the time a snow layer was deposited. Over time, glaciers can grow to thicknesses of several miles. Ice cores have mostly been drilled in Greenland and Antarctica because there the ice sheets are thickest and therefore provide a longer record of past climates.

To date, the deepest ice core drilling is done in Antarctica, and it goes down more than three kilometers. Our ice-core records for Greenland currently extend back to about 130,000 years while for Antarctica they reach back to around 800,000 years. Scientists can extract the climate records from ice cores because their layer structure preserves annual layers of snowfall and with those seasonal differences in the properties of snow; not unlike the growth rings of trees do in their own way for our more recent past. However, the deeper one goes in an ice sheet, the more the layers are compressed, becoming more difficult to differentiate. Fortunately, we have technologies that are much better than humans at identifying and counting snow layers and they help us to determine the age of ice in a drill core with increasing depth. The annual layer records preserved in the ice cores as read out through sophisticated laboratory analysis provide temperature data, chemical composition of any trace elements contained in the ice, as well as aerosol composition including dust, pollen, or ash, such as from volcanic eruptions. Most importantly, the preserved air bubbles of ancient air allow a direct analysis of past atmosphere compositions including the presence and percentage amounts of greenhouse gases such as methane and carbon dioxide.

Geological climate data from marine sediment records obtained from deep sea floor drilling can retrace climate history on the order of ten million years with a resolution of around one hundred years. It provides us with information on temperature, chemical composition of ocean waters, biomass or vegetation patterns, geomagnetic field variations like we discussed them earlier in the context of sea floor spreading, as well as precipitation data, sea level changes, and solar activity records. Additional geological climate data coming from terrestrial sources add further information such as obtained from secondary mineral deposits formed in caves, glacial deposits and erosion, coastal sediments and erosion features, and more. Glaciological climate data from ice cores currently allow us to go back on the order of one million years in climate history, less than what can be obtained from marine sediments. However, ice core data provide a much better time resolution of our planets climate history down to single years, reflecting the tree ring like records that annual snowfalls preserved in the ice cores provide. From ice cores the same types of data can be extracted as from marine sediments but in addition to the chemical composition of ancient ocean waters, the air bubbles preserved in ice cores give us also the actual chemical composition of ancient atmospheres; and they also provide data on volcanic activity. Another type of climate data comes from biological sources. Prominent among them are tree ring data. They are more location specific, have an annual resolution of one year and can trace climate history on the order of ten thousand years. Data from tree rings provide information regarding temperature or precipitation, including geomagnetic data if fossilized, or about volcanic and solar activity. Coral growth provides a similar time window into climate history with the same yearly resolution as tree rings. Coral growth data add to our records of the chemical composition of ocean waters, sea levels, temperature, and precipitation. Biological sources of course also include the fossilized foraminifera found in core samples from sea floors discussed earlier. Information retrieved from them comes with the same time resolution and climate time horizon as the marine core sediments they were buried in.

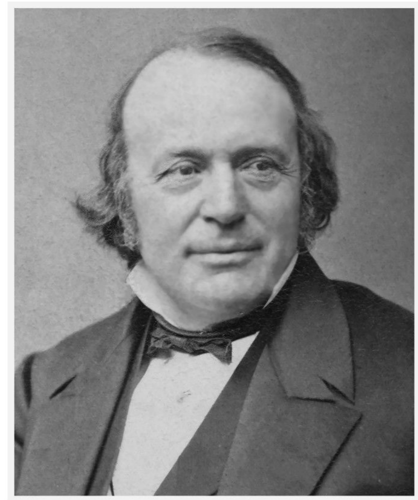
We know today that Earth has experienced several ice ages throughout its history, importantly over the past 2.56 million years of the Quaternary, which were critical for shaping human evolution. Ice ages are marked by alternating intervals of extensive glaciation, called glacials, and relative warmth, called interglacials. We currently live in a period of relative warmth, an interglacial that started about 10,000 years ago with the Holocene epoch (see plate VII). It is only a couple of centuries ago that scientists began to appreciate the significance of glaciations for Earth's past climates. We tend to forget that until quite recently our global maps of Earth still had literally white areas where we just did not know what was there. Young people growing up today find nothing special about the fact that humankind has set foot on the Moon only about a half-century ago. Back then, for quite a few of the older generation watching the first humans walking on the Moon, their childhood heroes still included daring explorers making the last great geographical discoveries on our own planet. Such as the ultimately futile search for the Northwest Passage or the excitement of the races to be first to set foot on Earth's North- and South Poles. The real extent of the Arctic- and Antarctic ice sheets became only known in the late nineteenth century. What caused the repeated expansion and contraction of these polar ice sheets for which there was increasing geological evidence remained a mystery. The discovery of the ice ages started in Europe where the generational memory of native populations whose ancestors had been living in alpine regions for hundreds of years had preserved the knowledge of glacier movements in their areas. Towards the middle of the nineteenth century, this would lead to the first formulation of ice age theories but it would take another century until science could provide a comprehensive explanation.

Discovering the Ice Ages

By the early nineteenth century, it had become clear that planet Earth once had been home to many species, a great number of which we only know from fossil finds. This invariably led to speculations as to what could have caused the loss of so many species. Catastrophic events as advocated by the *Catastrophism* hypothesis, frequently framed as biblical-type flood events, were popular explanations. Of course, floods must not have happened on biblical time schedules. They could well have happened at any time during the much longer geological time scales that became increasingly accepted with the theory of *Uniformitarianism*. Hence, it should be no surprise that the main proponent of *Uniformitarianism*, Charles Lyell, explained one of the geological otherwise not explainable facts with a modified version of biblical flood events. The geological peculiarities posing the problem Charles Lyell tried to solve were large rocks found in places where they evidently did not belong. Somehow these large boulders, some of them the size of small houses, had been transported over vast distances from the places where Earth's geology was made of such rocks. Lyell's widely accepted theory was that these boulders had been moved there by large icebergs carrying such rocks while they drifted in the biblical flood waters. One problem, though conveniently ignored by most, was that any such flood capable of achieving this would not just have to be of biblical proportion but would require much more water than all the surface water Earth carries. Some of these boulders were located high up in mountain areas, which of course implied a sea level of equal height for an iceberg to drift there and to leave boulders behind once the iceberg had melted. The problem is that there is no possible explanation as to where the water could have come from to produce a sea level of a few thousand meters above today's sea level; no wonder this problem was swept under the rug. Since otherwise Lyell's theory could explain most of the remaining observations, like the sediments that we now identify as end moraines of glaciers, it continued to be a widely supported hypothesis for quite some time. However, there were other problems with Lyell's theory. One of them being the grooved rock surfaces that the moving glaciers left behind. For many this did not matter much, but for others it would be the starting point to develop a completely new explanation: ice ages.

The man who was going to become the main advocate for the ice age theory had for quite some time no doubts himself that the existing theories explained most facts to his satisfaction. Even though, there were already others that had voiced the opinion that some of these facts as they found them in the European Alps could be explained by local glaciations. The ones who had a good idea of what really happened were actually the people whose communities through history had seen glaciers expand into their mountain valleys only to retreat again after a few generations. These people knew well that those glaciers had transported large rocks over several kilometers into places where they did not originally belong. Eventually, geologists picked up this knowledge and finally it reached some who could make a difference in changing the views of the broader community of geologists. One such person was the German-Swiss geologist Jean de Charpentier (1786 – 1855). It was through him that the young Swiss-American biologist Louis Agassiz changed his opinion in 1836 regarding ice ages and like any true convert, Agassiz did not just become a disciple of the ice age idea, he became a preacher.

In 1837, at a meeting of the Swiss Society of Natural Sciences, Agassiz proposed that Earth had been going through an ice age. The young scientist presented it as the explanation for the past extinction of Earth's life with the geology right at his doorstep providing proof for such an ice age. All around him in his native Swiss Alps, he could study the evidence for glaciers covering much of central Europe. Rock striations indicating the flow of glaciers, valleys carved out by gigantic ice masses, as well as boulders and end moraines left behind by retreating glaciers, all indicated to him that Earth in its past was subject to what he called a sudden intense winter that turned into an ice age. The concept of an ice age, as mentioned above, was not a new one and familiar to many of his colleagues but they never had considered it in the context of explaining the facts that *Catastrophism* supposedly already explained. In addition, they never had quite envisioned it on the scale that Agassiz proposed. We need to remember, that this was before the time of the great polar explorations of the late nineteenth and early twentieth century associated with the names of the Norwegians Fridtjof Nansen (1861 – 1930) and Roald Amundsen (1872 – 1928), the British Robert Scott (1868 – 1912) and Ernest Shackleton (1874 – 1922), and the American Robert Peary (1856 – 1920). The vast extent of the Greenland ice sheet was virtually unknown, and scientists knew even much less about the size of the Antarctic ice sheet. Understandably, many of his colleagues viewed a glaciation of most of the northern parts of Asia, Europe, and North America with considerable doubt. It also did not help that in his zeal, Agassiz even asserted that European ice sheets must have extended to the Mediterranean, for which there was simply no geological evidence. Eventually, it would take almost a quarter of a century until the ice age theory finally became fully accepted. The process started with Agassiz's 1840 publication of his two-volume work, *Études sur les glaciers* (*Studies on Glaciers*), in which he explained his ice age theory in detail for the broader community of naturalists to consider. Early converts included the doyen of English geology, William Buckland (1784 – 1856), who in short order also convinced his former student Charles Lyell of the correctness of the ice age theory. However, even after Lyell had dropped his iceberg drift theory in favor of the ice age explanation, others would continue to cling to Lyell's former theory for years to come.



Louis Agassiz
1807 – 1873

Agassiz's great ice age theory made him famous and he moved on to study the geological evidence for ice ages in the United States where he became a professor at Harvard University. Towards the end of the first half of the nineteenth century, natural scientists not only looked at the evidence for ice ages in the geological record, they had also started to connect past changes in Earth's climate with astronomy. The first to do so was the French mathematician Joseph Adhémar (1797 – 1862). In his *Révolutions De La Mer* (*Revolutions of the Sea*), published in 1842, he suggested that ice ages were related to the ~22,000-year cycle of the precession of the equinoxes. At the equinoxes, Earth's poles are at the same distance from the Sun and day and night are of equal length.

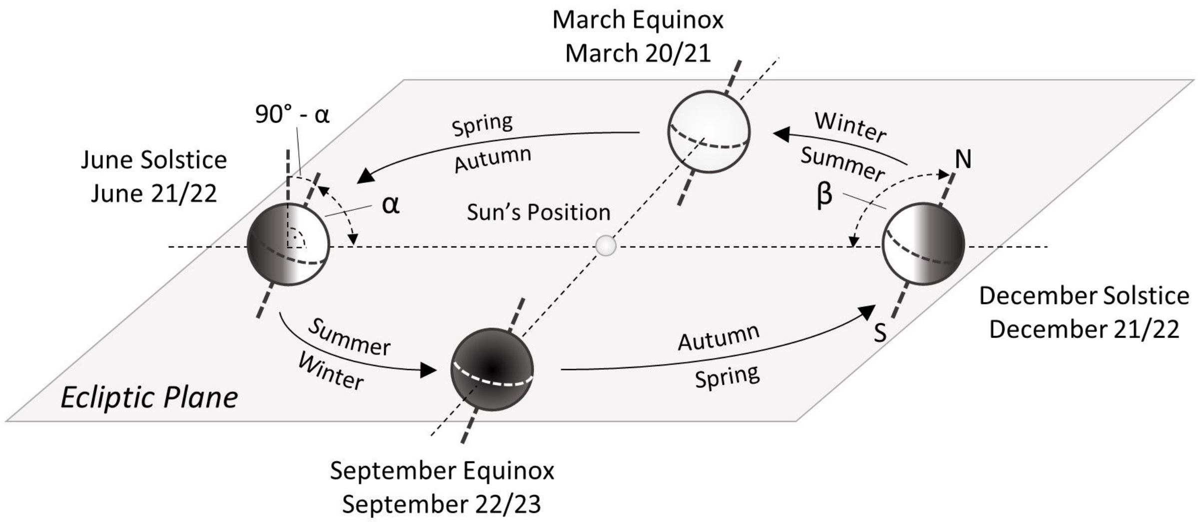


Figure 2.5: Earth's positions in its orbit around the Sun at the equinoxes and the solstices. The light and dark shaded part of Earth's globe show day and night side, respectively. Season labels above / below the ecliptic plane are for the northern / southern hemisphere. At the equinoxes, Earth's poles are at the same distance from the Sun and both hemispheres receive the same amount of sunlight. At the June solstice the northern / southern hemisphere receive the most / least sunlight; at the December solstice, the reverse is the case.

Hence, Earth's northern and southern hemispheres receive equal amounts of sunlight. The positions in Earth's orbit around the Sun where the equinoxes occur shift over time, returning in ∼22,000 years to the positions they have today. Adhémar also theorized that because the northern hemisphere during its winter receives less sunlight, snow would accumulate at the North Pole while the same would happen at the South Pole during the southern hemisphere winter. Because Earth moves faster through its orbit when closest to the Sun, northern hemisphere winters should be shorter than southern hemisphere winters and therefore more snow should accumulate at the South Pole. Before we move on to the development of the modern astronomical theory in the twentieth century, it will help to briefly review the key astronomical factors influencing Earth's climate. Fig. 2.5 shows a schematic picture of Earth orbiting around the Sun in its ecliptic plane. The angle between Earth's rotational axis and a perpendicular line drawn through the ecliptic plane is currently 23.4°. In fig. 2.5, this angle is identical to the angle $90^\circ - \boldsymbol{\alpha}$, $\boldsymbol{\alpha}$ being the angle at which at the June solstice Earth's rotational axis intersects the line connecting Earth and Sun. At the June solstice position, the northern hemisphere receives much more sunlight than the southern hemisphere, it is the beginning of the summer and winter seasons in the northern and southern hemispheres, respectively. From there, as Earth travels around the Sun, the angle between Earth's rotational axis and the line connecting Sun and Earth increases. It reaches its maximum value at the point of the December solstice, indicated by the angle $\boldsymbol{\beta}$ in fig. 2.5, where it equals 90° plus the 23.4° between the normal to the ecliptic plane and Earth's rotational axis for a value of 113.4°. Clearly, at this point the northern hemisphere receives much less sunlight ringing in the winter season while the southern hemisphere enjoys the beginning of summer. As Earth travels back to its June solstice position, the angle between Earth's rotational axis and the line connecting Sun and Earth decreases, reaching the value of α again at the next June solstice. In between the June and December solstices, Earth passes through the points of the March and September equinoxes. At both of these positions, the angle between Earth's rotational axis and the line connecting Sun and Earth becomes exactly 90°. At these points, a projection of Earth's rotational axis onto the ecliptic plane results in a line that is at a ninety-degree angle to the line connecting Earth and Sun. At the position of the equinoxes, both, the northern and the southern hemispheres get equal amounts of light, day and night are of equal length. One can mirror the graphic in fig. 2.5 through the ecliptic plane and rotate it by 180 degrees and give the same description as above, just with the words "northern" and "southern" exchanged. For historical reasons, the North is always depicted as "up" and the South as "down". If astronomy had started in the southern hemisphere, we can be confident that scientists on the southern half of the globe would have looked at themselves as being "up" as everything below their feet must have been perceived as being "down", including most certainly the other side of the globe. What fig. 2.5 clearly shows is that the seasonality we experience is due to the changing angle between Earth's rotational axis and the line connecting the positions of Earth and Sun. For mysterious reasons, many still seem to believe that Earth's seasons are dominated by the varying distance between Sun and Earth. That is definitely not the case. If Earth's axis were perpendicular to its plane of rotation around the Sun, there would be virtually no seasons at all. We have to thank the planet Theia for producing not only the Moon through its collision with Earth but also most likely for knocking Earth's axis out of its normal position with respect to the ecliptic plane by just a small amount so that we do have seasons. If Theia had collided with Earth in a different way, our seasons may be quite different, or we may not have any seasons at all. Earth's axis could for example have come to lie in the ecliptic plane oriented along the line connecting the positions of Sun and Earth. In this worst case, one side of Earth would always be on the night side in freezing cold, while the other would always be on the day side, constantly exposed to the Sun; pretty much the position Uranus is in today. This would likely not have been an ideal situation for life to develop, but one never knows, nature seems to be very creative in getting life to flourish even under harsh conditions. Fortunately, we do not have to worry about this as our Earth's rotational axis sits at a comfortable 23.4° off the normal to the ecliptic plane.

There are three basic cycles in Earth's rotation around the Sun that have an impact on the solar radiation that Earth receives and how this radiation varies across Earth's latitudes at different times of the year. The three cycles are related to Earth's ecliptic plane itself, the tilt of Earth's rotational axis with respect to its ecliptic plane, and changes in the direction of Earth's rotational axis as projected onto the ecliptic plane. We will start with the first cycle, changes to Earth's orbit, the ecliptic plane itself. Earth's elliptic orbit is almost a circle with an eccentricity of only around 0.017. If the eccentricity would approach zero, Earth's orbit would become a perfect circle. An eccentricity between zero and one describes an ellipse and if the eccentricity would reach the value of one the ellipse would open up to become a parabola; if that were ever the case Earth would stop orbiting the Sun and fly off into space.

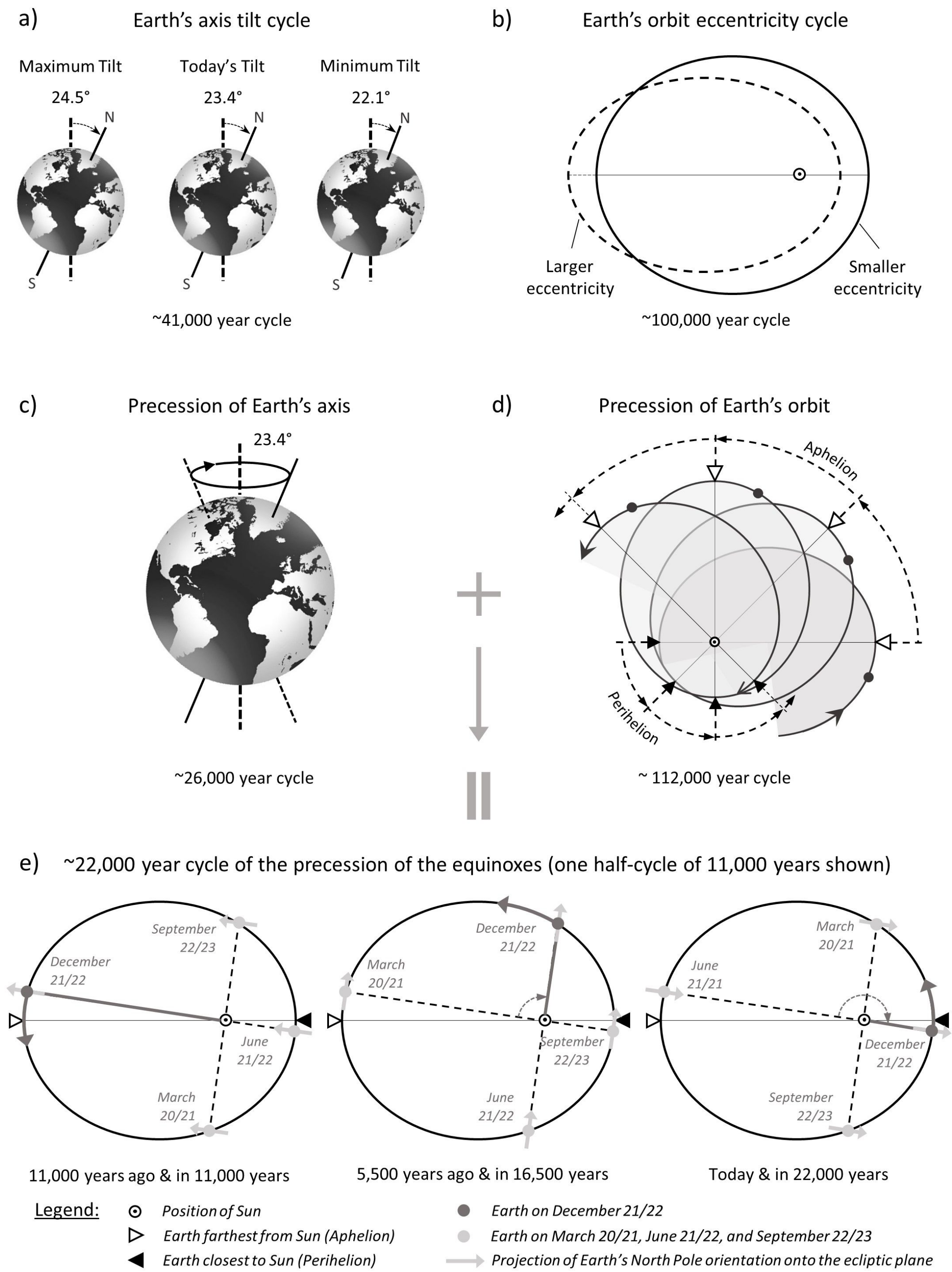
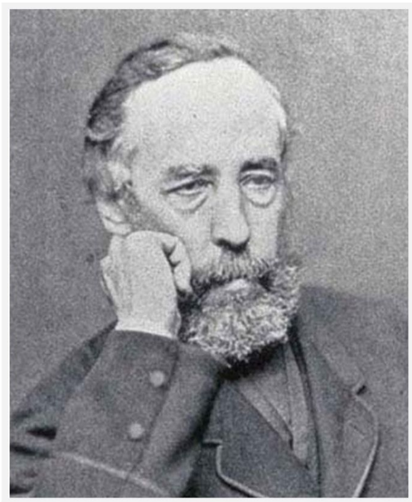


Figure 2.6: Astronomical cycles impacting Earth's climate: (a) Earth's axis tilt cycle of ∼41,000 years; (b) Earth's eccentricity cycle averaging a ∼100,000 period; (c) Earth's axis precession of ∼26,000 years; (d) Earth's apsidal precession with a ∼112,000 year cycle; and (e), the ∼22,000 year cycle of the precession of the equinoxes resulting from (c) and (d). The drawings very much exaggerate the eccentricity of Earth's orbit.

An eccentricity of 0.017 is quite small as the major axis of such an ellipse is less than 0.015% longer than the shorter axis. However, the eccentricity of Earth's orbit does not stay constant but oscillates over time. Past eccentricity values of Earth's orbit have varied in a range between being even closer to zero than it is today and about 0.068, still very close to a circle with a 0.23% difference between the major and minor axis of the ellipse. Currently we are at the low end of eccentricity values in this range, on our way to the minimum of the current eccentricity cycle before eccentricity will increase again. Earth's eccentricity cycle is actually the result of three different variations effecting eccentricity which oscillate on 95,000, 125,000, and 413,000 year cycles and combine to produce on average a roughly 100,000 eccentricity cycle (fig. 2.6b); or more precisely a mean value of 95,800 years. As eccentricity varies, the major axis of Earth's ellipse does not change, and eccentricity increases or decreases by the minor axis decreasing or increasing in length. Not only does Earth's eccentricity change over time, the orientation of Earth's orbit around the Sun changes as well. This motion of the Earth's ellipse whereby the line connecting its closest point to the Sun, the perihelion, and the farthest point from the Sun, the aphelion, rotate around the Sun is called apsidal precession; the apsides being the two ends of the line connecting perihelion and aphelion, the major axis of Earth's ellipse. The apsidal precession of Earth's orbit around the Sun occurs on a 112,000-year cycle (fig. 2.6d).

Earth's rotational axis has not always been inclined at 23.4° towards the normal of the ecliptic plane. The magnitude of this inclination has oscillated in the past within the range of 22.1° – 24.5° over a 41,000-year cycle (fig. 2.6a). For a given cycle, the minima and maxima have been higher than the 22.1° and lower than the 24.5°. The last time this angle was at a maximum value is about 8,700 years ago and in another 11,800 years it will have decreased from its current value of 23.4° to the minimum value of the current cycle. Without the stabilizing influence of the Moon, Earth would most likely experience much larger changes in the tilt angle of its rotational axis. A larger tilt will produce more pronounced seasons with an increased amount of sunlight received at higher latitudes at the hemispherical summer solstices and a correspondingly decreased amount of sunlight received at higher latitudes at the hemispherical winter solstices. With the ecliptic angle at the low value of its range, the four seasons as experienced in temperate and higher latitudes would be less pronounced; if Earth's rotational axis were not tilted at all there would be no seasons to talk about, it would be equinoxes all year round.

Not only does the magnitude of the inclination of Earth's rotational axis towards the ecliptic plane vary, but its orientation also changes over time as Earth's axis itself rotates around the normal of the ecliptic plane. This movement is known as the precession of Earth's rotational axis. It takes about 26,000 years, 25,772 to be exact, for one full cycle of this precession to complete (fig. 2.6b). Largely, this precession of Earth's rotational axis is caused by the combined gravitational pull of Sun and Moon exerted on Earth's equatorial bulge; gravitational interactions with other planets matter as well but to a much lesser degree. Finally, the combination of the ∼112,000 year cycle of the precession of Earth's orbit around the Sun with the ∼26,000 year cycle of Earth's axis precession produces the observed cycle of the precession of the equinoxes (fig. 2.6e), averaging to around 22,000 years with a precise mean of 21,700 years and including major frequency components around 19,000 and 23,000 years.



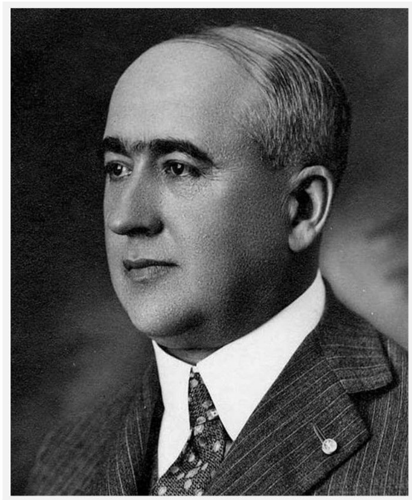
James Croll
1821 – 1890

Fig. 2.6e illustrates the clockwise precession of the equinoxes and solstices with a ∼22,000-year cycle as they shift over one half-cycle of this precession, looking back into time as well as forward into the future when the same constellations are expected to return again. The Scottish scientist James Croll was first to develop an astronomy-based theory to explain Earth's climate changes. In doing so, he built on the work of the French mathematician Le Verrier and also on Adhémar's work. Le Verrier had developed methods and calculations that led him to believe that the discrepancies with respect to the observed and calculated orbits of planet Uranus were due to another yet undiscovered planet; as we saw in the previous chapter, the planet whose orbit Le Verrier predicted turned out to be the planet Neptune. Now Croll used Le Verrier's methods to calculate how the eccentricity of Earth's orbit around the Sun had changed over time. Le Verrier's calculations had shown Croll that Earth's orbital eccentricity over the past 100,000 years had varied between low values close to the current eccentricity value and high values of up to about 0.078; equating length differences between Earth orbit's major and minor axis of 0.015% and 0.31%, respectively

Croll's theory proposed that during times of high orbital eccentricity winters tend to be colder because Earth is farther away from the Sun than during times of low orbital eccentricity as is currently the case. Croll analyzed Earth's orbital eccentricity over the past three million years and predicted that Earth's orbital eccentricity changed cyclically with intervals of high eccentricity lasting many tens of thousands of years alternating with long intervals of low eccentricity. His data showed long cyclical variations with a high eccentricity orbit in place approximately 100,000 years ago and reaching a low eccentricity orbit some 30,000 years ago with only a small increase in eccentricity since then. A larger eccentricity means that the Earth will be closer to the Sun at its perihelion, but because of Kepler's second law, it will also move faster through it. The average increase or decrease in temperature due to Earth's orbit having a higher eccentricity and putting it closer or farther from the Sun at its perihelion or aphelion is therefore only small. This was already clear to Le Verrier who had shown that the total sunlight received by Earth in its orbit around the Sun over one year is changed only very little by variations in its orbital eccentricity. Croll was aware of this difficulty and therefore argued that the radiation Earth receives during each season varies strongly with eccentricity. He conjectured that while the total integrated energy Earth receives during one year in a low eccentricity orbit is not very different from that received in a high eccentricity orbit, the respective seasonal changes are more pronounced. The amount of solar radiation currently impinging on Earth's atmosphere in its low eccentricity orbit is roughly 7% higher at the summer solstice than it is at the winter solstice. At a higher eccentricity, this difference gets more pronounced. For the maximum orbital eccentricity value Croll used, the difference between sunlight falling on Earth's atmosphere at summer and winter solstices is around 35%. The amount of sunlight impinging on Earth's surface is one variable but a significantly more important one is how much of that energy actually gets absorbed and how much is reflected back into space. Importantly, Croll recognized the continued accumulation of snow over many seasons and the resulting expanding snowfields as a major contributor in reflecting sunlight and thereby reducing the solar radiation absorbed by Earth in a positive feedback loop. Less sunlight results in more snow and more snow results in more sunlight reflected back making for cooler winters – a positive feedback loop.

Croll had essentially discovered two facts that are critical in understanding climates and climate variations, the importance of Earth's surface in reflecting back sunlight, the measure of which we call today albedo, and the importance of feedback loops in understanding climates.[52] In a way, Croll was forced to conclude, correctly as it turned out, that the changes in solar radiation his astronomical theory implied were not the primary cause of the ice ages. The overall magnitude of temperature changes due to variations in Earth's orbital eccentricity were just too small. However, he also was correct in understanding that these small changes were sufficient triggers that could amplify seasonal changes, leaving more snow-covered areas during intervals when Earth was farther away from the Sun at winter solstices, which would in itself, through the increased albedo of Earth's surface, lead to lower temperatures. Croll was on the right track but some information was just not available to him or to anyone else in his time. He also suspected that changes in the Earth axis tilt were important, but he had no data to tell him how big those variations were and on what time scales they occurred. His theory, summarized in his book *Climate and Time* published in 1875, was received with a lot of interest. Geologists started to look for evidence of the last ice age he had predicted to have ended around 80,000 years ago but what they found indicated that it had actually ended only 10,000 years ago. In addition, technology at the time did not allow testing of one of Croll's key predictions, alternating glaciation periods in the northern hemisphere with an 11,000-year cycle. Eventually, Croll's ice age theory would be largely disregarded for almost a century until it became clear that the 100,000-year cycle Croll had predicted was indeed one of the major cycles reflected in past ice ages.

Almost a quarter century would go by before the next major step in the development of the astronomical theory to explain the ice ages. The Serbian mathematician and climatologist Milutin Milanković published his first works on this topic in his native Serbia before the First World War but it would eventually take his whole career before he completed it by 1941. In 1920, Milanković published *Théorie mathématique des phénomènes thermiques produits par la radiation solaire* (*Mathematical Theory of Heat Phenomena Produced by Solar Radiation*) and that brought his work to the attention of one of the foremost climatologists of his time, the Russian-German Wladimir Köppen (1846 – 1940). Incidentally, Köppen also happened to be the father-in-law of Alfred Wegener. This collaboration benefited both sides as it gave Köppen and Wegener a tool to understand past climates while Milanković work benefited from the climatology and geology expertise Köppen and Wegener shared between themselves. In this way, the latter two helped Milanković understand that the variation of solar radiation is not most important at the highest latitudes, but around latitudes where snow does actually melt during the summer season but with a small reduction in solar irradiation may not do so completely anymore.



Milutin Milanković
1879 – 1958

If more snow accumulates during winter in areas where it does not melt anyway this will have little impact. However, if the area where snow stays on the ground longer increases because of less solar radiation received during the summer, more sunlight will eventually be reflected back into space as snow accumulates in these areas. Therefore, Milanković calculated the variation of summer solar radiation at three northern latitudes of interest that cover this critical area of summer snow melt in the higher regions of Europe for the past 650,000 years. When Köppen saw these curves, he was able to match them with the pattern of glaciations observed in the European Alps that he was familiar with. These alpine glaciation periods, still being learned by European students today, are named after four rivers, namely, Gündel, Mindel, Riss, and Würm. Both, Köppen and Milanković, were intrigued by how well the temperature minima in Milanković calculations matched up with the spacing in time and the duration of these four alpine glaciations that had been establish earlier in the twentieth century. Eventually, Milanković astronomical theory would become one of the key tools to enable the reconstruction of past climates but only by late in the second half of the twentieth century. It was unknown to Köppen and Milanković as well as to anyone else at the time, that the 650,000-year date assumed for the start of the Pleistocene epoch, which had ushered in Earth's modern ice ages, was wrong. Nor was it known then that the chronology of the four alpine glaciations was not on solid enough ground and would be successfully challenged in the 1970's. Also, much sooner than that, other scientists rightly started to point out that there were more factors in addition to solar radiation to consider. It would take several decades and an international coordinated effort among the leading science institutions in this field to finally work out the ice age chronology of the past two million years. Starting in the 1950's scientists could access new techniques and as frequently happens with a lot more data becoming available, the climate puzzle of Earth's past began to look more complicated. Clearly, ice age patterns were not as simple as the four alpine glaciations would have suggested. Many more ice ages or ice age phases were identified and more often than not, the geological record was somewhat difficult to read as anyone ice age destroys much of the geological record of the ice age preceding it. Adding weathering into the equation made it even more difficult to establish the ice age chronology from land-based geology. The answer would have to come from undisturbed sedimentary cores sampled from the bottom of the sea. It took until the early 1970's for a clear picture of Earth's climate history over the last 2.59 million years to emerge, the period that geologists and climatologists refer to as the Quaternary. The first epoch in the Quaternary is the Pleistocene, including most of human pre-history, and the second is the Holocene which started 11,700 years ago, roughly coinciding with the beginning of the last phase of the Stone-Age, and it is still the period we live in (see Plate VII). The difficulty in establishing Earth's climate history over the Quaternary lay with the deep Pleistocene and the challenge of finding an accurate chronometer for precise dating.

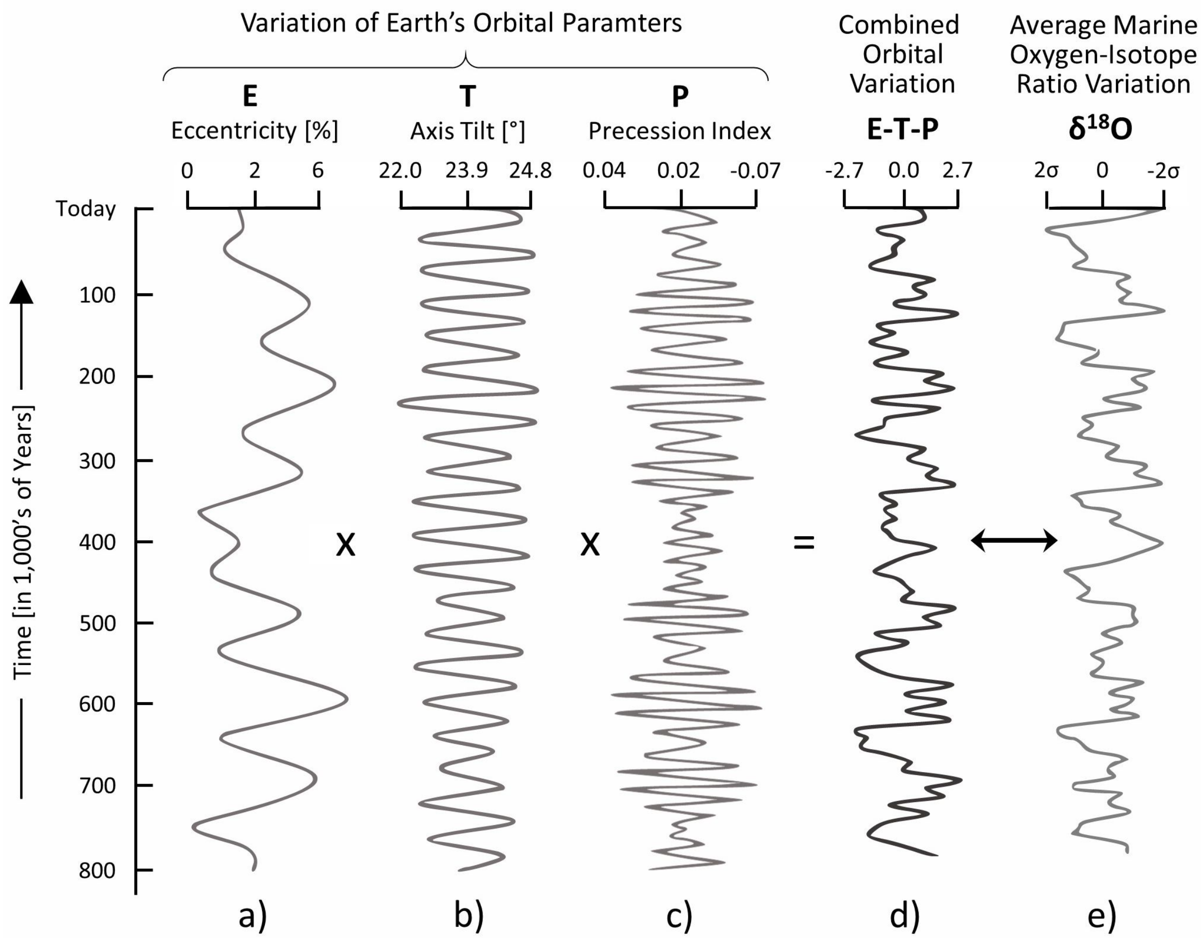


Figure 2.7: Proof positive for the astronomical theory of the ice ages. From left to right the graph shows for the past 800,000 years: (a) Earth's orbit eccentricity variation on a ∼100,000 year cycle; (b) the ∼41,000 cycle due to Earth's axis tilt variation; (c) the precession cycle of Earth's equinoxes with ∼19,000, ∼22,000, and ∼23,000 year frequencies; (d) the total variation in solar radiation received at a latitude of 65° N through the combined effect of (a), (b), and (c); and, (e) the variation in the oxygen isotope ratio recording changes in ice sheet volumes. There is a very good correlation between the curves (d) and (e) with low and high solar radiation periods being matched by larger and smaller ice sheet volumes. For data sources see Refs. [17–20].

To establish the chronology of Earth's climate throughout the Pleistocene using drill core samples from the bottom of the sea, any evidence in these samples that could provide information on Earth's climate had to be connected to the true age of the sediment they were found in. Like with sea floor spreading, magnetic material properties provided the key to understanding. In this case, it was the drill core layers which recorded the chronology of changes in Earth's magnetic field direction. By using such so-called magnetic reversal boundaries they could identify in cores drilled from deep sea floors, scientists were able to date climate changes as for example recorded by the foraminifera discussed earlier to a precise age derived from Earth's recorded magnetic history. The graph in fig. 2.7 shows the data that provided proof positive that the astronomical theory was indeed producing the correct long-sought explanation for the chronology of ice ages in the Pleistocene [17–20]. But it was a modified version of Milanković original theory including now the roughly 100,000 year cycle of Earth's eccentricity variation that James Croll had first identified, shown in fig. 2.7a. Fig. 2.7b and fig. 2.7c show the variations of Earth's axis tilt and the precession of the equinoxes over the same time frame. All of the three variations in Earth's orbital parameters are combined in fig. 2.7d to calculate the solar variation at the latitude of 65° N. As discussed earlier, the plot of the oxygen isotope ratio between O-16 and O-18 in fig. 2.7e is a measure of the amount of Earth's surface water contained in ice sheets. This oxygen isotope data came from drill cores taken at the same latitude for which the solar radiation curve was calculated. The chronology of the samples in the drill core that provided the isotope data was established using magnetic reversal boundaries from that core. As can be seen clearly, there is an excellent correlation between the data records in curves fig. 2.7d and fig. 2.7e with low and high solar radiation values in curve fig. 2.7d matching larger and smaller amounts of water bound in ice sheets in curve fig. 2.7e. It took more than 150 years for science to understand what triggered the glaciation periods in our current ice age, which started with the Pleistocene 2.59 million years ago. We are living today in an interglacial period, 11,700 years into the Holocene. Eventually our current warm period should be followed by another glaciation period.

The discovery of the ice ages was about understanding the ice ages of the Pleistocene, but ice ages have been a frequent occurrence throughout Earth history. Sometimes almost the whole Earth seemed to be in an ice age. This was for example the case during the Sturtian glaciation also called the Sturtian ice age in the Cryogenian period some 700 million years ago, an event some refer to as *Snowball Earth* or *Slushball Earth*, occasionally also as *White Earth* as much or all of Earth was covered in ice. For any ice age in Earth history, we can assume the astronomical theory to be equally relevant. However, what was different in Earth's distant past are the continental configurations. Like how much landmass there was and what its distribution looked like with respect to North and South Poles. As we have already seen, when much of Earth's landmass was consolidated in the supercontinent Gondwana and lodged around the South Pole we find evidence for an ice age. This so-called Karoo glaciation or Karoo ice age happened during the Permian period more than 250 million years ago and it is this ice age that left its traces on several continents now separated that would provide the evidence convincing Alexander du Toit that South America and Africa where once connected. Such a glaciation event should not surprise, as more landmasses close to any of the poles allow ice sheets in the respective hemisphere to grow, thereby increasing its albedo.

The continental configuration and the scientific framework of the astronomical theory will remain the key factors for predicting future ice ages. However, there is now also a new factor. Since Milanković put the final touches to his theory, the human factor has become much more important when trying to project how Earth's climate will likely evolve. Without the human factor it looks like we can expect our current warm climate period, which already is an exceptionally long one, to last for another 50,000 years before the next glacial period will produce a cooler climate. However, depending on the long-term impact of human induced climate change, that may turn out not to be correct.

Earth's Climate System

Earth's climate system is complex. The astronomical theory explained how variations of Earth's orbital parameters triggered Earth's ice ages; and it was a critical milestone in discovering one of the fundamental factors shaping Earth's climate – solar forcing. However, there are several other factors equally important in understanding past, present, and potentially future Earth's climates. During the second half of the twentieth century, we vastly improved our understanding of what scientists and media alike refer to as Earth's climate system. This was mostly owed to a combination of advances in a number of scientific fields which over time became interwoven into a new scientific discipline, climatology. Already the effort to proof or disproof the astronomical theory of the ice ages had been a large scale supranational and interdisciplinary scientific effort, the result of which was that by the 1980's the astronomical theory had become the accepted scientific model to explain Earth's ice ages. Climatology has not rested on these laurels but developed a broad based, data driven understanding of how the components of Earth's climate system interact. This includes the atmosphere, the oceans, the biosphere, the land surfaces, and the parts of Earth covered in ice, aptly named the cryosphere.

By volume, nitrogen and oxygen are the main components of Earth's atmosphere making up approximately 99% of the dry atmosphere. Nitrogen, with about 78% the most abundant component of Earth's atmosphere, is absorbed from and returned to the atmosphere by the nitrogen cycle of Earth's biosphere. Similarly, life also keeps Earth's oxygen atmosphere supplied with an oxygen content of currently around 21%. The rare gas argon makes up close to another 1% and then there are more minute amounts of neon, helium, and hydrogen contained in our atmosphere. Importantly, the lower atmosphere of which the above percentage numbers are indicative also has variable components including water, carbon dioxide, methane, nitrous oxide, and ozone. Water as a volume percentage is higher at sea level where it averages around 1% while its average in the entire atmosphere is around 0.04%. Carbon dioxide is currently also around a 0.04% level and as we know since the beginning of the industrial revolution the amount of carbon dioxide in the atmosphere is on a continued increase.

Carbon dioxide is a greenhouse gas, which means that it absorbs a portion of the infrared radiation coming from Earth and reemits it in all directions with some of it ending up reheating Earth. Greenhouse gases in essence act as blankets in Earth's climate system that retain a good part of the heat Earth's surface receives from the Sun which otherwise would be lost to space. Hence, greenhouse gases are not intrinsically bad for life on Earth, they likely have been critical in keeping Earth's surface sufficiently warm early on when the Sun was much weaker. True, most of the surface heat necessary to provide the conditions for liquid water on early Earth came from the heat in its interior retained from the time of its formation as well as from continued radioactive heating. However, if Earth did not have a greenhouse gas blanket back then maybe more of that heat would have been lost to space. A potential consequence of which could have been that liquid surface water may not have existed as early in Earth history as it did. Maybe thereby delaying the start of life or creating conditions for a very different beginning of life on Earth. Greenhouse gases by themselves are neither good or bad, they have been part of Earth's atmosphere since it first formed.

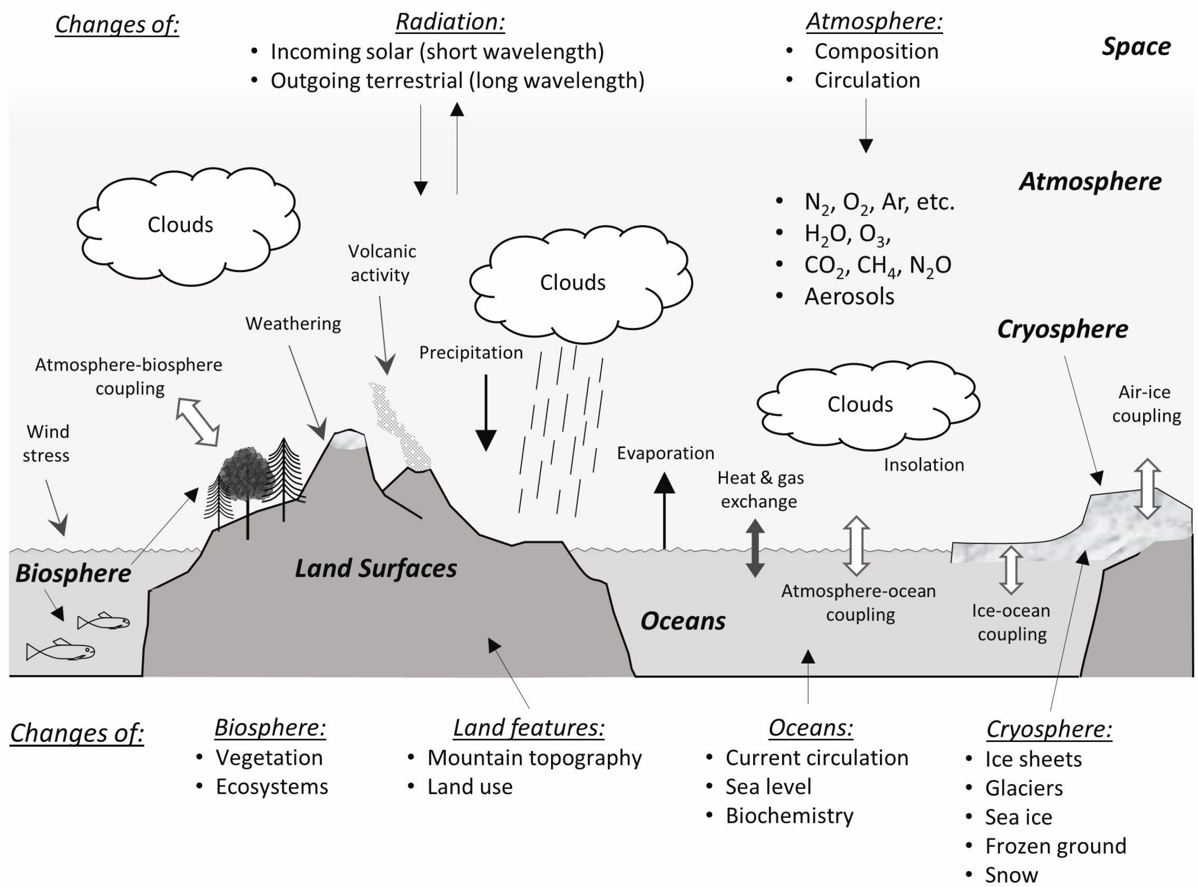


Figure 2.8: A simplified schematic of major components of Earth's climate system. The human climate factor is not included in this diagram.

However, their concentrations do matter very much. Our fragile human civilizations depend on Earth's climate to stay within a reasonably narrow range. Throughout Earth history, carbon dioxide variations in the atmosphere strongly correlate with major climate changes. In what is called the biological carbon cycle, the amount of carbon dioxide in the atmosphere reflects a balance between natural carbon dioxide absorption and storage such as by plants or the oceans and the amount of carbon dioxide given back to the atmosphere as plant matter decays or oceans warm. This biological carbon cycle does not include human contributions; nor does it consider major natural carbon dioxide sources like volcanic activities which occasionally can add large amounts of carbon dioxide as well as other dangerous climate gases and particulates to the atmosphere. The amounts of methane and nitrous oxide in the atmosphere are even smaller, averaging around 0.00018%, a percent fraction of eighteen in one hundred thousand or 0.18 thousandth of one percent for methane; and just below 0.000004% for nitrous oxide, less than a percent fraction of four in a million or 4 millionth of one percent. Tiny amounts indeed but both chemicals are quite potent greenhouse gases, much more so than carbon dioxide, specifically so because both of them can stay in the atmosphere for a very long time. There are natural biological sources for both greenhouse gases, but the human introduced contributions have taken over. Our use of fossil fuels, livestock farming, and landfills account for more than two-thirds of human contributions to methane in our atmosphere. As human caused global warming continues and the vast permafrost areas start to thaw, we can expect a substantial release of methane from these natural methane storage areas; and as microbes decompose once frozen organic matter they will produce more methane and carbon dioxide. Permafrost soils are not the only methane reservoirs on Earth. Enormous amounts of methane hydrates remain sealed away in our oceans at various depths with the right combinations of temperature and pressure. If the temperatures of our deep oceans rise, we could eventually see some of that methane released into the atmosphere.

Sometimes, climate change skeptics, to use a polite euphemism, contend that it is ridiculous that such small amounts of greenhouse gases, as we find for methane in today's atmosphere, can contribute to global warming. It is best not to respond directly and better to offer them a dish with two choices of spice, a pinch of regular pepper ranging somewhere below one hundred Scoville heat units and a pepper that comes in at the high end, say at more than one million Scoville heat units.[53] This should teach them that the amount of a substance says nothing about its potency, for spices and greenhouse gases alike, to pretend otherwise is just playing stupid.

Earth's atmosphere contains a small amount of water. Like carbon dioxide, water is also a greenhouse gas; and just like carbon dioxide, it has been a part of Earth's natural greenhouse blanket for a long time. However, an increase of water in the atmosphere is a climate reaction to the human induced increase in carbon dioxide rather than a direct result of human activity. As the human induced climate change results in higher surface temperatures, more surface water evaporates leading to an increase of water vapor in the atmosphere. This is an example of a positive climate feedback loop where a human induced change triggers a climate reaction, which amplifies the change set in motion by humans in the first place. In addition to raising carbon dioxide and methane levels in the atmosphere, gases that in very small amounts are part of Earth's natural atmosphere, a number of other very potent greenhouse gases have been added exclusively by human industrial activities. Prominent examples include chlorofluorocarbons, or CFCs for short, and sulfur hexafluoride, with the chemical formula SF_6. The former have a number of industrial applications, many of them in refrigeration but for a long time also as propellants in spray cans. SF_6, while less known, is practically the most widely used insulation medium in high voltage switches and incidentally the most potent greenhouse gas currently known.

Earth's atmosphere is characterized by a number of layers that are separated by changes in atmosphere temperature, temperature profile or both. Starting from Earth's surface these layers are the troposphere, the stratosphere, the mesosphere, the thermosphere and the exosphere.[54] These layers are separated by boundary layers: the tropopause between the troposphere and the stratosphere; the stratopause between the stratosphere and the mesosphere; the mesopause between the mesosphere and the thermosphere; and the thermopause between the thermosphere and the exosphere. Fig. 2.9a shows the average temperature profile of Earth's atmosphere up to the lower thermosphere. The heights of the lowest layers, the troposphere and the tropopause, vary significantly with geographical latitude and with season. The lower boundary of the stratosphere reflects this variation but its upper boundary varies little.

Depending on latitude and season, the troposphere extends from Earth's surface to altitudes of about six kilometers at the poles and to roughly eighteen kilometers in the tropics. Meteorologists differentiate an additional boundary layer between Earth's surface and the troposphere, the so-called planetary or atmospheric boundary layer. In this layer, which can be between a few hundred meters and a couple of kilometers thick, the behavior of the atmosphere is directly influenced by contact with Earth's surface. That clearly is important to predict short-term changes in weather patterns such as wind directions and strength, moisture, and temperature. In the troposphere, air temperature decreases with altitude from Earth's surface temperature to just below -60 °C. As air at lower temperature is denser, this results in heavier air sitting on less dense, warmer air. It is the instabilities resulting from such inversions that are behind the air movements that produce the phenomenon we call weather; weather is almost exclusively a phenomenon of the troposphere. The troposphere contains approximately 75% of the atmosphere's mass and about 99% of the total mass of water vapor and aerosols.

Separated by the tropopause, the stratosphere sits on top of the troposphere extending upwards to fifty kilometers. It is a much more stable layer than the troposphere because of its temperature profile. The average temperature in the stratosphere is just below -60 °C near the troposphere and about -15 °C near the top of the stratosphere. This temperature inversion results in warmer air sitting on top of colder air and therefore inhibits vertical air currents. It is this behavior, which makes the stratosphere a much more stable layer than the troposphere. Incidentally, avoiding the turbulence of the troposphere is one of the reasons jet aircraft fly in the high troposphere or the lower stratosphere.

Importantly, the stratosphere contains Earth's ozone layer. The thickness of this ozone layer, which sits between twenty and thirty kilometers above Earth's surface varies seasonally and geographically. The ozone layer derives its name from its high concentration of ozone, a molecule made of three oxygen atoms that can absorb ultraviolet solar radiation harmful to most organisms. Because of that, the formation of this ozone layer has been critical for the evolution of life on Earth. Without it, life on land would not have evolved as we know it and maybe, if the ozone layer had not developed, life would never have moved onto land where it would have been exposed to the ultraviolet radiation from the Sun. We can only wonder if intelligent lifeforms such as us would have evolved if life had remained constrained to the oceans. Interestingly, the ozone layer also serves another important function as the heat released in the process forming the ozone layer produces the temperature inversion in the stratosphere. Without this temperature inversion of the stratosphere, we most likely would see much more severe weather patterns than we do. This is because our weather then would not just be determined by conditions in the troposphere, a layer of between six to eighteen kilometers thickness but by the combined troposphere and stratosphere, with an effective thickness of up to fifty kilometers.

The next layer upwards in the atmosphere is the mesosphere layer, which sits on top of the thin stratopause boundary layer, separating it from the stratosphere. The mesosphere layer starting just above a fifty-kilometer altitude extends out to about eighty-five kilometers from Earth's surface. Throughout the mesosphere air temperatures decrease again with increasing altitudes up to the mesopause boundary layer where this trend stops and is reversed in the next layer, the thermosphere. The upper part of the mesosphere is the coldest part of Earth's atmosphere reaching temperatures well below -100 °C.

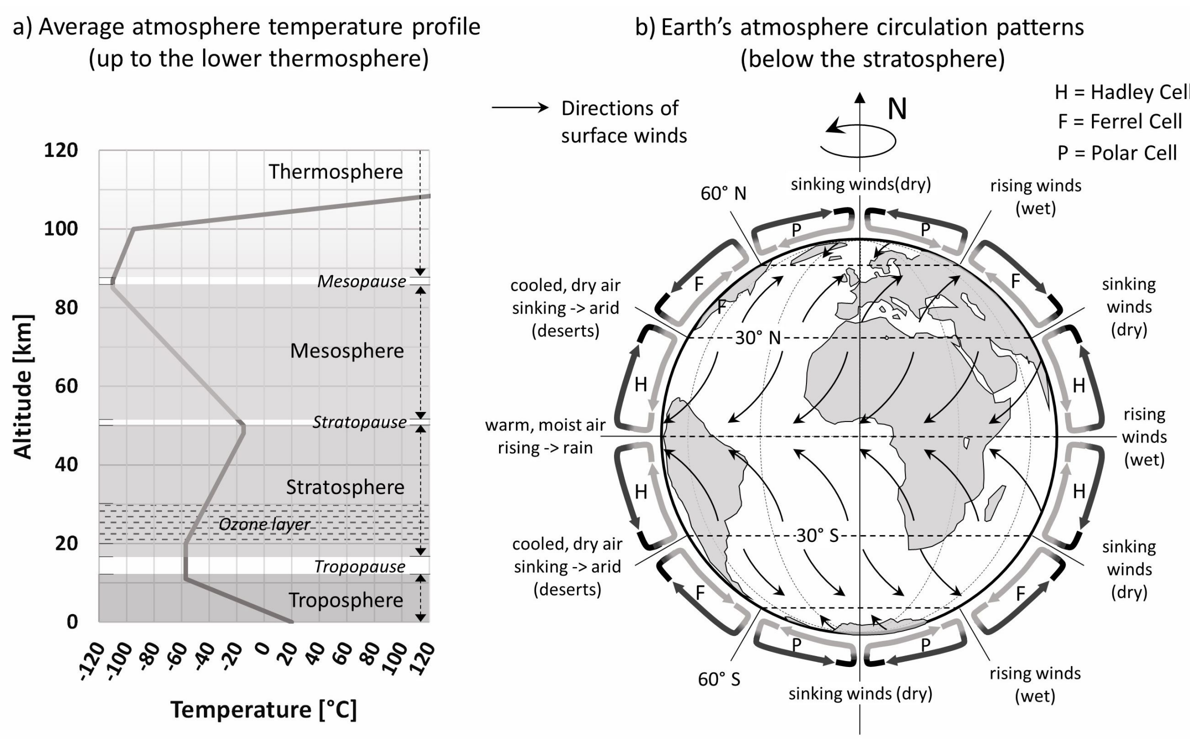


Figure 2.9: The average temperature profile of Earth's atmosphere up to the lower thermosphere (a) and a schematic of Earth's atmosphere circulation patterns below the stratosphere (b). The layer thickness for the troposphere and the tropopause varies significantly with geographical altitude and with season but the overall profile shape remains similar.

It is in the upper mesosphere and the lower thermosphere between around seventy-five and one hundred kilometers above Earth's surface where meteors burn and we see them as shooting stars high up in the atmosphere.[55] The thermosphere, separated from the mesosphere by the thin mesopause layer, extends to around six hundred kilometers above Earth's surface. Temperatures in the thermosphere highly depend on solar activity with absorption of high-energy radiation from the Sun resulting in temperatures reaching in excess of 2,500 °C in parts of the thermosphere during daytime. However, this temperature is not the air temperature one would measure in this part of the atmosphere because the atmosphere at this altitude is so thin that its gases are too diluted to conduct heat; temperatures at such altitudes reflect the energetic states of molecules and atoms and not the temperature of an atmosphere we could feel. It is in the upper part of the thermosphere, beginning at around 500 kilometers, where we would start to run into satellites as low orbit Earth satellites are typically placed in a range between 500 to 2,000 kilometers above Earth's surface. The high-energy radiation in this part of the atmosphere is also responsible for producing the electrically charged layers, starting at the bottom of the thermosphere, which mark the lower boundary of the ionosphere.

Finally, above the thermosphere, separated by the thermopause, we find the exosphere, which is the outermost layer of Earth's atmosphere and extends to about 10,000 kilometers above Earth. In the exosphere, molecules and atoms are still bound to Earth by

gravity but their density has now become even lower and collisions between them are too infrequent to resemble a gas in any way. At theses heights atoms and molecules follow their individual trajectories just as satellites do.
The ionosphere is the electrically conducting part of Earth's upper atmosphere shaped by ionizing radiation coming from the Sun. The ionosphere starts in the thermosphere and extends out to about 1,000 kilometers above Earth's surface into the exosphere. The ionosphere is not a uniform shell but is itself decomposed into a number of layers with varying charge densities characterized by different ionization processes and recombination rates. The ionosphere plays an important part in atmospheric electricity and forms the inner edge of the magnetosphere created by Earth's magnetic field. It is in this part of the ionosphere, overlapping with the thermosphere and the lower part of the exosphere, where we can observe the beautiful phenomenon of the polar lights. Polar lights are the result of charged particles created at high altitudes, which have to follow trajectories prescribed by Earth's magnetic field towards Earth's magnetic poles, forcing them into collisions with the gas molecules of the upper Earth atmosphere.
The basic description of Earth's atmosphere properties given here may leave the impression of Earth's atmosphere system being mostly static. Nothing could be further from the truth; Earth's atmosphere is a very dynamic system. As already mentioned earlier, the atmosphere layers are not constant, but their thicknesses and temperature ranges vary with geographical latitude and season. We know that for example the temperature profile of the lower atmosphere can locally invert, resulting in warmer air sitting on cooler air. A large and growing number of cities and industrial regions experience this atmospheric phenomenon with increasing frequency. We call it smog and it can linger for days or even weeks before weather conditions eventually help reverse the situation. In principle, such weather inversion situations are normal and to be expected. However, what is not normal is that we are treating our atmosphere as a dumping ground for all kinds of pollutants and turn those naturally occurring weather situations into something much worse. We cannot do much about the weather, but we can stop polluting the air that we need to breathe to stay alive. If we could do that, such weather inversions would not be able to inflict the harm they do, including several millions of premature deaths every year. The ozone layer of Earth's atmosphere, embedded in the lower part of the stratosphere layer, is critically important for life on Earth as it provides protection from harmful ultraviolet radiation. It is also a quite fragile part of the atmosphere as we ourselves have demonstrated towards the end of the twentieth century when we managed to thin the ozone layer through our use of CFC's, which we started in the 1930's. The danger to life on Earth from a depleted or punctured ozone layer resulted in a drastic reduction of CFC emissions starting in the 1990's. But given that CFC's can stay in the atmosphere for up to roughly one hundred years it will take some time until the ozone layer will have fully recovered. Variations in thickness and properties of the upper atmosphere layers such as the thermosphere are due to changes in solar radiation. In addition, because of the charged particles in the ionosphere, starting in the lower thermosphere, changes in solar-driven geomagnetic activity are also an important factor.
The circulation of air within the atmosphere is a major factor in conditioning Earth's climate. Earth's lower atmosphere has several circulation zones as shown in fig. 2.9b that extend from either side of the equator to the poles. The first ones extending from the

equator to approximately 30° S and 30° N are known as Hadley circulation cells. They are named after the English physicist and meteorologist George Hadley (1685 – 1768) who first formulated an accurate theory describing the trade winds, also illustrated in fig. 2.9b, and the associated circulation patterns, the Hadley cells. To a seafaring nation such as the British, understanding the trade winds was of course important and Hadley gave the correct explanation in 1735, although this would not be recognized until almost sixty years later. Because of Earth's rotation, the surface wind patterns driven by the Hadley cells, called the trade winds for a good reason, are not directed straight towards the equator, parallel to the lines of longitude, but blow towards the equator from the northeast in the northern hemisphere and from the southeast in the southern hemisphere (see fig. 2.9b). A second set of circulation cells extends from 30° N to 60° N and 30° S to 60° S. They are known as Ferrel circulation cells named after the American meteorologist William Ferrel (1817 – 1891). They were proposed by Ferrel in 1856 to explain the prevailing surface wind patterns at mid-latitudes, the westerly winds, often just called westerlies, observed in both hemispheres between latitudes of 35° - 60°. Because of Earth's rotation they blow towards the Polar Regions from southwest in the northern hemisphere and from northwest in the southern hemisphere (see fig. 2.9b). Airflow in the Ferrel cells is reversed from the airflow observed in the Hadley cells. And it is also the reverse of the airflow observed in the polar cells that extend from 60° N and 60° S to the respective poles. In the Hadley circulation cells, warmer and moister air rises at the equator, loses moisture as it does so, and then sinks as colder and drier air at the higher latitude end of the cells. At the same time, closer to the ground replacement air flows back in these cells into the opposite directions completing the atmospheric circulation. A similar mechanism works in the polar cells. Unlike the Hadley and polar cells, Ferrel cells do not have an inherent mechanism powering their air circulation. Rather, they derive much of their energy from the circulation of their neighboring Hadley and polar cells, the very reason why airflow in Ferrel cells is the reverse of Hadley and polar cells. Because of that, climatologists consider Ferrel cells secondary circulation cells.

The combined effect of the three cell types moves warm and moist air from low latitudes to high latitudes with cooler and drier surface air flowing towards the equator. As indicated in fig. 2.9b, this is an important factor in defining Earth's climates. Arid climate conditions prevail where the cool and dry air descends at latitudes around 30° N and 30° S; it is there where Earth's desert areas can mostly be found. Around the equator and around 60° N and 60° S, wet and moist conditions prevail; this is where we find tropical climates girding the equator and temperate and wet climates around 60° N and 60° S. Of course, those conditions are modulated by local geographies, including closeness to coastal areas or other large bodies of water, mountains, vegetation, and of course relative latitude. Air movements in the three types of circulations cells are contained within the troposphere, with air in the Hadley cells at the equator rising up to no more than around fifteen kilometer and to lower altitudes in the Ferrel and polar cells. Of course, traders had known well before Hadley and Ferrel provided the correct explanations that they needed to sail to the New World on a southeastern route using the trade winds and come back a northern route pushed by the westerlies.

The oceans are critical components of Earth's climate system. In many ways they serve as Earth's climate buffers. Compared to land areas, the oceans warm more slowly during

summertime and cool more slowly during winter. This is due to water having a tremendous capacity for storing heat, one of the highest of any common substance we know. This is not only critical for life in the oceans and in our lakes and ponds but also for Earth's climate system. It takes more than ten times the energy to increase the temperature of one liter of water by 1 °C than it takes to heat the same mass of copper, one kilogram, by 1 °C. The heat capacity of a substance also determines its ability to transfer heat, something we painfully experience when we are not careful in draining our spaghetti and hot water vapor condenses on our skin, burning it by releasing its stored heat. Earth's oceans cover today 71% of its surface area and contain about 97% of its surface water. The distribution of landmasses and oceans over Earth's surface is not evenly balanced. In the southern hemisphere, oceans dominate the latitudes between 30°and 70° S where they make up between more than 80% and close to 100% of Earth's surface depending on latitude; only farther south above 70° S do land areas dominate. In the northern hemisphere oceans cover between 30 - 50% in latitudes between 50°and 70° N only exceeding 50% and rising towards 100% as one moves further north from 70° N. The high heat capacity of water and the enormous masses of water contained in our oceans allows them to store vast amounts of heat without a large increase in temperature. Because of that, Earth's oceans play a central role in stabilizing Earth's climate system.
We can observe our ocean's stabilizing functions so-to-say in a life stream when we play back the record of the past several decades. As the rising concentration of atmospheric greenhouse gases over that period has increasingly prevented heat radiated from Earth's surface from escaping into space, most of that excess heat has been stored in the oceans, more specifically in the first few hundred meters of surface water. The result is that the respective upper ocean heat content has increased significantly over the past decades. In essence, most of the additional heat accumulated in Earth's climate system due to global warming is stored away for now in the oceans. It is estimated that warming of the upper oceans accounts for a little less than two third of the total increase in the amount of stored heat in Earth's climate system over the past fifty years and warming of the deeper ocean, from 700 meters down, adds just below one third. Altogether, during that time, Earth's oceans have stored away more than ninety percent of the additional heat accumulated due to global warming. Oceans do not heat up uniformly. They receive about twice the amount of solar radiation along the equator as compared to the poles. Oceans and the atmosphere interact in a complex way. The ocean surfaces anchor the humidity and temperature of the air above it, thereby creating air flows in between the extremes of rising hot and humid air and sinking cool and dry air. This process and Earth's rotation produce much of the weather systems whose winds in turn push the ocean surface waters and thereby the ocean current patterns in the upper one hundred meters of the ocean. At the same time, differences in temperature and salinity drive deep ocean currents as well as surface currents, scientists refer to them as thermohaline currents.[56]
The global system of ocean currents resulting from the deep thermohaline currents and the thermohaline or wind-driven surface currents sometimes is referred to as the oceans conveyor belt system (see fig. 2.10). There is no specific beginning or end for our oceans conveyor belt so we will just start looking at it in the Norwegian Sea, where warm surface water coming from the Gulf Stream heats the atmosphere. The Gulf Stream is a surface current mostly driven by wind and to a lesser extent by salinity differences.

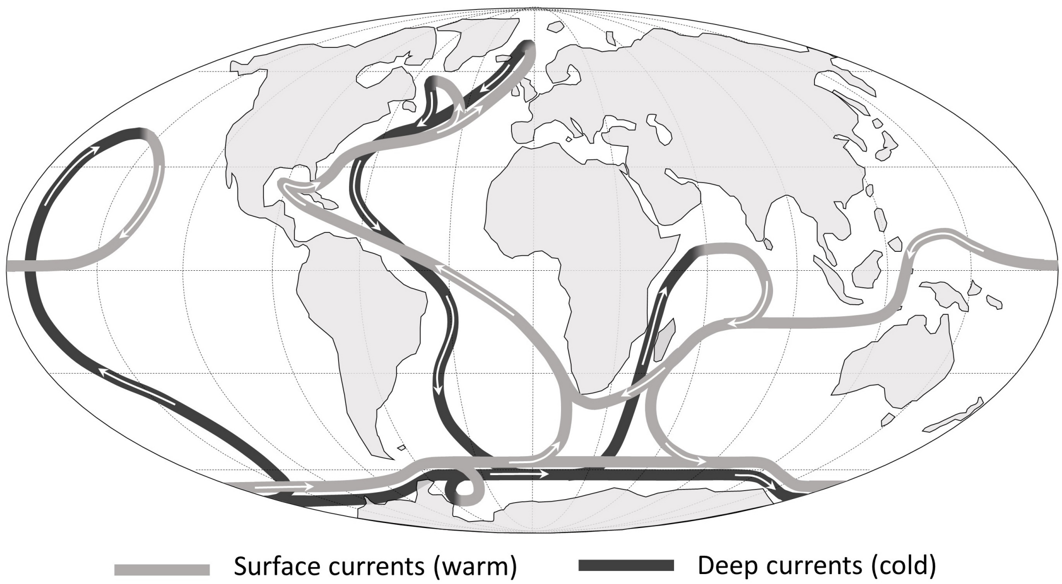


Figure 2.10: The currents of Earth's ocean circulation system adapted from Refs. [21,22]. Earth's oceans conveyor belt system is driven by differences in water temperatures and salinity, the thermohaline engine, and by wind forces.

The waters of this warm ocean current flowing north cool and as the density of this cooling water increases it sinks down in the water column towards the bottom of the sea. As the cooler water sinks, the Gulf Stream continues to bring more warm surface waters to the north for which the cooler waters moving towards the bottom make room by flowing back south as a cold deep-water current. This deep cold-water current then keeps flowing past the equator all the way to Antarctica, encircling it, and branching off into the Pacific and Indian Oceans. Of course, somewhere the deep-water currents have to connect back to the surface currents. In locations where this happens, ocean surface waters are pushed aside by the prevalent wind patterns making room for water rising up from beneath to replace it. The respective process is called upwelling and occurs in open ocean waters and along coastlines.[57] This mixing process can take place regardless of water temperature or salinity as it is not thermohaline driven but is a replacement of surface waters being transported away. The replacement water rising from the deep ocean is cold, less salty, and typically very rich in nutrients. This is why the fishing grounds around Newfoundland for example have been so productive; the same applies to an upwelling area producing the nutrient rich waters in the Antarctic. Much is still to be learned about these upwelling processes but they are difficult to study as they happen over long time scales. This is because the oceans conveyor belt while moving enormous amounts of water does so very slowly.

The ocean current patterns resulting from the thermohaline and wind driven processes are determined by the distribution of landmasses and temperature differences. The ocean currents transport heat from the equator to the poles where it much more readily escapes into the atmosphere than at the equator, providing positive feedback for the currents of warm waters flowing to the poles. There likely will be profound consequences for Earth's climate system if our ocean conveyor belt system should ever change in a major way; or even in a minor way. We already can see today what seemingly marginal increases of ocean temperatures can do. The impact of global warming is not only evident in rising sea levels driven by melting glaciers and increasing water temperatures. It also clearly manifests itself in atmospheric changes caused by warmer oceans such as increased frequency and ferocity of hurricanes. For communities that happen to be in the path of these destructive natural forces this has already devastating consequences today. But for humanity overall the associated negative consequences are still limited. However, this could change dramatically if continued global warming results in disruptions of today's ocean conveyor belt system. Earth's ocean circulation system works very slowly and changes to it from continued global warming are likely to develop on the time scale of many decades and longer. However, eventually the accumulated heat will be redistributed in the oceans and with more and more warm waters flowing to Polar Regions this will have obvious consequences for the polar ice sheets. As the ice sheets melt, the salinity of the oceans will change.

With salinity and temperatures changing, the fundamental constellations driving today's thermohaline current patterns are very likely to change too. The question is only be by how much. If such a change for example disrupts the Gulf Stream, much of Western Europe would experience colder and not warmer climates, despite overall global warming. Global warming is a complex phenomenon and there will be climate change losers and winners. Some of those on the losing side include populations on low-lying islands, which slowly vanish beneath the rising oceans. Those people losing their homes and the lands where countless generations of their ancestors lived, what have they done to deserve such a fate? They did not pollute our atmosphere with the greenhouse gases producing this rising ocean. The ones first to suffer from the consequences of climate change are the ones that contributed the least to global warming. With the changes we inflicted on Earth through our industrial activities over the roughly past two hundred years, we have already committed Earth to the additional warming this will produce. However, the delay provided by the ocean climate buffer gives our civilizations hopefully more time to prepare for the severe impacts this will have on coastal and island communities for sure but also on the future of human civilization itself.

Sea levels will not only rise as the Arctic and Antarctic glaciers melt, but they will also rise because of the temperature increase of the ocean. Water has its highest density at around four degrees Celsius and as it warms up from there the volume of water will expand. Warmer oceans therefore translate into higher sea levels. Sea levels have been rising in modern times and are now on average higher by about twenty centimeters than a century ago. That may not seem much, but it is. Until humans started to change Earth's climate conditions beginning with the industrial age, sea levels had varied very little for several thousand years. Twenty centimeters over roughly one hundred years is a small increase when compared to the future sea level rise that has already been dialed-in but not materialized yet; things will definitely get worse.

Currently, sea levels are rising on average by about three millimeters each year. By the end of this century, ocean levels are expected to be between $\sim$0.5 to $\sim$1.2 meters higher than they were in 2000. However, if global greenhouse gas emissions continue to rise, it is unlikely that global warming can be contained to less than 2 °C above pre-industrial levels as many nations agreed to in Paris at the end of 2015. In that case, the outlook changes. But why should it matter much if global temperatures rise by an additional degree until 2100? Well, the answer is that a one-degree increase in global temperatures is estimated to produce a more than two-meter rise in sea levels. Most of humankind lives in coastal areas and many of our most fertile crop regions are in low-lying areas such as river deltas. What will a couple of meter sea level rise do to the billions of people living practically at sea level and what will it do to worldwide food security once many of the now highest producing agricultural areas are lost? Importantly, it is not just the indirect impact of the oceans on our civilizations due to rising sea levels that we should be concerned about. Earth's oceans as a major food resource for much of humanity are at risk. The accelerated degradation of marine ecosystems, already underway and likely to get even worse, is bound to severely diminish the bounty of seafood humanity has come to rely on harvesting from the oceans.

Earth's oceans are not only a heat buffer but also a carbon dioxide buffer. As everyone likely knows from experience, carbon dioxide dissolved in bottled carbonated drinks will bubble if the pressure is lowered by opening the bottle; importantly this effect is stronger if the temperature of the liquid in the bottle has increased, as for example when one opens a bottle of lukewarm sparkling water. This is because water can store more gas at lower temperature or at higher pressure. The same is true of the oceans and their ability to dissolve oxygen or carbon dioxide. Let us look first at carbon dioxide. Our oceans hold an enormous amount of carbon dioxide. Scientists estimate that about fifty times the amount of carbon dioxide contained in the atmosphere is sequestered away in the oceans. Earth's oceans can sequester so much of it because the bulk of carbon dioxide in our oceans is not stored in gaseous form. Carbon dioxide reacts with water to form carbonic acid. This keeps the amount of dissolved gaseous carbon dioxide low so that the oceans can continue to take up additional carbon dioxide. However, this comes at a price and that price is the acidification of the oceans due to the carbonic acid that is the root cause of the demise of many once thriving coral reefs. Also, as the oceans continue to warm, their ability to dissolve carbon dioxide is going to decrease until eventually surface waters are saturated with carbon dioxide. This will happen if oceans warm to a degree that they stop supporting the currents that now transport carbon dioxide rich waters to colder polar areas that still have capacity to store it. When and to which extent this happens is a critical parameter in climate models. However, aside from our ocean's ultimate carbon dioxide storage capacity and the associated ecological consequences, it must be understood that the oceans only act as a buffer not as a final storage. The carbon dioxide exchange between the ocean surface and the atmosphere is a balanced one, ensuring that the carbon dioxide sequestered in the oceans eventually will be released back into the atmosphere once the carbon dioxide gradient between ocean surface waters and the atmosphere favors this; like it does when upwelling cold water is already so saturated with carbon dioxide that it vents into the atmosphere.

Another predicted major impact of continued global warming is the suffocation of Earth's oceans, the depletion of oxygen in our oceans. It took many hundreds of millions of years in Earth's evolution to oxygenate the oceans. This has been the prerequisite for the evolution of the first primitive non-bacterial lifeforms in Earth's oceans. Our oceans ability to circulate oxygen from surface waters to the less oxygen rich layers of the deep sea depends on the capability of ocean water to dissolve oxygen and eventually transport it to deeper waters. The effectiveness of both very much depends on water temperatures. The way oxygen is delivered to life in the deep of the sea is via cold oxygen rich water in Polar Regions sinking towards the bottom of the sea. With polar waters warming as ocean currents transport increasingly warmer ocean waters to the poles, polar waters can retain less oxygen. Warm water is also less dense than colder water. Because of that, these already less oxygenated warm polar waters will also be less efficient in oxygenating deep-sea waters. The results are increasingly oxygen-starved parts of the deeper ocean that at some point will not continue to support life. The spread of such oceanic dead zones because of oxygen depletion is in addition to the impact of coral reef destruction due to ocean acidification. It may even surpass the severity of the latter, as with the exception of those sulfur-loving lifeforms around black smokers, life in the oceans just cannot exist without oxygen. We may be well on our path to destroy the oceanic biospheres before even having come close to understanding the extent and complexity of life in our oceans.

Life, Earth's biosphere, is an integral part of Earth's climate system. Life created Earth's oxygen rich atmosphere in the first place, and largely it is plant life, which maintains the composition of the atmosphere by inhaling carbon dioxide and exhaling oxygen. Throughout the history of life on Earth, its biosphere has helped condition Earth's climate. And it has changed with and in reaction to Earth's climate changes. As discussed above, much of the carbon dioxide that we have pumped into the atmosphere since the industrial revolution has been sequestered in the oceans. Another good part has been absorbed into the biosphere, incorporated into the living biomass. However, we may be pushing the limits of what the biosphere can help us compensate for, specifically so as we continue to deforest large areas on our planet and continue to bury large parts of Earth's surface under the concrete of our civilizations. On a geological time scale, the capability of life to remove carbon is impressive. Some of this calcium carbonate, accumulated over millennia by the deposition of countless dead organisms on the sea floors, has been folded up by geological forces into the mountain landscapes we enjoy today, including for example the European Alps. However, increasingly acidic oceans reduce the ability of organism to sustain calcium carbonate shells or skeletons. And that in turn reduces the amount of carbon dioxide that can be sequestered in this way from surface waters, eventually to be deposited onto the sea floors when these organisms die. The result is predictable. As life in the ocean removes less carbon dioxide from surface waters, the oceans can take up less carbon dioxide from the atmosphere.

One more positive feedback of the biosphere adding to global warming is the warming of soils on land. It triggers increased growth and respiration rates of microorganisms, which convert carbon from dead matter buried in soils into carbon dioxide. An often-cited example of a negative feedback that could potentially help reduce the impact of climate change is an increased uptake of carbon dioxide. Plants need carbon dioxide for photosynthesis to meet their energy needs and an increased amount of carbon dioxide in the air will have a major impact on plant life. We know that plants do react to changing amounts of carbon dioxide. If there is less or more carbon dioxide in the atmosphere, plants need more or less pores in their leaves respectively to bring in carbon dioxide. Incidentally, this is one example for how paleontologists have been able to read the record of the relative carbon dioxide amounts prevalent in Earth's past. To a lesser degree, in addition to pore numbers, the size of the pores also changes with the atmosphere's carbon dioxide concentration. In reaction to the roughly 25% increase in carbon dioxide levels observed over the last ∼150 years, some plants have already reduced density of leaf pores by about one third. Using the analogy of crops grown in greenhouse conditions producing higher yields, it is sometimes argued that higher atmospheric carbon dioxide levels should benefit plant growth. The argument essentially asserts that an increase in carbon dioxide would result in a greening of Earth. However, this is an oversimplification as it extrapolates greenhouse conditions in a way that cannot be sustained in nature and does not consider other negative impacts from climate change that are likely to limit plant growth.

On a geological time scale, the arrangement of Earth's land surfaces has had a profound influence on climate. As much of the landmass on Earth has several times been concentrated along the equator into supercontinent formations, ocean-atmosphere dynamics, ocean currents and the prevailing climates dominating the landmasses would have been very different from today. Already Charles Lyell suspected that a concentration of landmasses in polar areas should cool Earth's climate. Similarly, we know today that a concentration of landmasses in the equatorial region tends to produce very warm climates but with the downside of large internal continental areas likely to be arid and inhospitable. The movement of continents clearly does not play a role in the relatively short-term climate change that human industrial civilization has induced. However, our use of the land contributes to climate change and the resulting global warming is likely to have a significant impact on the future use we can expect to enjoy from the land. Earth's land surfaces make up less than one third of its total surface, the rest being covered by water, but they account for about two thirds of the total solar radiation absorbed by Earth; the remaining one third being absorbed by Earth's oceans. The heat capacity of Earth's rock surfaces is only about one fifth of water, so land surfaces heat up much more readily and cool much faster than Earth's ocean covered surfaces. Because Earth's surfaces account for such a large fraction of the solar radiation absorbed by Earth, variations in the surface albedo of Earth's land surface have a significant impact on the dynamics of Earth's atmosphere and on prevailing climates. From the poles to the equator, Earth is divided into several major climate zones: the polar and subpolar zone; the temperate zone; the subtropical zone; and the tropical zone. Because of how the landmasses are distributed on the two hemispheres, temperate zones are predominately found in the northern hemisphere, subtropical zones are more balanced but still prevalent in the northern hemisphere while the landmasses with tropical and polar and subpolar climate zone are distributed more evenly.[58]

The development of modern human civilizations was intricately linked to temperate and subtropical climate zones and to a much lesser extent to tropical zones. The impact of climate change on agriculture in these areas will vary with geography but to a greater extent it will depend on how successful any human efforts to mitigate the consequences of climate change can be. If human societies just continue with business as usual, many of our today most fertile agricultural areas will see a significant impact. As always, there will be winners and losers. Winners could include the parts of Earth where climate change will locally improve conditions such that farming may become possible or more productive. Losers will include some of the fertile agricultural areas where the land may become much less productive or may eventually become unusable for agriculture; or may be just lost to the rising oceans. On balance, if nothing is done to stop making things worse, it will be very difficult to maintain the amount of fertile agricultural land under cultivation today. And in the longer-term, vast populations that today live in coastal areas will have to relocate. Some unfortunate island populations are already seeing their lands being swallowed by the sea knowing that in the near future the places where their ancestors had lived for many generations will be gone for good. However, it is not only agricultural or coastal areas and islands about which we must be concerned. Global warming can affect vegetation patterns in ways that could reduce Earth's green surface, Earth's lungs so to speak, to an extent where their vital role in Earth's carbon dioxide cycle is compromised beyond recovery.

This carbon dioxide cycle is quite a bit more complex than we thought. For example, when we refer to the Amazon basin as being a part of Earth's oxygen producing "lungs", this is not quite correct, even though the net result is just that. Fact is that as much oxygen as the Amazon rain forest produces the life it hosts practically consumes all of it. From what we understand today, much of the fertility of the Amazon basin derives from desert dust that reaches it from the deserts of North Africa. We only discovered this phenomenon once we could see it from the proper vantage point, which is from outer space. This Sahara Desert dust carries with it a substantial amount of phosphorous, a legacy of large inland seas once covering much of northern Africa. It is this phosphorous carrying dust that provides the fertilizer for the immense richness of Amazonian plant life and that in turn provides the basis for the unequaled biodiversity of Amazon's animal life. One of the most desolate areas on Earth, North Africa's Sahara, fertilizes one of the most biodiverse, if not the most biodiverse area on our planet. Scientists estimate that of the almost two hundred million tons of dust that leave Africa across the Atlantic, nearly thirty million tons fall on the area of the Amazon basin each year.

As pointed out above, Amazon's rich animal life practically consumes all of the oxygen its plant life produces. So how then does the Amazon basin function as one of Earth's lungs? It does so indirectly. Trees suck water up from the ground and transport it to their leaves in the forest canopy where a good amount of it evaporates. This happens on a massive scale creating what scientists refer to as a river in the clouds, flowing in the opposite direction to the Amazon River on the ground. Amazon's invisible river in the clouds transports enormous masses of water from east to west, towards the foothills of the Andes; if it were a real river, it would be the largest river on Earth by far. Once this river in the sky reaches the Andes, the clouds rise and the water they carry starts to precipitate, feeding all the tributaries of the Amazon. As this water washes down into the Amazon tributaries, it carries with it the richness of the Andean soils. Eventually this continuous stream of nutrients carried by the Amazon River reaches the Atlantic Ocean where is supports countless microorganisms including trillions of diatoms, single-celled photosynthesizing algae, who use it to build their silica shells.

It is these diatoms and not the Amazon rain forest itself which produce much of the oxygen we breathe. Estimates as to how much of our oxygen all the diatoms in our oceans produce come in mostly around twenty percent although some are quite a bit higher. In addition to producing a substantial amount of our oxygen, diatoms also do sequester enormous amounts of carbon dioxide. When diatoms die, they sink to the bottom of our seas where over time they build up enormous layers of silica. Layers like those of the former inland seabed in what is today the Sahara Desert, where winds picks up the pulverized remnants of diatoms living millions of years ago and blow them across the ocean to fertilize the Amazonian basin. Saharan dust of ancient sea floors built from diatom remnants fertilizes the Amazon basin, which in turn provides diatoms the means to grow shells to build new sea floors made of diatoms. It is an amazing cycle, producing much of our oxygen and sequestering enormous amounts of carbon dioxide, and it has been going on for a very long time.

There are other similar cycles, equally intricate and just as important to maintain the detailed balance required to sustain higher life on Earth. A balance that humans have been disturbing for quite a while but only recently to the extent that it may be threatening our long-term future on this planet. Not only are we poisoning the air we breathe and changing Earth climate, but eventually we may also run short of water and that on our blue planet Earth. Water shortages for human consumption and for irrigation purposes could be another dire consequence of continued global warming. As precipitation patterns change, water is likely to become even more scarce in places we may need it most, like our most fertile areas, and may become more than plenty in places where it is of little use to us or may even become a major problem as we can witness with increased and more severe flooding events. Water is already a very limited resource in many places today. Resource shortages of water or arable land, or land just vanishing below the rising oceans can put severe strains on human societies. Can we have any confidence that our civilizations will be able to address these challenges without major conflicts, given our troubled conflict-laden past?

The regions of Earth covered in ice sheets, glaciers, sea ice, soils remaining permanently frozen year-round, or areas covered in snow are part of what scientists refer to as the cryosphere. Only about 8% of the Earth surface are today permanently covered by snow or ice but seasonal snowfalls can easily double this area. Because of the relatively smaller area permanently covered in ice and the large landmass that is covered by seasonal snowfall, the seasonality of the cryosphere is much more pronounced in the northern hemisphere. Today, only about 4% of the northern hemisphere is permanently covered in ice; 3% of which comes from the ice-covered Arctic Ocean and 1% of which comes from the Greenland ice sheet. Seasonal snowfall increases this 4% cryosphere area to about 24% during northern hemisphere winters. The situation is quite different in the southern hemisphere where most of the cryosphere is made up by the Antarctic ice sheets and seasonal variations are mostly due to increases and decreases of sea-ice around them. Overall, during southern hemisphere winters the cryosphere makes up about 13% of the southern hemispheres surface, only a little more than half of the extent of the cryosphere in the northern hemisphere during winter. As we know from the discussion of ice ages, the large seasonal component of the cryosphere only translates into an increase of its permanent components over a long time period; provided the necessary solar forcing prevails as conditioned by Earth's orbital configuration. Its permanent components therefore react only very slowly to climate changes. However, they do so now at a quickening pace as we can clearly see when we compare the extensions of mountain glaciers around the world as documented over the past roughly 150 years or so. Almost all of them have retreated to a fraction of their former size and some of them are gone for good. We can follow the accelerating reduction of the Greenland and Antarctic ice sheets now in real-time.

Ice sheets are made up of glaciers, which are formed over many years from accumulated snow that gets compacted as the pressure of the snow above compresses the older snow layers below. Masses of glacier ice covering in excess of 50,000 square kilometers are called ice sheets. So technically, there are only two ice sheet areas, which however make up a large part of the cryosphere, the ice sheets in Antarctica and in Greenland. Recalling the discussion of ice ages and the solar forcing that triggers them it should come as no surprise that these large ice sheets form at the poles and expand from there. In the past, the ice sheets anchored at both of Earth's poles extended much farther to the equator. The glaciers that built up at lower latitudes during the glacial periods of the ice ages in high altitudes reached extensions that moved them from being just glaciers to becoming ice sheets as for example happened in the European Alps and North America.

Long before the arrival of humans, Earth's climates were quite different from anything humans ever experienced throughout their species evolution. Several times throughout Earth history, much of the continental landmass was concentrated around the equator with little to no landmasses located in Earth's Polar Regions. For snow to accumulate and eventually to build up to ice sheets there has to be of course first solid ground for snow to fall on; unless Earth goes through a much more extreme cold period where ocean surfaces start to freeze over, which has also happened in Earth's past. So we are looking today at a cryosphere that may be typical for our period, the Quaternary, but was quite different in the past.

Sometimes, the argument is put forward that there was life on Earth during very different climate periods, much warmer and much colder than today. Therefore, the argument goes, we should not be too concerned with climate change. The premises of the argument is of course correct, but the conclusions drawn do not hold up. If we were still cave-dwelling or savanna-roaming early humans maybe there would be something to the argument as our pre-historical ancestors have persevered; at least one human species has and that is us. During the last 10,000 years, as humans emerged from pre-history, started agriculture and created the first human civilizations, Earth's climate variations were, when compared to the long-term Earth climate variations, rather benign. There were few humans around then to start with and they always could move if local conditions changed such that their environment could not sustain them anymore. In our multi-billion and exceedingly complex human civilizations we are much more vulnerable to climate change. Our modern civilizations certainly have many powerful technologies that our ancestors could not even have imagined. However, it would be the pinnacle of hubris for us to believe that we can control the impacts of severe climate change without putting the very fabric of our modern civilizations at risk. Climate conditions have indeed been very different in the past but we only have seen our human civilizations thrive within a very narrow band of climate conditions. There certainly is no guarantee that human civilizations can thrive under more extreme climate conditions, they may or they may not. Therefore, we should be quite concerned about Earth's shrinking cryosphere, the melting polar ice sheets and the melting glaciers.

The long-term threat of rising oceans is very real, and the potential consequences are dire. Many of our civilizations most populous cities and much of our most fertile lands are only slightly above sea level. They will all be lost if the polar ice sheets should ever melt. A twelve-meter sea level rise can be expected if the Greenland and West Antarctic ice sheets melt away. If eventually the East Antarctic ice sheet would also melt completely, we can expect oceans to rise by an additional sixty meters. All of this could come with an alteration of sea currents, even a complete disruption of the oceans conveyor belt system, which may change climate conditions locally in much more dramatic ways. Regions that today are conducive to agriculture could become arid while other areas could see a lot more precipitation. As today's permafrost areas thaw, new land may become available to agriculture. However, these frozen soils store enormous amounts of carbon in the form of organic matter that has not decayed since it captured carbon dioxide from the atmosphere a long time ago. Scientists estimate that permafrost soils worldwide contain about twice the amount of carbon dioxide that is currently in the atmosphere. It is not only additional carbon dioxide added to the atmosphere, which are of concern when permafrost soils melt. As its organic matter thaws, microbes will convert some of the carbon dioxide into methane, which is a much more potent greenhouse gas. Recent discoveries indicate that this process converting carbon dioxide captured in frozen plant material into methane may happen much faster than we had assumed.

Current climate models only account for a gradual thawing of the permafrost soil where the upper layer of the soil that thaws during summer season becomes gradually thicker. This process activates microbes that decompose the buried organic matter and thereby releases carbon dioxide and methane into the atmosphere. Because at the same time the thawed ground also can support vegetation, we can expect some of the carbon dioxide to be recaptured through photosynthesis by this added vegetation. While this is not the case for the methane being produced, the added vegetation still limits and slows the expected overall contribution to global warming from released carbon dioxide. However, a new process has been observed that is driven by melting of ice deep in the soil. Unlike the gradual seasonal thaw, this process called flash thawing very much accelerates the release of ancient carbon dioxide into the atmosphere.

All parts of the cryosphere are shrinking, and they shrink at an accelerating rate. The question is not if this is happening but how far it will likely go. How much of a reduction in Earth's cryosphere is already baked in, meaning, we cannot really change it anymore? What can we do to stop making things worse? While we do not have full answers to those questions yet, we know enough that we can start doing something about it. That will need the political will and backbone to do what is right, not just for us but more importantly for future generations. However, such wisdom and strength of character is a rare commodity, not only among politicians. That there are still leaders in our societies that question the reality of what is happening is not only sad, it is dangerous; not so much for our generation but for those who come after us.

Earth can live with very different climate conditions and global temperatures and carbon dioxide levels have varied significantly over geological time. Fig. 2.11 and fig. 2.12 show Earth's global temperature and carbon dioxide levels over the past 541 million years and the last one million years, respectively. One can only imagine what the reaction of a scientist from the nineteenth or early twentieth century would be if they could see this kind of data. It is extraordinary indeed that modern science and its ever-improving methods can give us now a meaningful understanding of climate conditions well into the distant past. The data records used to put together fig. 2.11 and fig. 2.12 come from a number of research teams across the globe (see Refs. [23–32]). As this invaluable work continues, we can expect some of the uncertainties in the data sets that are clearly still present to be reduced.

Fig. 2.11 and fig. 2.12 show, that over the last 541 million years global temperatures have deviated much from the modern thirty-year average for the years 1960 to 1990; and similarly, carbon dioxide levels have been much higher and lower over the same time frame. To determine Earth's global temperature, scientists analyze data from land- and ocean-based meteorology stations to calculate average values for each cell in a grid box network that spans Earth's surface. This data is then combined to calculate average temperature values for the northern and southern hemispheres separately before they are merged into a global average. Several organizations from different geographies have been contributing to this effort for many years now and with the availability of satellite-based temperature measurements truly global temperature mapping is now possible. The data of different organization show some variations, which is to be expected for such studies, but the overall agreement between the different data sets is good. Unlike politicians, scientists have their own data continuously fact checked by their peers, whose research and measurements may use different techniques and methods. The beauty of science is that measurement data and the respective methods used to acquire them are scrutinized constantly and over time, this ensures an ever-increasing data quality and convergence towards highly trusted and verified data sets. Earth's global temperature data are no different in this respect. This process has helped to improve the data set for modern times where we do have actual measurements available as well as for the data used to estimate the prevailing global temperature conditions many million years ago. The same applies for the determination of atmospheric carbon dioxide concentrations in modern times and in Earth's distant past.

This section started with a brief review of the methods scientists have at their disposal to understand past climates and temperatures and it is from such work that we have the global temperature and carbon dioxide data for pre-modern times back to 541 million years ago. Clearly, the uncertainties of such data for the distant past will be larger than for what we can measure in our time. However, as the curves in fig. 2.11 and fig. 2.12 show, the variability of Earth's global temperature and carbon dioxide levels over time is significantly higher than the potential uncertainty of the data. For the earliest data in the temperature and carbon dioxide records, this uncertainty is indicated in the respective graphs by the light gray bands the data is embedded in. These uncertainty envelopes should be seen as guides to the eye and rough measures of the average data uncertainty but not as the true measure because data variability over much shorter time intervals within these periods could be higher or lower. The important fact is that the large temperature and carbon dioxide level variations observed over these periods are well beyond any potential data uncertainties. For now, we have much fewer data from these distant pasts than for more recent geologic periods. Over time, we can expect this to change. New insights and data will likely provide for a more detailed understanding of global temperature and carbon dioxide levels in Earth's distant past. However, the overall shapes of the temperature and carbon dioxide records we already have now are unlikely to change much. So what does the data in fig. 2.11 and fig. 2.12 tell us?

Quite clearly, global temperatures were on average warmer than they are today. We can also see large temperature swings over long time periods. On what looks like a roughly one-hundred-million-year cycle, global temperatures varied between about ∼14 degrees warmer and ∼6 degrees Celsius colder than the modern-day average (see fig. 2.11a). Currently, we are closer to the lower level in these long-term oscillations. The carbon dioxide level variation mirrors this temperature cycle but not perfectly as it shows more details around the long-term trend than the temperature data. We should not expect a perfect match of the carbon dioxide trend to what the temperature data shows. Changes in carbon dioxide level itself, caused by climate changes or other factors, are forcing factors driving temperature changes; in turn, they can lead to additional carbon dioxide level changes in either direction. We also need to be keenly aware that the temperature and carbon dioxide records shown here for Earth's distant past also reflect the gigantic changes Earth has seen over geological time with plate tectonics reshaping Earth's surface on a massive scale. We have no such experience in our short human history to which we could compare any such changes. What is normal in terms of temperature or carbon dioxide changes on a geological time scale may not be something that would support the evolution of complex life such as us. Therefore, while it may be interesting to point out the much bigger changes in temperature and carbon dioxide in the very distant past, much more relevant for us are the temperature and carbon dioxide records of the more recent past in Earth history. Specifically, the stretch of time that supported the evolution of primates, which eventually would lead to us. It is on that scale where the changes we are making to Earth's climate are very significant.

As we have seen earlier, the solar forcing mechanism discovered by Milanković is an important factor in driving climate changes. During these climate cycles, changes in vegetation and in ocean temperatures will result in different carbon dioxide levels in the atmosphere depending on whether positive or negative feedback in the carbon dioxide cycle prevails. The overall effect is that less carbon dioxide is found in the atmosphere during cold periods than during warm periods. This correlation is evident when looking at the correspondence between maxima and minima in the temperature and carbon dioxide data in fig. 2.12a. Carbon dioxide levels rise and fall along with changes in solar forcing but because carbon dioxide is a greenhouse gas, it also acts as an amplifier of solar forcing. However, there are also additional factors, other than solar forcing, which can produce a significant change in atmospheric carbon dioxide levels, such as for example increased volcanic activity. This has happened on a massive scale several times in Earth history producing much higher carbon dioxide levels. How Earth's global temperatures respond to such events not only depends on the increased amount of carbon dioxide in the atmosphere but also on the amount of other greenhouse gases such as methane present in the atmosphere. In addition, when estimating the overall climate impact, we must also consider the dust in the atmosphere resulting from such increased volcanic activity as well as higher levels of gases such as sulfur dioxide or hydrogen sulfide. In short, while solar forcing will be reflected in the carbon dioxide records as well as the temperature records, carbon dioxide forcing from events such as increased volcanic activity have their own imprint on temperature records. If just carbon dioxide is added to the atmosphere, as we have been doing since the start of the industrial revolution, the result is a predictable rise in global temperatures, which we have been observing now for some time.

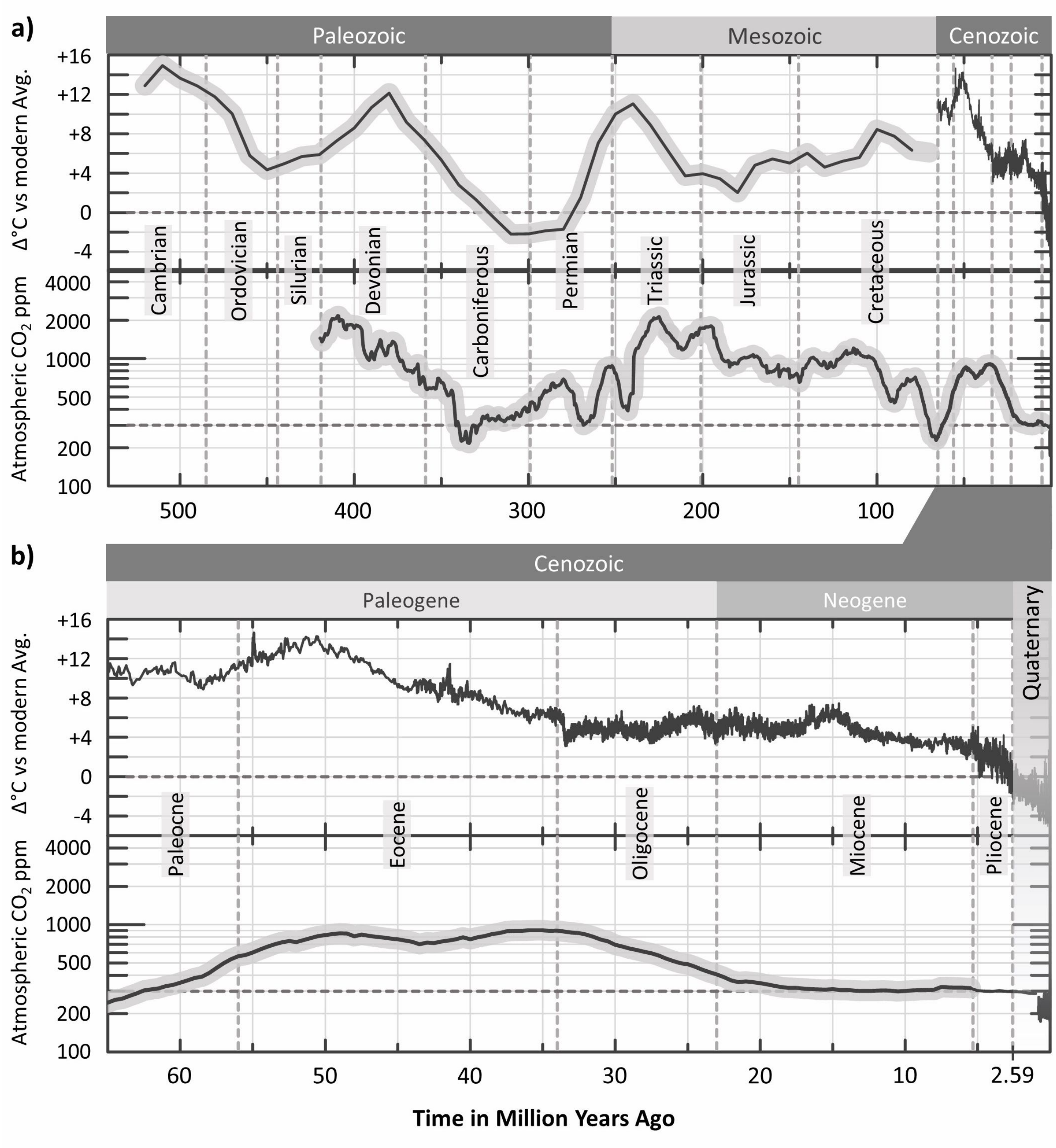


Figure 2.11: Variation of average global temperature vs. the modern thirty-year average (1960 − 1990) and atmospheric CO_2 levels in parts per million (ppm) over the past 541 (a) and the past 65 million years (b). For data sources please see Refs.[23–32].

The overall increase in global temperatures due to the carbon dioxide we pumped into the atmosphere over the last roughly 200 years looks small, but its impact is already clear and more importantly, we are now locked into continued global warming for many generations to come, even if we stopped putting more carbon dioxide into the atmosphere immediately. Unfortunately, we do the opposite as we produce ever more carbon dioxide.

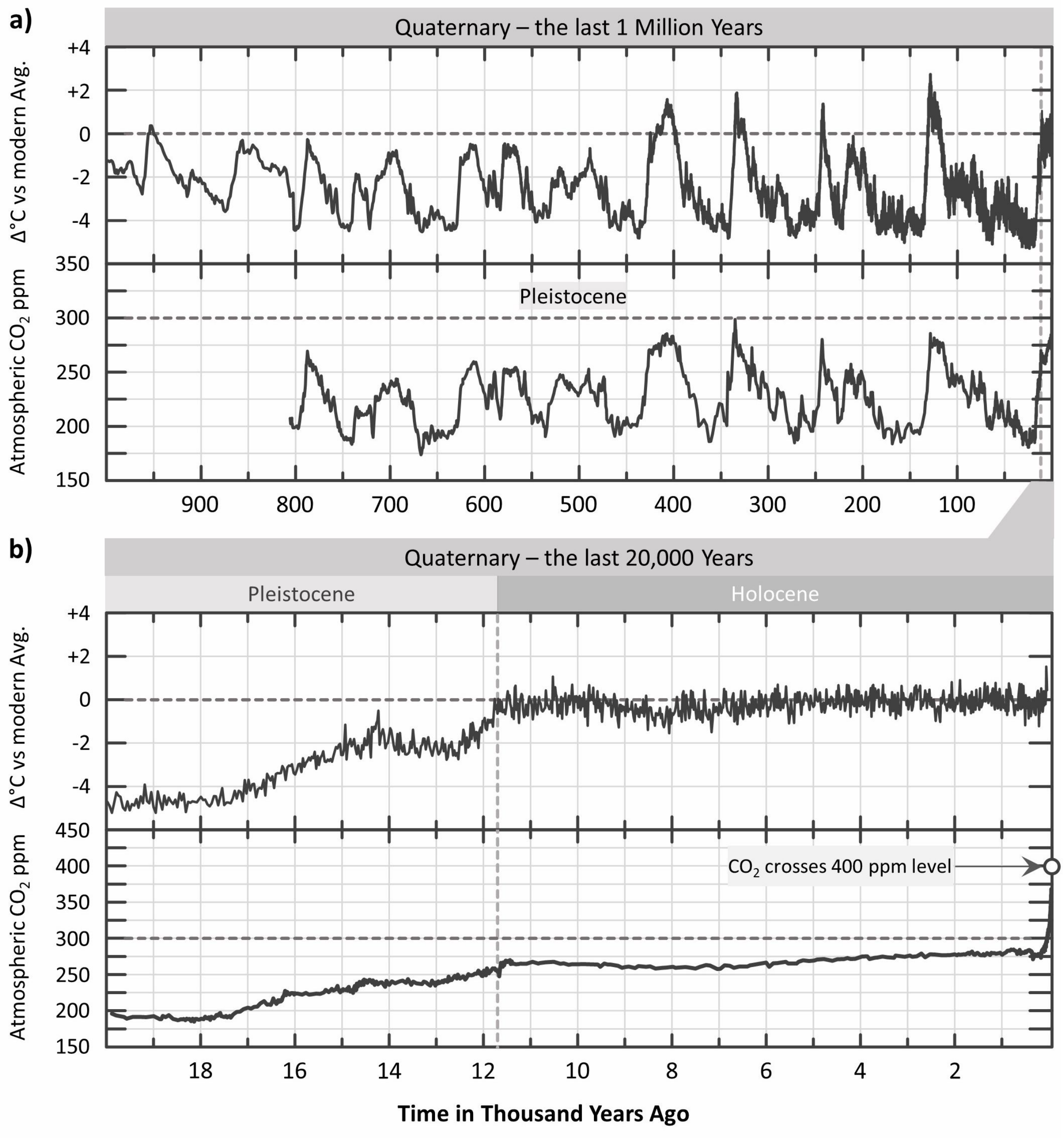


Figure 2.12: Variation of average global temperature vs. the modern thirty-year average (1960 – 1990) and atmospheric CO_2 levels in parts per million (ppm) over the past one million years (a) and the past 20,000 years (b). For data sources please see Refs. [27–32].

Fig. 2.12 shows the last one million years in our current geological period, the Quaternary, on a more detailed level. The temperature and carbon dioxide data in fig. 2.12a cover the time period of the glaciation events that we looked into earlier. The clearly visible large temperature and carbon dioxide variations over the past eight hundred thousand years have their origins in the solar forcing cycles discovered and explained by Milanković.

The temperature and carbon dioxide data are mirror images of the insolation variation and the average marine oxygen isotope ratio variation from fig. 2.7, which were the first proof of Milanković theory. And they should be as a good amount of the temperature data is based on oxygen isotope variation. The carbon dioxide data are an additional independent confirmation as they do come from analyzing ancient atmospheres, preserved in tiny ice bubbles within ice cores. What immediately stands out in fig. 2.12b is how remarkably little global average temperature and carbon dioxide levels have varied over the last roughly 10,000 years, the geological epoch we call the Holocene. It is likely not a coincidence that with the beginning of the Holocene we see the first human steps towards agriculture. Our distant pre-historic ancestors during the Pleistocene, the epoch preceding the Holocene, had to face much more varied and harsher climates than our more recent ancestors. As we look at the data of our species recent past, we can see that carbon dioxide levels have been rising since the industrial revolution, breaching the level of 400 parts per million in 2016. Not since many million years has our planet seen such high carbon dioxide levels.

Sometimes the argument is made that we should not worry about global warming because it barely registers yet on the curve shown in fig. 2.12b and Earth has supported life at much higher average temperatures for many millions of years. Indeed, we have to go back to the late Oligocene for similarly high carbon dioxide levels. As we continue unabated to pump carbon dioxide into the atmosphere, we are well on our way to create climate conditions that may have prevailed in the Eocene. If we still were just one animal species among many others an Eocene climate may not even matter at all and the argument that we should not worry about a much different climate would not seem to be that far fetched. Life has flourished on Earth throughout the past 541 million years with many species becoming extinct and many new ones evolving in a constant coming and going, punctuated by several mass extinctions that cleaned the slate for new life to take center stage. Humans have only been around for a very brief period of time and modern humans did not arrive until some 200,000 to 300,000 years ago. The duration of modern civilizations on Earth is very much shorter than that, only a few thousand years. Maybe we just were fortunate and the prevailing climate over this short period has been reflective of a stable equilibrium which when pushed a little out of its balance always regains it. What happens if this changes? How exceptional is the recent 10,000-year climate stability that humanity has enjoyed? We do not know, and we do not know either if our complex human societies, expected to level at a worldwide population of roughly eleven billion by the end of this century, will be able to prevail through much more challenging climate conditions which could last for many thousands of years or longer. In this context, it is instructive to observe just how powerless even the richest and technologically most advanced societies on this planet are in the face of larger natural disasters. This should make us skeptic as to how resilient our societies will be in the face of future potentially much bigger natural disasters that we can expect as certain consequences of the ongoing climate change.

On a few pages it is not possible to give a detailed account of how vastly more we know today as compared to a few decades ago about the components of Earth's climate system and how they interact: the atmosphere, the oceans, the biosphere, the land surfaces, and the cryosphere. During this brief period we have transitioned from reconstructing past climates and understanding our current climate conditions to projecting the future of Earth's climate. Enabled by increasingly powerful computers, scientists could finally distill their understanding of Earth's climate system into ever more sophisticated computer models to simulate Earth's climate history as it was being unveiled by paleoclimatology. Climate models calibrated using Earth's climates of the past have enabled us to make prediction about what Earth's future climates may look like. The first climate models to integrate the couplings between the different components of Earth's climate system and to model positive and negative feedback mechanisms go back to the late 1960's.[59] These first models already discovered the central role that carbon dioxide plays in Earth's climate conditioning and they made the first predictions of the impact a carbon dioxide increase would have on global temperatures. Much of the understanding of how the various climate components worked and interacted was already there and since then has only become more accurate. However, what changed dramatically over the past fifty years is the computing power that is available to run such models at ever-increasing detail. The most complex models are general circulation models that divide Earth's atmosphere and oceans in discrete small cells and simulate the evolution in these cells and their interactions with neighboring cells over time. Ever-faster computers and ever-larger computer storage has allowed the discrete cells to be shrunk to very small sizes and to make the discrete time intervals of these models so small that they essentially become real-time continuum simulations in space and time.

Some still deny the reality of human induced climate change, frequently deriding its scientific evidence as *"garbage in equals garbage out"*. The arrogance behind such statements is astonishing to say the least but more than anything else, it reflects a person's ignorance. In today's world, someone who had no opportunity or willingness to familiarize her- or himself with the basic concepts of science will with increasing frequency be in situations where she or he has to make a choice. This seemingly uncomfortable choice is between acknowledging that she or he does not have the expertise in a given area and therefore wisely consults experts; or she or he outright challenges science itself, essentially mistaking it for a belief system rather than for what it really is, our best understanding of the real world. This is unfortunate, specifically if such a person is close to or at the seat of power where it is being decided what to do about global warming and climate change. It may be easier in this situation to ignore the facts and deny that human induced global warming and climate change are realities, even more so if this is also promoted by the money interests that today seemingly hold the keys to move someone into positions of political power. We will have to see how long this can go on as it is becoming increasingly clear what the potential consequences will be. Starting with the industrial revolution, we can trace an increasing impact of human societies on Earth's climate. Human induced climate change is a real and indisputable fact. Those who doubt it do not know what they are talking about or worse, they know what they are talking about but choose to lie and deceive. There may still be some lost scientists out there paid by special interests who seek to debate the established facts and thereby try to insinuate that human induced climate change is not proven. They are dead wrong and discussing this topic with any of these people further is a waste of time. The debate that we need is really a very different one: Does the human induced climate change matter and what can and should we do about it? We will come back to this topic later.

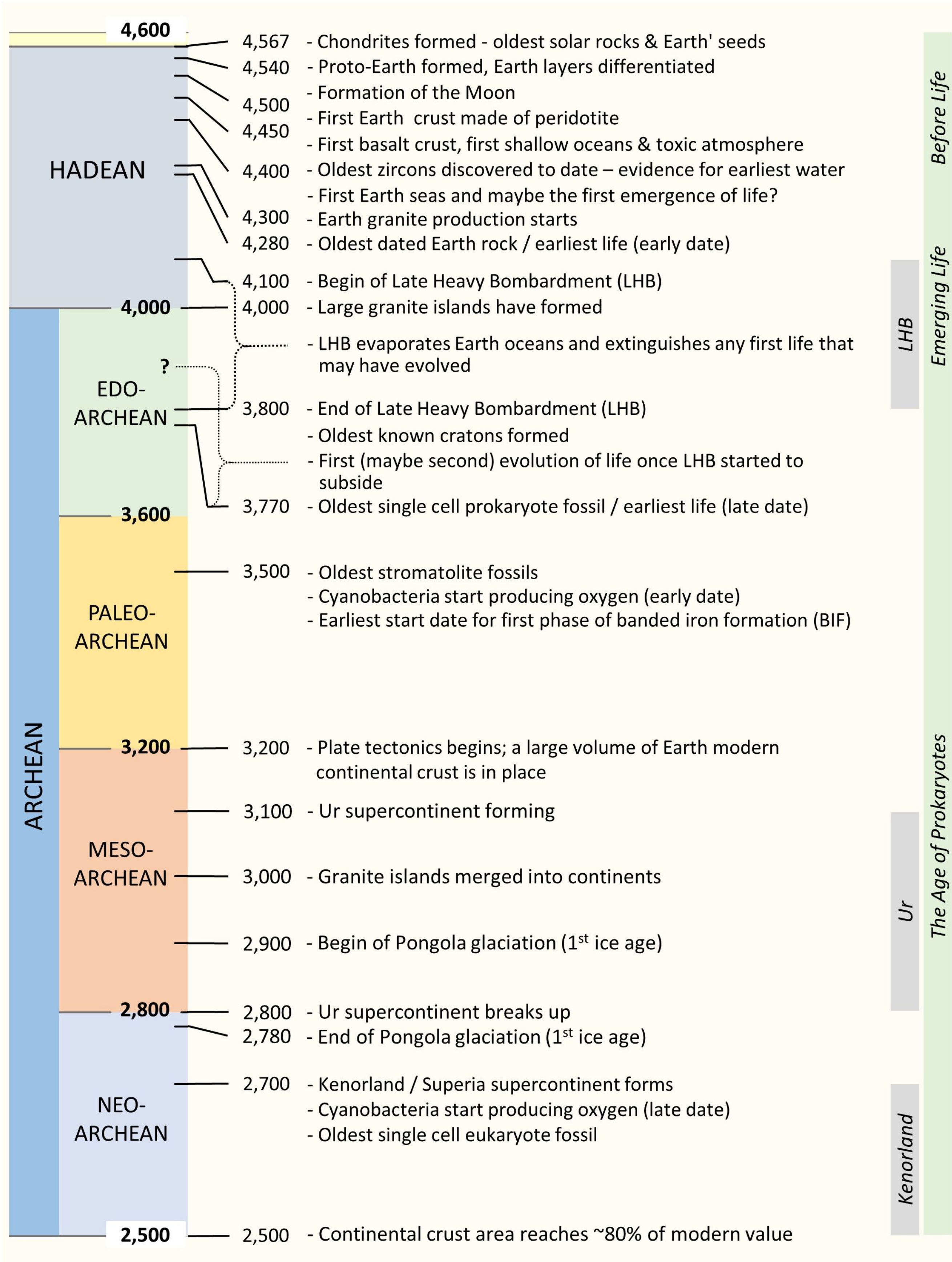


Plate I: Geological time scale: 4,600 - 2,500 million years (logarithmic scale).

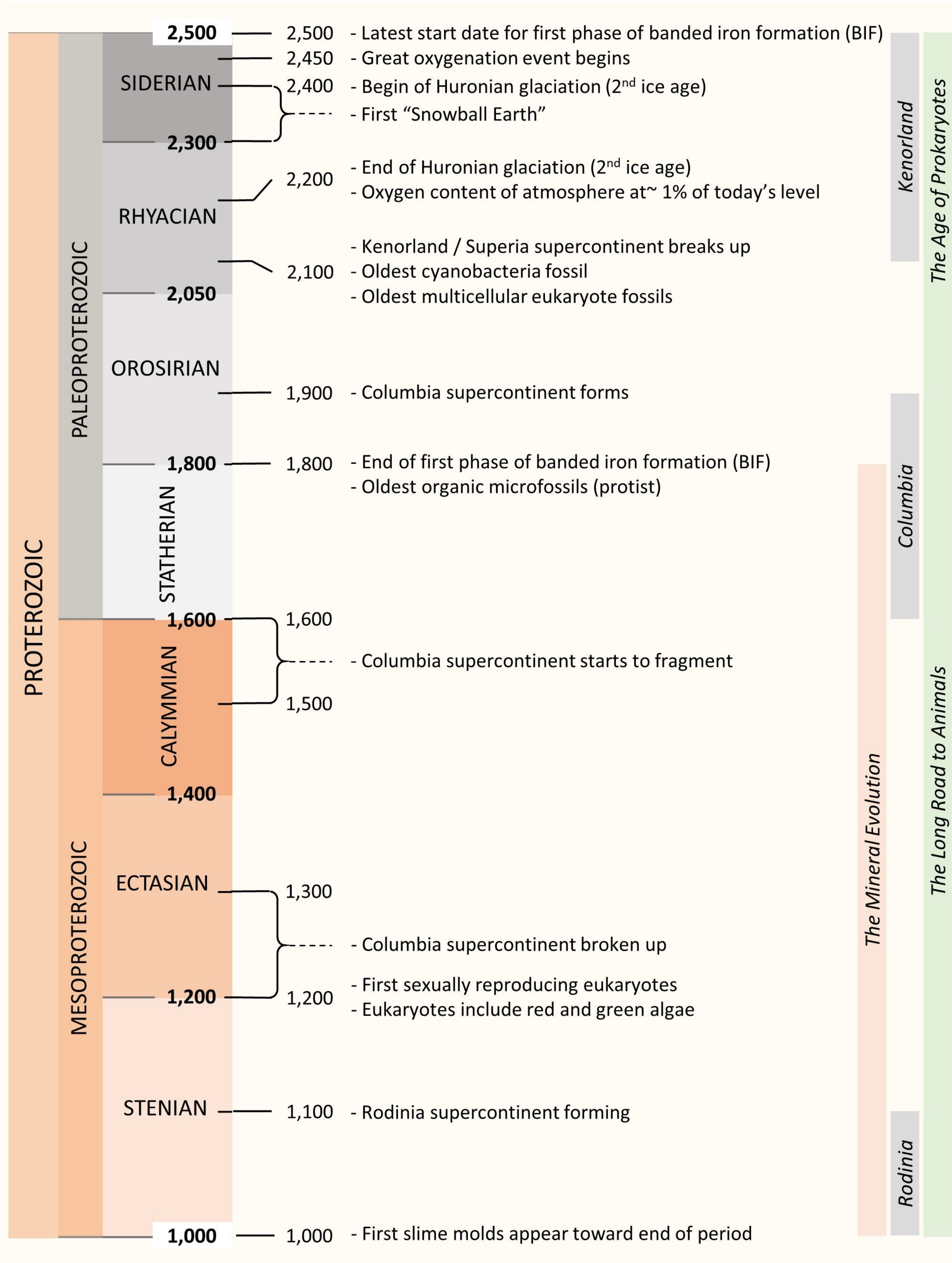


Plate II: Geological time scale: 2,500 - 1,000 million years (logarithmic scale).

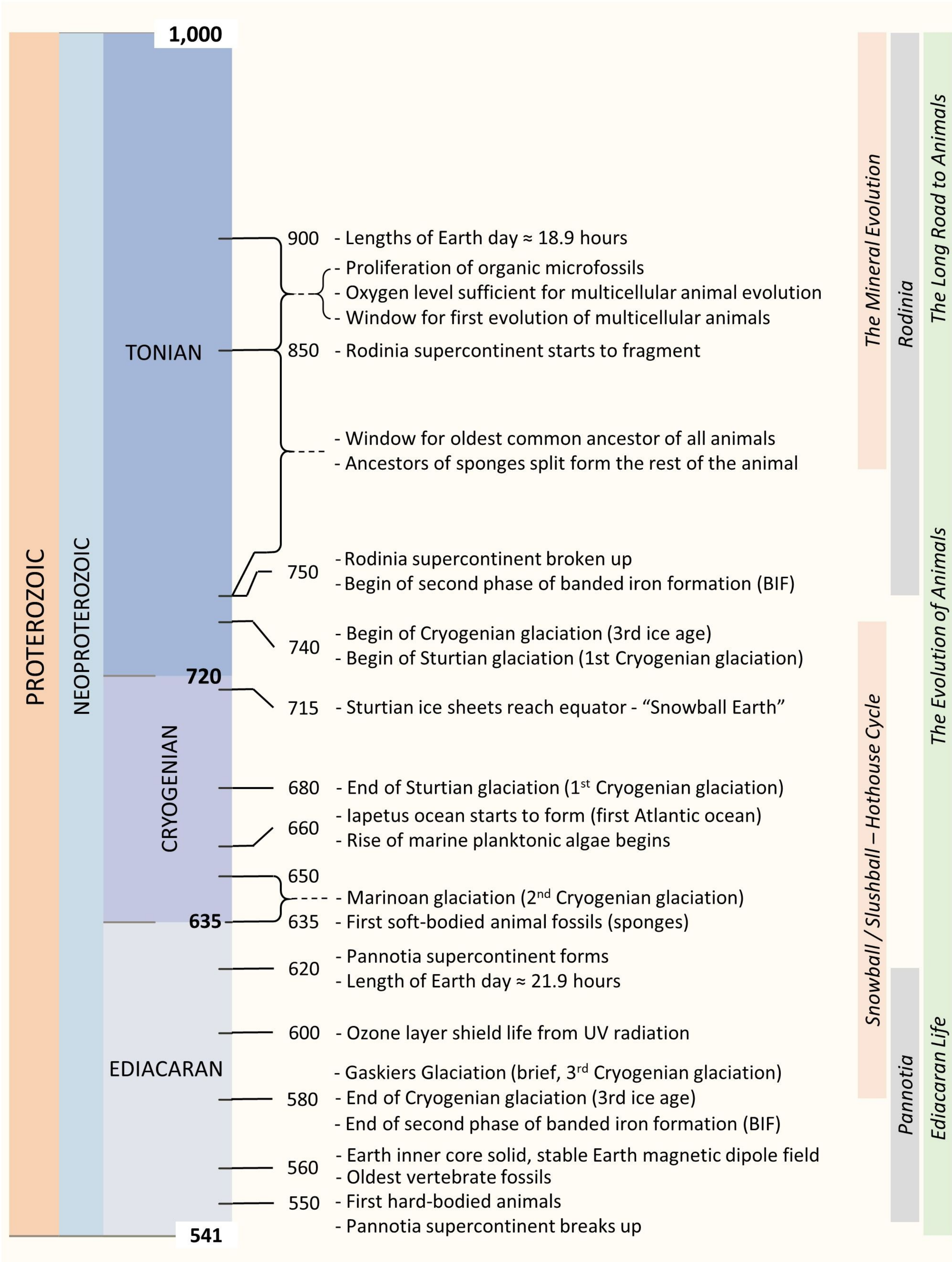


Plate III: Geological time scale: 1,000 - 541 million years (logarithmic scale).

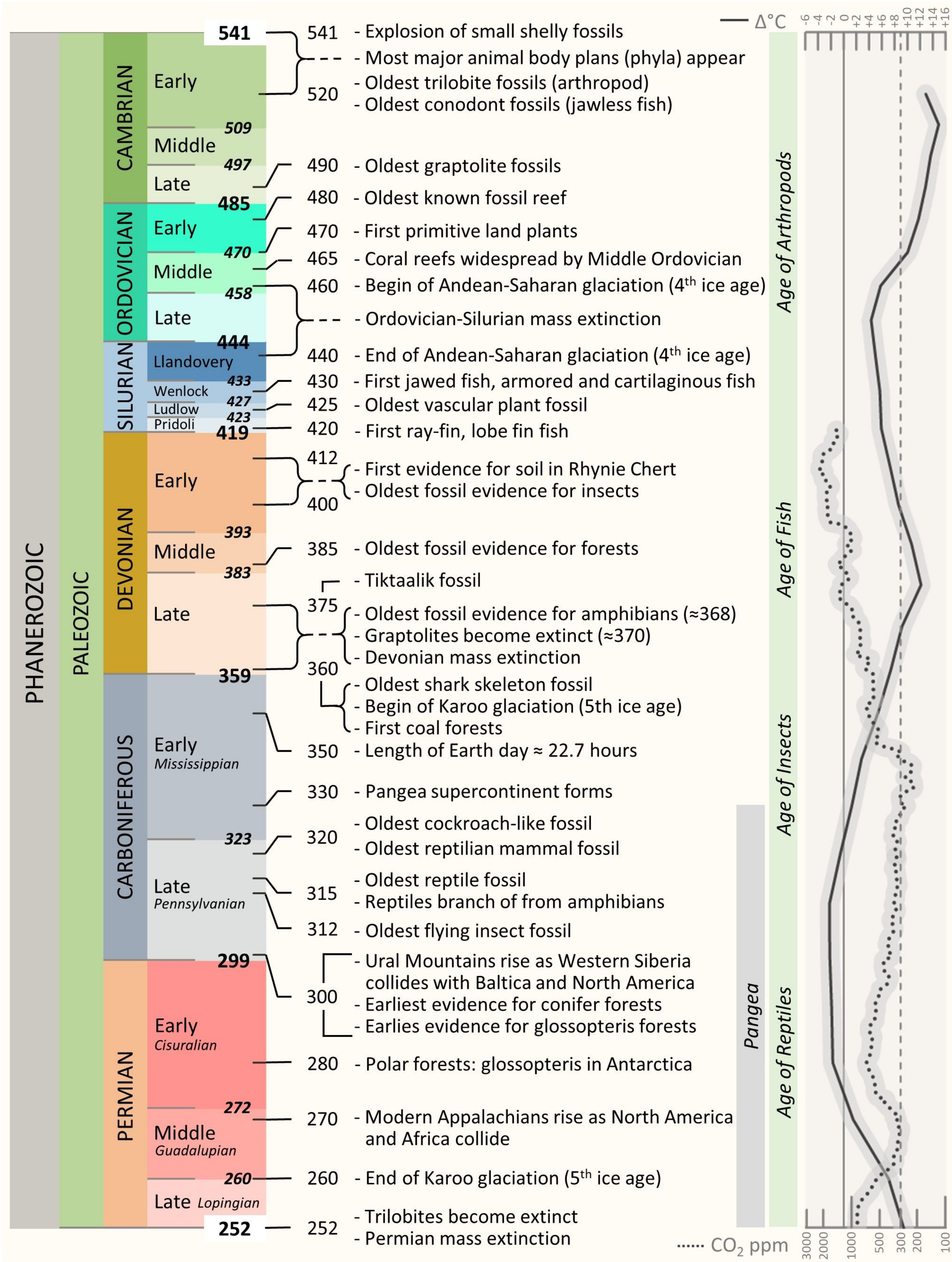


Plate IV: Geological time scale: 541 - 252 million years (logarithmic scale).

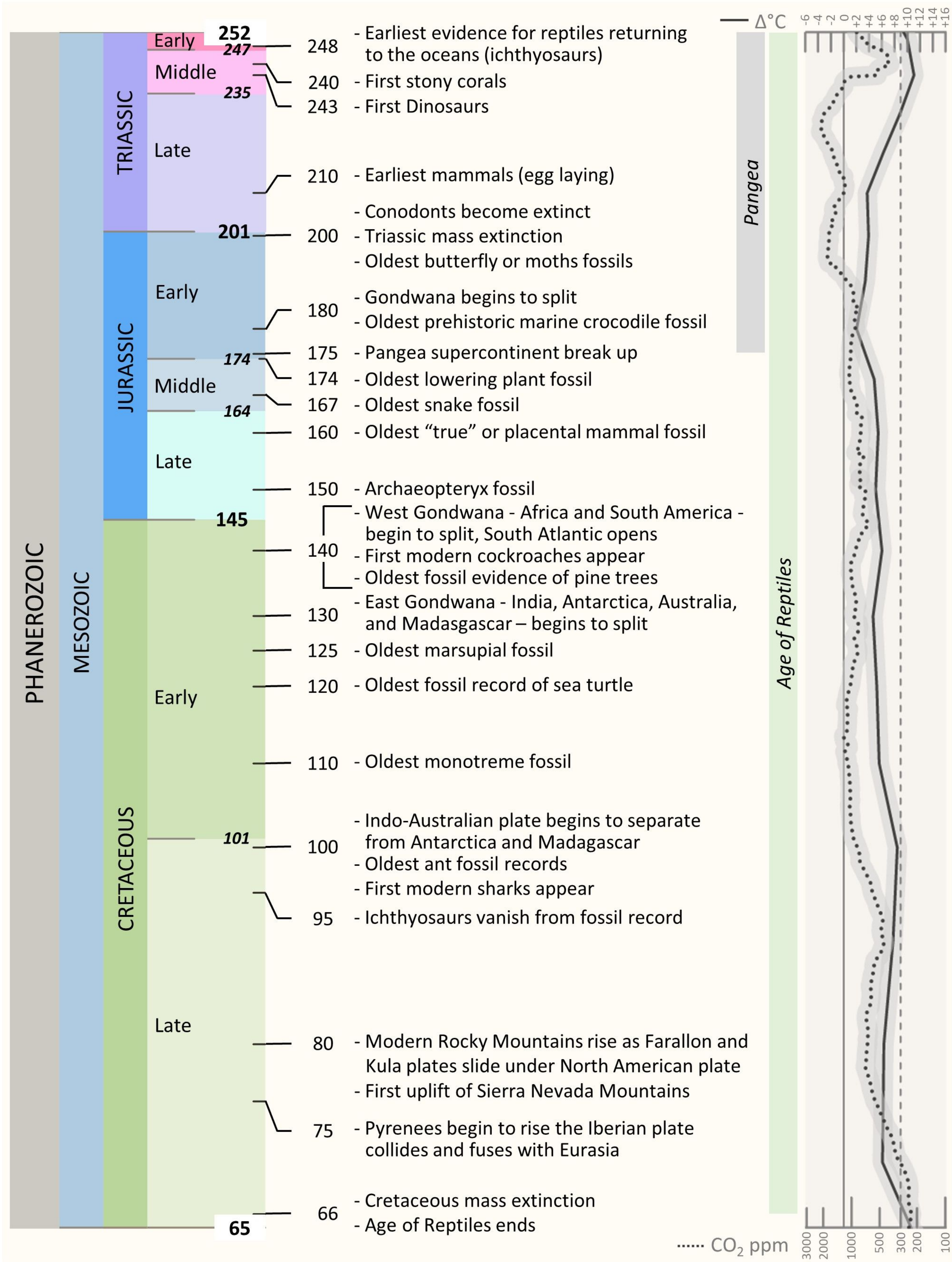


Plate V: Geological time scale: 252 - 65 million years (logarithmic scale).

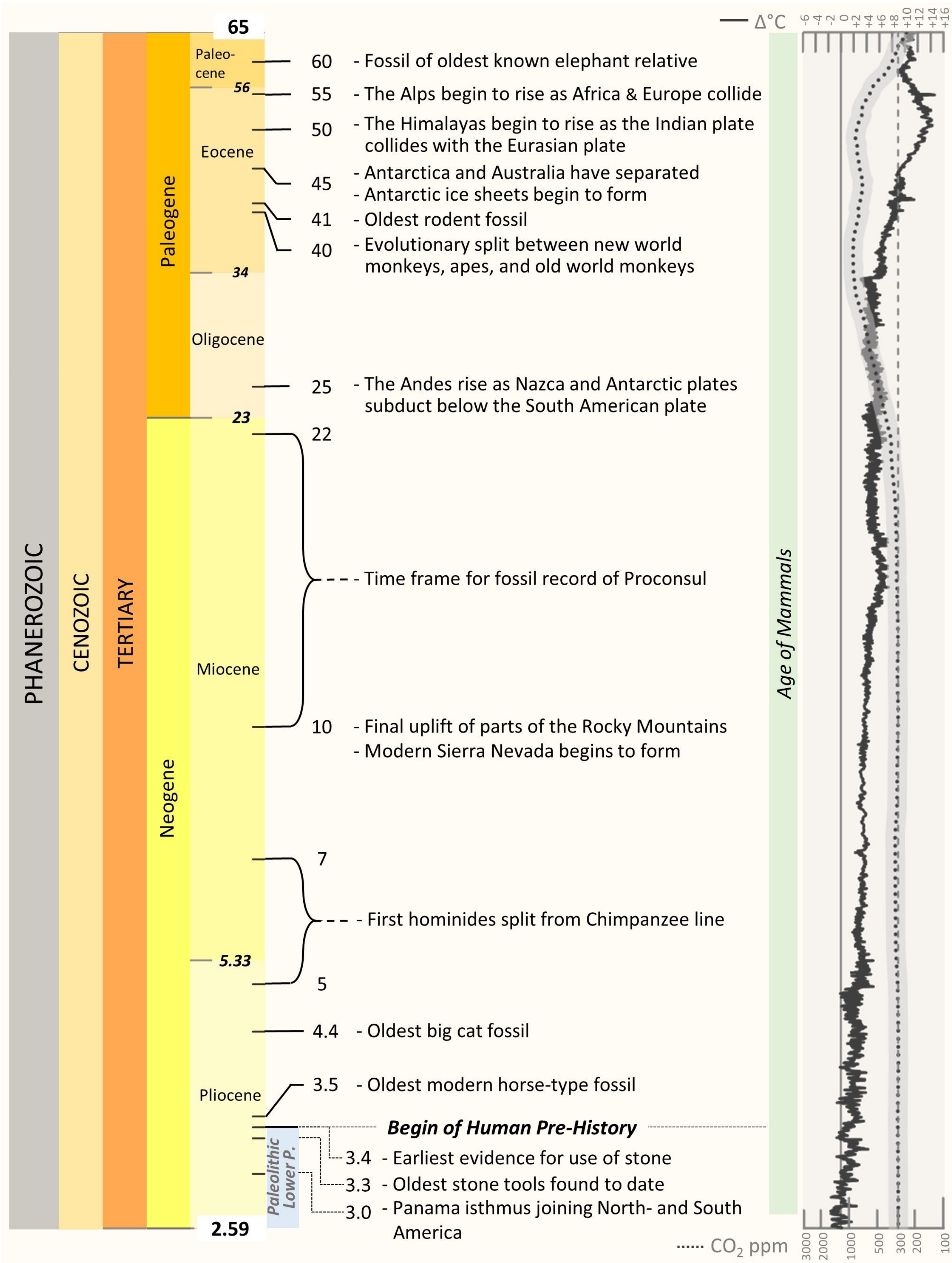


Plate VI: Geological time scale: 65 - 2.59 million years (logarithmic scale).

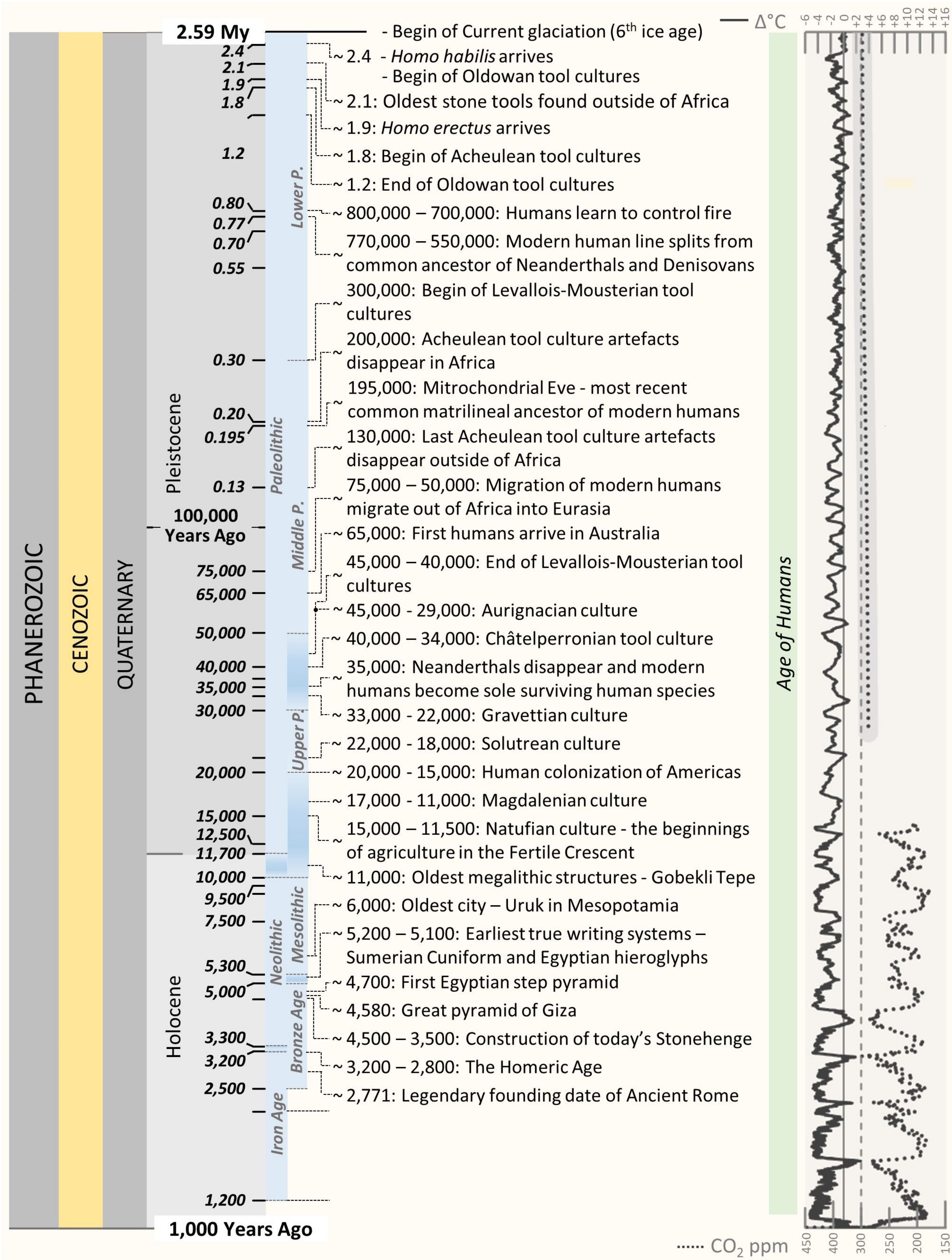


Plate VII: Geological time scale: 2.59 million years to 1,000 BCE (logarithmic scale).

Life On Earth

Spring Life

In the first chapter of this book we explored how our view of the universe changed since humans first gazed inquisitively up into the sky. Then, in the previous chapter, we looked into how our understanding of Earth evolved, from mere speculation prevailing until some three centuries ago, to the detailed knowledge of our planets history and its inner workings that we have today. Along the way, well into our very recent past, we had to give up cherished preconceptions about the universe and our home planet Earth. This is unlikely to change. Every generation builds on the knowledge of those preceding it, improves and adds to our human understanding, and identifies new challenges for future generations to explore. Some of the new knowledge that became available through a much-improved understanding of the cosmos we live in and the planet we inhabit, also has set the stage for a more profound understanding of life on Earth. We now know that evolution does not start with the first emergence of life but traces its roots back billions of years, the vast stretch of time that it took to put into place the preconditions for life. The evolution of life as we know it has been preceded by cosmological, stellar, and planetary evolution. Without several generations of stars seeding the universe with the chemical elements required for life, we would certainly not be here today. In a very true sense, we and all other life on Earth are indeed made out of stardust. Similarly, we would not be here without a habitable planet forming in our solar system, planetary evolution and the evolution of life are joined at the hip. Maybe there are other ways for life to evolve in the universe, different life than the one we know, able to emerge without a planet such as ours. We will have to leave this as a possibility. Here we are concerned with the evolution of life on Earth as it started some 4,000 million years ago on our planet. When we picture the evolution of life it is the so-called higher lifeforms which usually get most of the attention and this chapter may be guilty of this too. However, the masters of our planet Earth are the lifeforms that have been around for much longer than these lifeforms more visible to us. It is the small and to our naked eyes often invisible world of single-celled organisms that all higher lifeforms very much continue to depend on.

The objective in this chapter is not to provide a sweeping account of life on Earth, this would not be possible, but to paint a broader picture that helps to understand how we as humans do fit into this larger canvas. Importantly, this has to include an outline of how we came to understand what we know about life on Earth today. The way humans have evolved and now seemingly control much of this planet for better or worse also makes us responsible for what happens to other lifeforms on Earth. Both, with regard to what we do and what we neglect to do. Life may also have evolved somewhere else in this universe. From all we know about the universe, we practically have to assume that from a statistical perspective this is almost a certainty. Unfortunately, this same knowledge and the laws of physics as we know them today also tell us that any life existing in other regions of the universe, however far or widespread among the stars, may never connect. So, we must look at life on our planet as if it were the only life that there is in the universe. This is the perception that should inform how we behave with respect to other lifeforms on this planet and this is how we should hold ourselves accountable.

For what follows in this chapter it will be helpful to take a look back into the evolution of biological thought and to familiarize ourselves with some of the basic vocabulary of biology and botany just as we would do when traveling to a foreign country, the language of which we may not know. The word *biology* derives from Ancient Greek combining the word for life, *bios*, and the ending *logía* for study or discourse. It first came up sometime during the eighteenth century and was firmly established by the early nineteenth century, referring to the field of science concerned with the study of all living organisms. The origins of biology reach deep into the human desire to understand ourselves, as human beings and as to how we relate to the world around us. Like all modern sciences, biology has evolved into an edifice with a large number of rooms added over time, many of which have become their own scientific disciplines. A detailed historical account of how biology became the science we know today is beyond the scope of this chapter. Rather more, the objective here is to get a sense of the broader outlines that eventually led us to where we are today with respect to understanding the living world around us.

Biology, the science of life, feels different to most of us when compared to mathematics, physics, astronomy, cosmology, chemistry, geology, mineralogy, paleontology, or climatology; some of the sciences prominently featuring in this book. The reason for this is that it relates to us in a much more direct way simply because the science of biology is our effort to understand life itself of which we are an integral part. Sometimes, one also sees the so-called exact sciences, think of mathematics and physics as two prime examples, differentiated from the supposedly not so exact sciences, the humanities, social sciences, and arts, with biology deemed to being closer to the exact sciences but not exactly one of them. This, however, is a very misguided view. All of our human efforts to study the world around us and ourselves systematically, following the common set of rules by which we judge science to be science, are part of the same scientific endeavor. It just so happens that they target different areas of our human understanding of the world. For some sciences, this meant that they had to start out with the grandest of all challenges right from the beginning while others could increase their learning in an incremental way, tackling easier challenges first before they also ran into the big challenges. To appreciate this one just needs to look at biology vs physics. Many challenges in physics allow scientists to isolate and idealize the problems at hand into simpler and easier models we

can describe mathematically. This is how physics has incrementally built our knowledge about much of the nonliving world. In the process of doing so, physics has invented many new ways to explore the world which go far beyond what our senses allow us to detect.
Based on our steadily increasing knowledge about the physical world around us, engineers have been able to build ever better measurement instruments. These allow us today to measure infinitesimally small things on the scale of elementary particles as well as infinitely large things on the scale of the universe. With our own senses, we probably can estimate distances from roughly one millimeter or one thousandth of a meter on the low side, to maybe several thousands of meters on the high side, spanning roughly six orders of magnitudes. However, our physical measurement instrumentation enables us to measure distances over roughly forty orders of magnitude. This represents a ten-billion-trillion-trillion-fold increase, which put in numbers is a one followed by thirty-four zeros, over our sense perception. Similar to this extended length scale our ability to measure time, using one second as the basic unit of measurement, now stretches over roughly thirty-eight orders of magnitude. Compared to that, our human sense to perceive time may stretch from a second to historical times, which we can generously put at a few thousand years, resulting in roughly eleven to twelve orders of magnitude in our human perception of time. On these vastly enlarged time and length scales, we can now measure physical properties to which our senses are otherwise completely oblivious. Only now, with these powerful measurement tools at hand have physical sciences discovered their very own big challenges. Such as for example what the universe is made out of, because as it looks today, we really only know a small fraction of it. Finally, the physical sciences are running into the same kind of conundrum that biological sciences had to tackle from the very beginning. However, they still get a pass in some respect, as unlike the biological sciences they can deal with this in a different way by stating for example that it makes no senses to talk about time and space before the *Big Bang*. In a certain way, physics, just like biology, now also has to address the challenge of creation. However, since there exists only one universe and we cannot see beyond it in space or before it in time, these questions may indeed be useless to ask. Not so for biology. Life did somehow start on planet Earth and at some point, developed into the lifeforms we see today. Hence, for biology there is a before and after with respect to life; for physics there is only an after with respect to the *Big Bang*.
The challenges with understanding life on Earth, how it originated and evolved, are manifold. As we look back in history, biologists, as well as natural scientists and philosophers before them, have broadly used two approaches. The first one is frequently referred to as the reductionist approach and the second one as the systems approach. The reductionist approach is close to how physics works. It takes the immensely complex phenomenon of life and seeks to understand it by analyzing the most basic physical mechanisms that make it work. Cell biology, which studies the functioning of living cells, their parts, structures, and their life cycles would be a good example for that; molecular biology would be another one. To use a simple analogy, this is no different from physicists having first to explore atoms before they can hope to understand the properties of solids such as metals, insulators, or semiconductors on a fundamental level. However, it is not possible to understand such a complex system as life on Earth represents using only a reductionist approach. Understanding evolution, ecosystems, how living beings interact with one

another and with their environments requires a system level approach. Such systems have very many degrees of freedom, characterized by a broad range of parameters often not easily separated, if at all; and most of them are difficult or next to impossible to study in a quantitative way in traditional laboratory settings, at least for now. Of course, comparisons between sciences such as physics and biology have their limitations because unlike atoms, many living beings have a mind of their own.

As we have seen already in the previous chapter, to understand the evolution of life one must consider the broader context of the planetary evolution that Earth has experienced and how those two sets of evolution, life and planet, are deeply connected with each other. Not only understanding the evolution of life but what life really is, that is quite another matter. While certainly difficult, answering this question for example for single-celled organisms is seemingly straightforward. However, once we enter the roam of self-aware life that becomes a completely different story. There is another dimension to biology, where as a science it is much closer to the traditional humanities than to physics or chemistry. We can only mention this here in passing. However, from the little we will say about this it will be clear that we must be very skeptical with respect to any claims of human exceptionality, like us being the only self-aware or self-conscious lifeforms. Those reflect more our still very limited understanding of other lifeforms than anything else.

All sciences work within their own ethical and moral contexts, most of it shared across disciplines. However, nowhere are they as strong and evident as in the medical sciences and all other branches that sprung from it, including biology and some of its children which increasingly are becoming sciences of their own, such as genetics. This is so for a good reason. The ethical and moral dilemma physicists faced when atomic bombs exploded over the cities of Hiroshima and Nagasaki was an indirect one. They were the enablers of the bomb, but they did not make the decision to drop the bombs on civilian populations. In a sense, physicists were one stage removed from culpability. Not so in the medical profession or for biologists, though that may be less obvious for the latter. Biologists working with living animals have to make a choice about what they have in front of them. A living being, so to say an individual of a species that is as unique as they are themselves, or if they perceive what they experiment with as a biological machine. For a long time, for most of its history, biology has treated other lifeforms as biological machines and nothing more. Today, this clearly is not the case anymore. However, we are still a long way from translating the understanding that biology gave us of animal life into social norms that would respect the fact that animals just like us feel pain and angst, and yes, that many of them have emotions. They may not be emotions like ours but still, they are emotions. Every pet owner will likely attest to that, but it took scientists several hundred years to acknowledge it; and not all of them have done so yet. As our understanding of our animal relatives, not only what they are but more importantly who they are, continuous to improve, humanities relationship with the animal world it belongs to will have to change dramatically. Given how cruel our species can even be to fellow humans this will not happen overnight but eventually we will stop torturing and killing animals.

There is another very critical difference between biology and the physical sciences not concerned with life itself. A given type of atom will always be the same regardless of where it happens to be in the universe, and it will not change in time unless it is unstable and radioactively decays into smaller daughter atoms. Just consider a hydrogen atom that has been around since a few hundred million years after the *Big Bang* when electrons first combined with protons to form hydrogen atoms. Unless it happened to become part of a helium atom in the nuclear furnace of a star, this very same hydrogen atom is still around today. While physicists have continued to find many more elementary building blocks of nature than they may have imagined one hundred years ago, the total number of these building blocks is still small. The *Standard Model* of particle physics counts seventeen elementary particle types, physicists call this their particle zoo. Now, for practical matters, the zoo of biology is home to countless species. Currently, the estimate for the number of classified species is just under nine million and the assumption is that more than eighty percent of all species currently living on Earth are still undiscovered. However, an estimated roughly ninety-nine percent of all species that ever lived on Earth have become extinct. Add those numbers and the total number of species in the biologist's species zoo comes out somewhere on the order of 4,500 million; it is a numerical coincidence that this equates to just about one species on average for every year that planet Earth has existed.

It is easy to see that it is a very different challenge to understand the natural laws governing the evolution of so many millions of species as compared to formulating the laws that rule how the physicist's particle zoo behaves. Even chemists have it far simpler with a periodic table that includes only one hundred fourteen elements, ninety-four of them occurring naturally and twenty-four of them synthesized by humankind. The approach to unearth the natural laws of life clearly had to be very different from how scientists approached physics or chemistry. Over time, with the improved instrumentation provided by the other sciences, biology was also able to develop its own reductionist approach to dissect and study the elements of life down to the molecular and atomic scale. Anyone thinking biology is a less advanced science than physics or chemistry or mathematics because it contains less formulas and is in many aspects more of a descriptive science has it backwards. The nature of the challenge posed to biologists and naturalists in history required a very different approach, and it has rewarded us with a broad understanding of the nature of life at that. Where there was a blank canvas for a long time, only slowly filled in with color and often in need of correction, we now have a rich tapestry of life in front of us. Since humans first seriously entertained biological thoughts, we have come a long way. However, our understanding of life and the living world is far from complete. Similar to the physical sciences, it is only a few hundred years ago that biology became a true science, finally leaving behind the ages where religious or other societal constraints could impose limits on the exploration and our understanding of life as we find it in nature.

Biology in Pre-history and Antiquity

The human desire to understand the roots of life goes back a long time. Western cultures throughout most of their history have largely had a utilitarian view of nature, which they inherited from antiquity. Most likely this utilitarian perspective on life traces back even further than that as the ancient Greeks were adept at absorbing knowledge from all around the ancient world and from those preceding them. However, some cultures took a less utilitarian and hierarchical approach, notably some Asian cultures. Some 10,000 – 15,000 years ago, humans became pastoralists and herders and eventually began embracing agriculture and along with it, new religious beliefs that came to fore. However, a long time before that, we do not know for how many tens of thousands or even hundreds of thousands of years, animism used to be the foundation of our distant ancestor's spiritual lives.[1] By the time new religious beliefs arose, animism had become part of the very fabric of human cultures, deeply ingrained in everyday lives and beliefs. Animism is the religious belief that practically everything in nature has its own distinct spiritual essence. Our word for animals derives from the same root as the word animism. Natural deities likely abounded in pre-history and still were prominent at the time of the ancient Greeks as they identified almost each of their Olympian Gods with a specific animal. All of this changed with the rise of the three monotheistic religions of Judaism, Christianity, and Islam in the Middle-East. Christian theology was quite clear that animals were put on Earth by their God to serve Man. Animals in a sense had their souls taken away and became nothing else but property; the lucky ones ended up as pets, the not so lucky ones on our plates. In some parts of Asia, things were different. The Buddhist worldview requires its believers not only to refrain from killing animals but even more so, to be kind to them. Similar, Hinduism is opposed to killing animals for meat as well, but not every believer is a vegetarian. As we move further east towards China, things change again more to a utilitarian perspective of nature. Common to all beliefs was that nature and all it contains had a creator. They only differed in how their beliefs allowed them to deal with all other life in this creation.

As we know today, humans developed some practical medical knowledge at a much earlier time than we were giving them credit. Anyone doubting this just has to look at the first aid kit found in 1991 alongside the ice mummy of a roughly 5,300-year-old Bronze Age human. The corps of this mid-forty man, popularly nicknamed Ötzi, was preserved in the ice of the Ötztal Alps of South Tyrol at about 3,000 meters above sea level, where he seems to have been ambushed and killed. His medical kit included a fungus that has antiseptic properties and can, if swallowed, also act as an antibiotic. This indicates that even as early as then humans had not only an intuitive but also a very practical medical knowledge of how to treat small wounds with antiseptic fungi; or what to swallow to treat parasitic worms which seems to have been the case with Ötzi. Through trial and error, voluntary or not, humans must have gained and improved such knowledge over many generations. A number of sources are likely to have helped humans to acquire this kind of practical biological know-how that was critical to their survival. Understanding which plants were edible and which could be harmful, potentially even lethal, or which ones had non-nutritional benefits were certainly two of these. Injuries ancient humans sustained or illnesses they contracted and the efforts they made to cure either, at whichever possible way, must have increased our ancestor's knowledge about human biology to some extent. Of course, there is also the butchering of all kinds of animals that humans killed to sustain themselves. That certainly must have produced an intimate familiarity with all the various animal parts, a familiarity that many of us do not have any more today as we usually buy our meat in nice and cleanly packed portions in a grocery store just like it had never been part of an animal. Humans have slaughtered animals for ritual purposes throughout known history, a practice that unfortunately continues in some places to this day. The motivations for this practice were manifold, at the center likely fear of the gods, hope to garner or regain their favors, and the desire to predict the future. Among other things, we can imagine that from our ancestor's perspectives these practices helped to ensure the benevolence of the gods for various human enterprises including fecundity of man and beast or good harvests; to atone the gods if they seemed to be angry with humans; or to divine the future from the entrails of dead animals. Whatever the reasons were for ritual slaughter, reading entrails also creates familiarity with animal anatomy.

With the Bronze Age, around 3,000 BCE, the earliest human civilizations had aggregated into territories ruled by city-states and united under the leadership of their respective ruling class. This happened in Mesopotamia, bordered by the rivers Euphrates and Tigris, or in the lands bordering the river Nile, the Indus River, or the Yangtze River. None of these civilizations seems to have pursued the study of botany or biology in a systematic way. However, warfare in these early civilizations became a state enterprise and that must have included the need to treat wounded soldiers. In warfare, city-states such as Ur in Sumer or the early pharaonic state in ancient Egypt, could only afford to lose so many soldiers before they lost the ability to defend their city or territory, its population, and likely most important, its ruling class. Therefore, it became quite essential to make sure that less severe wounds did not become lethal. There must have been a considerable practical knowledge of human anatomy to treat for example fractures of wounded soldiers. All early civilizations seem to have used a wide variety of herbal treatments. For most illnesses however, the only possible cure was likely still thought to come through divine intervention. In the case of ancient Egypt, developing the process of mummification and its practice over many centuries must also have shed some light on human anatomy. We know that there were physicians treating patients in these civilizations. Their toolkits certainly must have included herbs. Medical practitioners back then also seem to have used quite a number of animal extracts and animal body parts for treating various ailments. In some countries, animal derived potions or treatments are part of traditional medicine, the practice of which continues alongside modern medicine. The vast majority of such animal derived treatments are from a medical perspective useless at best. However, this practice is culturally deeply ingrained and laden with superstitious beliefs it continues into modern times. The unfortunate consequence is that this perpetuated obsession with animal derived potions in modern societies, particularly in some parts of Asia, practically ensures the future extinction of many already endangered species. Animal potions did not help to heal any more patients back then than they do today but physicians in the early Bronze Age likely believed they did.

Interestingly, in some cases, a physician's toolkit must also have included surgery instruments as we do find references to those in written records with respect to drilling and filling teeth. Dentistry back then must have been a rather unpleasant procedure as knocking a patient out before the operation or sedating her or him with alcohol or hallucinogenic drugs were possibly the only types of anesthesia available at the time. It must have been quite risky to be a physician too. In many cases, perceived malpractice, which would have been hard to avoid given the medial knowledge and practice back then, was often punished severely. This was the state of affairs the ancient Greeks found when they saw what was still practiced in Egypt or passed down from the Sumerians through those who succeeded them in Mesopotamia to eventually become part of the Babylonian canon of wisdom and knowledge.

The ancient Greeks admired both, the Chaldeans as they called the Babylonians, as well as the Egyptians. *The Histories* of Herodotus give a vivid picture of the fascination Egyptian culture held for the ancient Greeks. However, at its core, it remained alien to them and the same is likely true for Babylonian culture. The ancient Greeks, as inquisitive with regard to nature as they were with respect to almost anything else, were not satisfied with this state of affairs and tried to understand how the living world works. The Greek pre-Socratic philosopher Anaximander of Milet believed that all living things evolved from sea creatures as opposed to them being the creation of gods. He also reasoned that humans must have originated from another species. His rational was that human offspring had a much longer path to adulthood and during this extended adolescence phase could not care for itself. Specifically, young children could not find food for themselves for a long time. This is very different from almost all other animals where young must fend for themselves much earlier. Anaximander's conclusion was that this could not have been always so. Otherwise, there could be no humans because they would likely not have lived to reach adulthood and have children themselves. Hence, humans must have evolved from another species that had offspring that was completely self-sufficient from a young age. So a little more than 2,500 years ago, Anaximander was already entertaining nothing less than an evolutionary thought that got the broader outline right.

In antiquity, ritually dissecting animals had been part of many cultures to learn the will of the Gods. Divination of animal corpses, as it is often referred to, was not a Greek invention in any way and would go on to continue in many cultures in various parts of our world for a long time to come; it even may still be practiced by some today. The first person we know to dissect animals for a different purpose is Alcmaeon of Croton, a Greek who lived also in the fifth century BCE and possibly was a scholar of Pythagoras.[2] Croton, or Crotone as it is called today, is a small city in the Italian province of Calabria, which at the time was part of the Greek world of Magna Graecia.[3] Alcmaeon of Croton was not so much interested in anatomy but rather looking to find the seat of life and intelligence. He may have been a medical practitioner, but we do not know that for sure. We only have indirect sources informing us about him and his thoughts as none of his writings have survived. In a broad sense, he seems to have explored how our senses connect to the brain and later records credit Alcmaeon with having located the seat of reason and intelligence in the head, essentially identifying the brain as the organ of the mind. It is important to point this out because for the next almost two thousand years this assumption changed, relegating the brain to other functions such as for example cooling the blood and moving the seat of reason to the heart instead; that already was to begin with Aristotle himself.

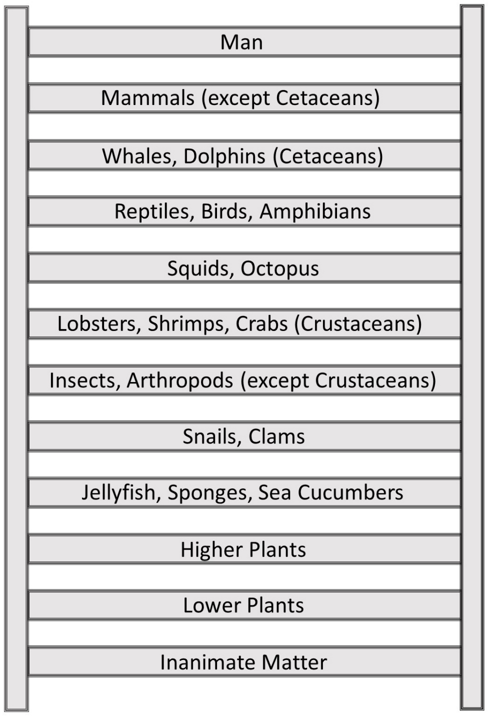


Figure 3.1: A simple version of the schematic known as Aristotle's *Ladder of Life*.

Empedocles (c. 490 – c. 430 BCE) was another pre-Socratic philosopher from the city of Akragas in Sicily in Magna Graecia.[4] We already met his younger fellow pre-Socratic philosopher Democritus in the context of his atomic theory of the universe. Democritus lived in Abdera in Thrace on the opposite end of the ancient Greek world.[5,6] He and Empedocles shared the view that all living organisms consist of what then were considered the four basic elements: earth, air, fire, and water. This view, which we already encountered from a different perspective in the context of the five Platonic Solids (see fig. 1.2), would come to predominate not only Greek thought but would long after that remain a bedrock of Western thought because it was the view that would be adopted by Aristotle. Hippocrates of Kos (c. 460 – c. 377 BCE) is likely the most famous physician of ancient times.[7] A substantial body of work on human anatomy and physiology has been ascribed to Hippocrates. Collectively, it is referred to as the *Hippocratic Collection*, but we do not know how much of this compilation comes from other sources and how much of it is genuinely from Hippocrates. Hippocrates subscribed to the same view as Empedocles and Democritus with regard to the four elements constituting life. The *Hippocratic Collection* includes the opinion that the brain's function is to cool the blood. Maybe this is where Aristotle got his view of the brain, which coming from an authority such as Hippocrates must have carried quite some weight. Once Aristotle added his weight to this opinion it would remain part of the medical and biological orthodoxy for almost two millennia.

We should not be surprised to find Aristotle among the founding fathers of zoology and biology. He knew of Anaximander's thoughts with regard to all life having evolved from sea creatures. He also was familiar with Anaximander's view on human evolution. However, Aristotle's interests lay much more in zoology and the biology of animals than in human biology. Similar to how he approached astronomy, Aristotle also used his great gifts of detailed observation and reasoning to understand the living world. Occasionally however, he included other sources without putting them through the same scrutiny that he applied to his original work. Aristotle's studies included animal anatomy, their similarities and differences, as well as animal behavior, but he was also interested in plants. Eventually he grouped animals and plants into a hierarchy, classifying animals with respect to function and complexity. The presumption of such a hierarchy derived from the idea that the entire natural world represents a continuum of life, from primitive to complex lifeforms. Later, this ordered hierarchy became the so-called *Scala Naturae* or *Ladder of Life*, which in medieval Europe would eventually give notion to the concept of the *Great Chain of Being*. It still underlies in many ways today's biological thought. Fig. 3.1 shows a schematic of a simplified version of Aristotle's *Ladder of Life*. Not surprisingly, Aristotle put Man on the top rung of the *Ladder of Life*. Next down on the rungs follow the respective groups Aristotle differentiated in his hierarchy: mammals; whales and dolphins; reptiles, birds, amphibians, and fish; octopuses and squids; mollusks; crabs, lobsters, and shrimp; spiders and insects; sponges; higher plants; and lower plants.[8] Clearly, today's modern biology disagrees with Aristotle's conclusions in many respects. However, in the way he viewed the continuum of life, he did understand a number of relationships in the natural world quite correctly. Aristotle's *Ladder of Life* would come to dominate the European perception of nature well past the Middle Ages. Aristotle was very systematic in all of his studies, one of the qualities that endeared him so much to later generations. So, it should come as no surprise that he also came up with what can justly be referred to as the first classification scheme of biology. Even though the word biology did not exist back then, biological classification is essentially the equivalent of what Aristotle did. Many more such classification efforts would follow and for a long time they would closely adhere to Aristotle's views until they eventually matured into the branch of biology that came to be called taxonomy.[9]

Where Aristotle led in zoology, Theophrastus (c. 371 – 287 BCE) his pupil and successor at the Lyceum would follow in botany.[10] He continued some of Aristotle's work on animal behavior but is known more for his works on botany. In a sense, Theophrastus did the same for botany that Aristotle had set out to do for zoology and he classified more than five hundred plants for their common and distinctive characteristics. The zoological and botanical legacy of Aristotle and his disciples included many deep insights; but they also were quite often wrong, frequently in areas where Aristotle seems to have copied from others. Aristotle's work in many ways summarized the knowledge of his time. Given the veneration for Aristotle in Western culture for almost two thousand years, it would turn out to be quite challenging to try to correct where he erred, as one just was not supposed to contradict an Aristotle for a long time to come.

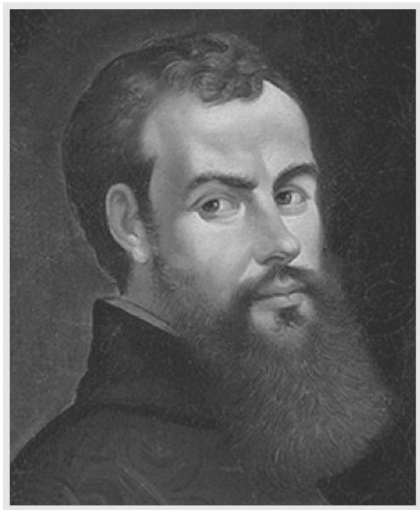
Aristotle represents the culmination of the ancient Greek's quest to understand nature and life. No other culture would match their intellectual efforts in these areas, and quite a few others too, for a long time to come. Certainly not the Romans who made great politicians, soldiers, engineers, and peasants but who were less inclined towards science. Their contributions to science, or better what we would call such endeavors today because science as we know it did not exist back then, focused much more on practical applications, engineering and agriculture prominently among them. We already mentioned Pliny the Elder's encyclopedic thirty-seven-volume work *Naturalis Historia* (*Natural History*) in the last chapter. Pliny devoted a large part of his encyclopedia to zoology. The part of zoology he was familiar with, farm animals or the fauna of his native Italian lands, he described reasonably accurate though repeating a number of Aristotle's errors. It was not below the dignity of a Roman noble to devote his time to the study of agriculture and the associated animal husbandry, and this was no different for Pliny.[11] Therefore, Pliny's familiarity with farm animals is no surprise. However, this practical knowledge of his native Italy's farm animals did not translate into skepticism with respect to legendary or mythical accounts of animal anatomy and behavior, which Pliny also included into his zoology. Whatever his sources were for the latter, they had little to do with reality.

Because much of the zoology part of Pliny's encyclopedia was devoted to such fantastic accounts, it thereafter was often referred to as Pliny's Bestiarium. Such Bestiaria remained quite popular until the late Middle Ages and some of this fascination lingered on for even longer than that.
Any major achievements in sciences during Roman times came often from Romanized Greeks such as the physicians Pedanius Dioscorides (c. 40 – 90) and Claudius Galenus (129 – c. 216) both from cities in Asia Minor, the former from Anazarbus and the latter from Pergamum.[12] Dioscorides work *De Materia Medica* (*On Medical Matters*), a five-volume compilation of herbal medicine, was one of the few works that was not lost in the turmoils of the fall of the Roman Empire. It would remain the most important work on pharmacology circulating in Western Europe throughout the Dark Ages. In many ways, it continued to be one of the most important resources of herbal medicine well into modern times. Claudius Galenus became known to posterity simply as Galen. His fame as a leading physician of antiquity was only second to Hippocrates and his medical authority would rule for more than one thousand years. He relied heavily on Aristotle, as did others, but he also studied animal anatomy in much detail. However, he was not going beyond the vivisection of animals. For Galen, as for most other physicians before him, dissecting a human body remained beyond his scope. Maybe more importantly, it was beyond what most in Roman society would have been willing to accept.
All ancient civilizations developed medical knowledge and not just the ones that we see as tributaries of what would become western medicine and biology. Much could for example be said for someone being better off, medically speaking, in parts of ancient China than in the contemporaneous Mediterranean World. The ancient Chinese medical practices of acupuncture and moxibustion predate third century BCE China, as by then they already were well-established treatments; acupuncture needles made of bone have been unearthed that date back to some one-and-a-half millennia BCE. Long before there were any European universities offering medical training, China's Tang Dynasty founded in 618 CE the Imperial Medical Bureau with what we would call today medical departments. Chinese medical practices were adopted by many of its neighbors, some of which however had their own medical practices such as we find them for example in Tibetan medicine and the traditional healing system of India, Ayurveda. The roots of Ayurveda are ancient, some date its origins back to the time of the Indus Valley Civilization around 3,000 years BCE. It certainly became a well-established medical practice during India's Vedic Age between roughly 1,500 to 500 years BCE.[13] In many ways, eastern holistic medical knowledge and practices were superior to what was available in the West. Over the millennia of medical development in the East, there must have been a good number of medical practitioners or physicians whose names are likely unknown to most in the West but who nonetheless were the equals of a Hippocrates or a Galen. However, the holistic approach would not lead humanity towards the modern sciences of medicine and biology. It was the reductionist Western approach which over several hundred years, improved by every generation, would lead us into medical and biological modernity. Therefore, the protagonists in the narrative to come all hail from cultures rooted in the West.

Biology from the Dark Ages to Enlightenment

With the demise of the Roman Empire, the study of biology suffered in the same way as all other sciences did. During the time period often referred to as the European Dark Ages, there was scarce new knowledge acquired and much was lost. Western Europe was reintroduced to much of Greek medical and biological learning through Muslim scholarship which had preserved this knowledge. Like in other fields, Muslim scholars at the time also added their own genuine contributions to medicine where they were certainly far ahead of Christian Europe. This included such eminent scholars as Ibn Sīnā and Ibn Rushd (1126 – 1198) known in the West as Avicenna and Averroës. In Muslim ruled Spain, Jewish scholarship prospered as well. Arab and Jewish physicians were in high demand as their practice was educated by learning through observation instead of following scholastic and Catholic Church doctrine by the letter. This certainly benefited their patients, and they took note. In Western Europe, we find Albertus Magnus (c. 1200 – 1280) and his pupil Thomas Aquinas but both of them remained constrained by catholic doctrine and the catholic interpretation of Aristotle's work. Zoology for example still rested on the authority of Pliny's Bestiarium. The first European universities were founded in the late eleventh century, the oldest one in Bologna in 1088. As universities spread over much of Western Europe, Christian theology reigned supreme over their curricula. Invariably, this made the works of Aristotle which the Catholic Church chose to incorporate into its worldview established orthodoxy, not to be contradicted by professors and students alike. For a long time, the Middle Ages have been seen as a dark period. Often, as noted above, they are summarily characterized as Europe's Dark Ages. However, that was not quite so. The later part of the European Middle Ages had their own Renaissance so to speak, sometimes referred to as the Renaissance of the twelfth century with the newly founded universities becoming the seeds for a revival of science. Eventually, this would lead to academic thought gaining some incremental independence from the Catholic Church; but it would take much longer for the leading minds to free themselves from religiously believing in the infallibility of Aristotle's or Galen's authority.

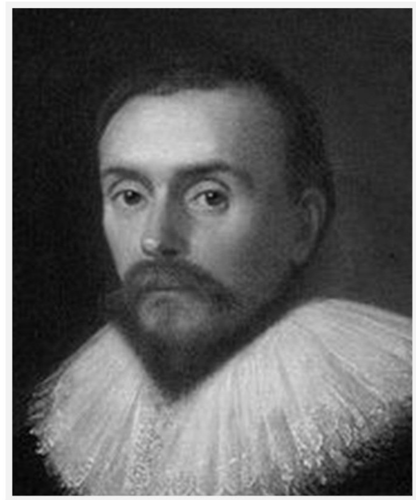
While we still find scholars tied to Aristotle in the fifteenth century, we also see the beginning of real empirical science even if that often came first in the form of alchemy and would not completely separate from it for a long time.[14] The Swiss physician and alchemist Theophrastus von Hohenheim (1493 – 1541), better known as Paracelsus is a good example for the latter.[15] Frequently called the father of toxicology, Paracelsus was a pioneer in exploring relationships between chemistry and medicine which in modern times evolved into what we call today pharmacology. Homology is an important concept in biology. It is based on the observation that the anatomies of different species are related in the sense that for example the fins of a fish, the wings of a bird, and our hands share the same evolutionary origin. When the French naturalist Pierre Belon (1517 – 1564) first recognized this kind of shared ancestry, even though he stated it not quite like that, it nevertheless signified the beginning of comparative anatomy. With Western Europe transitioning from the late Middle Ages to the Renaissance we see a rejuvenation in arts and a new appreciation of the pagan nature of the Greek world. It is then that we find the first detailed drawings of human anatomy, of which Leonardo da Vinci's (1452 – 1519) are likely the most famous examples.



Andreas Vesalius
1514 – 1564

More importantly, in the light of new empirical evidence, scholars increasingly questioned the authority of traditional sources such as Aristotle and Galen. Medical lectures at universities began to include human vivisections, even though the practice remained controversial. Andreas Vesalius, the Flemish physician frequently called the founder of modern human anatomy, published his major work, *De Humani Corporis Fabrica* (*On the Fabric of the Human Body*), in 1543 in Padua, coincidentally in the same year the revolutionary work of Copernicus was first printed. Instead of relying on Galen, Vesalius was relying on his own observations. The unprecedented detailed anatomical drawings in his works corrected many of the major errors he found with Galen's assertions. The mold had been broken. From now on, it was not a sacrilege anymore to contradict Galen or Aristotle on questions of biology or medicine. However, Vesalius did pay a price. The envy of some of his colleagues turned the prejudice that still existed regarding human vivisection against him, practically accusing him of grave robbing. While this marked the end of his research career, his work had also gained him benefactors in very high places including the Holy Roman Emperor Charles V, whose imperial physician he would eventually become.

Erroneous assumptions about blood circulation had been around since Galen's time, upheld by his disciples through the centuries. The Spaniard Miguel Serveto (c. 1511 – 1553), better known under the Latin form of his name as Michael Servetus, understood that blood circulated from the heart to the lungs, back to the heart, and from there to all body parts. However, it would still take a while before the discovery of separate arterial and venal systems transporting blood from and back to the heart. Since Galen's time the assumption was that the lungs served a respiratory function, but the essential lung function remained a mystery and would not become clear until after the discovery of oxygen in the late eighteenth century. Servetus was not only interested in biology but also in theology. Unfortunately, his books on theology were considered heresy and he became one of the many who fell victim to the religious fanaticism and intolerance that were the curse of Europe well into the seventeenth century. The Catholic Church had no monopoly on burning people alive. On his way to Italy, Servetus unwisely stopped in Geneva where his former protestant brethren proved to be quite capable of committing this heinous atrocity too. On October 27, 1553 his former friend and protestant reformer John Calvin (1509 – 1564), Jehan Cauvin in his native French, had him burnt alive atop a stake built from Servetus own books. The second half of the sixteenth century brought finally the first mostly correct understanding of human reproductive biology. The names of the physicians and anatomists associated with this progress are those of the Italian Gabriele Falloppio (1523 – 1562) also known by his Latin name of Fallopius, and his compatriot Girolamo Fabrizio (1537 – 1619) who studied under Fallopius and is better known under his Latinized name as Hyronimus Fabricius.



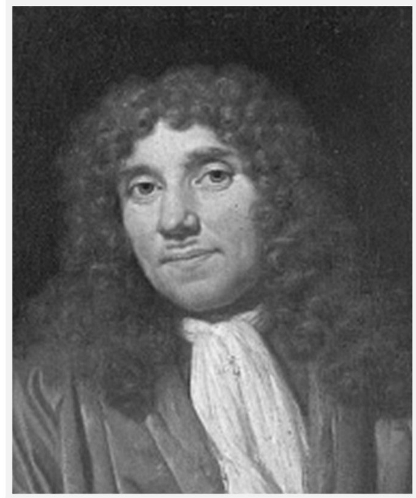
William Harvey
1578 – 1657

The English physician William Harvey, a student of Fabricus, would come to understand that the heart is a muscle, pumping blood into the body through the arteries. He also discovered that blood flowed from the left side of the heart to its right side via the lungs. His acclaimed *Exercitatio Anatomica de Motu Cordis et Sanguinis in Animalibus* (*An Anatomical Study of the Motion of the Heart and of the Blood in Animals*), summarizing his studies of the heart and of blood circulation, was published in 1628. It had always been a mystery what blood really was and how it was made, but clearly it was recognized as vital to life from time immemorial. For centuries, the orthodox opinion had been that the liver produced blood somehow from food which then would flow to the heart, which added to it the *Elan Vital*, the presumed life-giving quality of blood. Harvey simply observed that the quantity of blood was quite constant in human beings and that this would be difficult to maintain if the human blood volume would depend on the often quite erratic food consumption of humans. People seemed not to have less or more blood because they were eating less or more. With this insight, the assumed *Elan Vital* enrichment of the blood seemed not to be necessary anymore. Harvey was also very interested in reproductive biology or embryology, as we would say today. In 1651, Harvey published *Exercitationes de Generatione Animalium* (*Exercises on the Generation of Animals*), summarizing his studies of animal conception and embryo development. In many ways, this work is considered the beginning of modern embryology. His interest in reproductive biology led Harvey to suggest that mammals, including humans, reproduced by male sperm fertilizing female eggs. Harvey essentially believed in the doctrine of *"ex ovo omnia"*, which is the famous Latin phrase stating that all life originates with an egg, or "ovus" in Latin; a concept later to evolve into *Ovist Preformationism*. Mammalian eggs where not known in Harvey's time, and it would take about another two-hundred years before they would eventually be discovered.

The late sixteenth century sees the first classification efforts, essentially the beginning of taxonomy but that word did not exist back then. The Italian physician and botanist Andrea Cesalpino (1519 – 1603) published his sixteen-book treatise *De Plantis* (*About Plants*) in 1583. It is considered the first textbook of botany and it describes and classifies more than 1,500 plants. The Swiss botanist Gaspard Bauhin (1560 – 1624) first published his plant descriptions in his *Phytopinax* (*Plant Images*) 1596 and followed it with his *Pinax Theatri Botanici* (*Theater of Botanical Images*) published in 1623, which included all the plants previously described by others that he could get hold off; the total came to more than six thousand plants. The English naturalist John Ray (1627 – 1705) contributed to taxonomy in botany as well as zoology. In zoology his *Synopsis Animalium Quadrupedem et Serpentini* (*A Survey of Qaudrupeds and Reptiles*) published in 1693 still reflected much of the old Aristotelian classification scheme, although he started to diverge from orthodoxy in some areas. Ray's impact on the development of botanical taxonomy was more significant and longer lasting as over the next generations this work influenced scholars who would build the foundations of modern taxonomy. He was first to suggest a species concept as the fundamental unit of classification for lifeforms and his belief was that it reflected *Divine Creation*. In his religiously informed perspective, taxonomy was a systematic to be discovered in *Divine Creation*, not something man could create on his own. According to Ray's definition, the distinguishing characteristic of a biological species was that a species always reproduced itself as the same species; he saw species as fixed and permanent but allowed for the possibility of variation within a species. Defining species as being fixed and permanent was in agreement with the assumptions of *Divine Creation* that all species that existed were divinely created at one point in time in the distant past. In 1686, John Ray published his *Historia Generalis Plantarum* (*General History of Plants*) where he classified a large number of plants, close to 19,000 of them, according to their similarities and differences. One of the last great achievements before microscopy revolutionized biological research was the almost modern description of the female reproductive system by the Dutch Regnier de Graaf (1641 – 1673) who built on the works of Fallopius and Fabricius.

The invention of the microscope in the last years of the sixteenth or the first few years of the seventeenth century represent a true watershed for biology and medicine.[16] Once the first modern microscopes were available, a completely new word became visible and that very much would help to correct some long-standing misconceptions. One year after Harvey had died, the Dutch scholar Jan Swammerdam (1637 – 1680) conjectured, based on his microscopy studies, that blood was composed of cells. The seventeenth century also sees the first comprehensive descriptions of plant and insect life. Swammerdam was a pioneer in both areas, shedding light on the life cycle of insects from egg, to larva, to pupa, to adult insect. Spontaneous creation from rotten meat was a long-standing myth. While sounding strange today, it seemed like a natural explanation for flies emerging spontaneously from rotten meat. Already William Harvey had denounced spontaneous creation in his studies of reproductive biology, asserting his belief that all lifeforms, including even maggots and worms, originate from eggs. However, it was the Italian physician Francesco Redi (1626 – 1697), who experimentally disproved that rotten meat spontaneously created flies. At first, it looked as if it was rather the maggots that were spontaneously created, and not the flies. However, he correctly suspected that it was flies to start with, as they were initially dropping something into the meat that would evolve into maggots, which eventually turned into flies. To verify this, he conducted a controlled experiment where he put one piece of meat into a sealed container and another piece of meat in an open container. What he observed was that there were no maggots or flies emerging from the meat in the sealed container, but the usual process happened to the meat exposed to the environment. This may have been one of the very first controlled biological experiments ever undertaken. Redi's conclusion was clear with regard to spontaneous creation from rotten meat: it did not exist. However, unlike Harvey, he continued to believe in spontaneous creation in other areas.

Ornithology, the branch of zoology that concerns the study of birds, also got its scientific start in the seventeenth century with *Ornithologiae Libri Tres* (*The Ornithology in Three Books*), the work of the English scholar Francis Willughby (1635 – 1672), published posthumously in 1676. With it, Willughby pioneered classification of bird species according to their physical characteristics, a first in ornithological taxonomy.



Antonie van Leeuwenhoek
1632 – 1723

Another development in the seventeenth century were the early beginnings of paleontology. We have already met Nicolas Steno in the last chapter as one of the founders of modern geology. His fossil studies we mentioned there also make him one of the founders of paleontology, the study of life roughly up to the Holocene epoch (see plate VII). However, towards the end of the seventeenth century, it was the discovery of the microscopic world that took center stage and the significant discoveries about lifeforms of the distant past would only come later. The Dutch scientist and tradesman Antonie Philips van Leeuwenhoek may not have invented the microscope, but he very much improved it. Importantly, he used it to measure everything interesting he could get hold of that was too tiny for his eyes to resolve and made numerous observations adding to different areas in biology. His greatest insight was that an enormous amount of life surrounds us that is just too small for the unaided eye to see. The discovery of bacteria ranks as one of the greatest discoveries in biology ever made. In Leeuwenhoek's time, the term bacteria was unknown as it would only be introduced in 1828. But bacteria were clearly first observed by Leeuwenhoek in 1676 who published his discovery in a series of letters to the Royal Society of London. He summarized his study of microbial life or as he called it, the study of animalcules, Latin for very tiny animals, in *Arcana Naturae Detecta* (*Nature's Mysteries Disclosed*), published in 1695.

It is nigh impossible to talk about the beginnings of microscopy, when science finally began to lift the veil hiding until then the world of the invisibly small from human perception, without mentioning Robert Hook (1635 – 1703), the English natural philosopher or physicist as we would say today. Robert Hook made many contributions to science and he was a man of diverse interests, including microscopy. His *Micrographia* (*Micrographia: Or Some Physiological Descriptions Of Minute Bodies Made By Magnifying Glasses*), published in 1665, is a marvelous book. This work opened up completely new perspectives on life and of life as never seen before in this detail. It summarizes Hook's microscopy observations as well as some of his work in other areas and it has been specifically acclaimed for its detailed illustrations of insects and plants. It includes the first picture of a facetted insect eye and famously it includes the first accurate description of a cell. While Hook was the first one to describe a cell and in the process gave the cell its very name, his contributions in this area do not seem to go beyond that. Robert Hook was not the only one to produce marvelous drawings of nature thence unseen.

This applies no less to the most wonderful work of the German-born naturalist Maria Sibylla Merian. She made her observations of the insect and plant world directly in nature and then transferred it onto paper in what became truly the very first accurate scientific illustrations of many species. This includes a clear understanding and depiction of their different metamorphoses from caterpillar to mature adult insect, marvelously captured in outstandingly beautiful and detailed watercolors.

Maria Sibylla Merian
1647 – 1717

Merian published her two-volume book *Der Raupen wunderbare Verwandlung und sonderbare Blumennahrung* (*The Wonderful Metamorphism of Caterpillars and Strange Floral Nourishments*) in 1679 and 1683, respectively. Her insights in this area are truly remarkable considering that at the time there were still many who held the view that spontaneous creation provided a satisfying explanation as to where insects come from; essentially believing that insects emerged so-to-say fully formed from mud. Merian not only brought into focus the insect world of her native European lands, she also set out on expeditions to South America, vastly expanding our knowledge of the insect world in this then far away continent with her 1705 publication of *Metamorphosis Insectorum Surinamensium* (*Metamorphism of Insects in Surinam*). A remarkable achievement in any time, even more so as she did this as a female artist and scientist. When it was seemingly still customary to perceive women as inferior to men in almost all respects and women were discriminated in countless ways, she accomplished what very few men in her time would have been able to achieve.[17]

The Italian physician Marcello Malpighi (1628 – 1694) is sometimes referred to as the father of microscopical anatomy, physiology, and embryology. Another early user of microscopy, he made a number of first observations and some physiological features were eventually named after him. Naming him father of embryology is a stretch at best because Malpighi was a proponent of the doctrine of preformation. This doctrine stated that organisms are fully formed from the moment the organism was created. It asserted that all organisms including humans were created at the same time, and that succeeding generations of humans grow from so-called homunculi, Latin for little persons, essentially representations of small human beings assumed to have existed since the beginning of creation. With respect to his views on embryology, Malpighi clearly ended up on the wrong side of biology and history. More important than his work in the area of embryology was his work on plants. Malpighi's *Anatome Plantarum* (*Plant Anatomy*) published in 1675 is credited as one of the first significant treatises on plant anatomy as opposed to almost all other such work before, which had been focused on animals.

The English botanist Nehemiah Grew (1641 – 1712) was a contemporary of Malpighi and a no less skilled microscopist. He published *The Anatomy of Plants*, which he had begun earlier under the title *Philosophical History of Plants*, in 1682. It was the first time the inner structure and function of plants was presented in minute detail. He gave accurate descriptions of plant anatomy and also hypothesized that some plants may be hermaphrodites, combining both sexes on a single plant. Influenced by his work was the German botanist Rudolf Jakob Camerarius (1665 – 1721), or Camerer by his German last name. Camerer performed field experiments to study plant reproduction by cutting off parts of a plant. When such a plant produced degenerate seeds which did not develop correctly, Camerer concluded that he had cut off a part of the plant's reproductive physiology. His *De Sexu Plantarum Epistolam* (*Letter on the Sex of Plants*) appeared in 1694.

Antoine-Laurent de Lavoisier
1743 – 1794

With it, he was first to identify plant pollen and ovary as the male and female parts of a plant. The English clergyman and scholar Stephen Hales (1677 – 1761) published his *Statistical Essays* concerning plant physiology in 1727 and for animal physiology in 1733. The first volume, *Statistical Essays Containing Vegetable Statistics*, summarized his studies on the transpiration of plants, how water circulates through a plant. Hales experiments also led him to the conclusion that plants are interacting with the air through their stem and leave systems. Interestingly, he theorized that plants might draw on light for their growth. The second volume, *Statistical Essays Containing Hæmastatics*, is the first work reporting blood pressure measurements. In the eighteenth century, scholars continued to build an ever-encompassing encyclopedic description of insect, plant, and animal life. This included the French scholar and entomologist René Antoine Ferchault de Réaumur (1683 – 1757) and the Swiss naturalist Charles Bonnet (1720 – 1793). Through the course of the seventeenth century, entomology, the branch of biology devoted to the study of insects had become an established science. While Réaumur may be better known for the temperature scale he created his work in entomology is likely more lasting. His six-volume *Mémoires pour servir à l'Histoire des Insects* (*Studies on the History of Insects*), published between 1734 and 1742, summarized his studies of insect anatomy and continues to be held in high regard. In addition, Réaumur contributed to an improved understanding of digestion by showing that it is not merely a mechanical process, as was the traditional view, but also a chemical process. Bonnet's 1745 publication of his *Traité d'Insectologie* (*Treatise on Insects*) marks another milestone in entomology. His work in botany focused on plant leaves and it eventually led him to the conclusion that plants can sense their environments.

Towards the end of the eighteenth century the role of oxygen was finally discovered. It was the French chemist Antoine-Laurent de Lavoisier, one of the pre-eminent scientists of the eighteenth century and a founding father of modern chemistry who recognized oxygen for what it is and also gave it its name. The British scientist Joseph Priestley (1733 – 1804) had conducted earlier experiments that demonstrated the existence of a substance like oxygen but he could not free himself from the existing orthodoxy of the so-called phlogiston theory. Simply put, the latter explained burning by the liberation of a substance called phlogiston, which translated from ancient Greek means "burning". Interestingly, in this theory, corrosion was seen as a loss of phlogiston. Lavoisier learned about some of Priestley's work and followed up with his own experiments coming to a different conclusion than Priestley. He understood that he had found a new chemical element that provided a natural explanation for the observed effects. Therefore, it eventually was Lavoisier who effectively discovered oxygen and not Priestley. Lavoisier summarized all of his discoveries in his *Traité Élémentaire de Chimie* (*Elementary Treatise on Chemistry*) published in 1789. Understanding the role oxygen plays in burning and oxidation was fundamental to eventually discovering its role in biology and evolution.

Carl Linnaeus
1707 – 1778

Sadly, Lavoisier, one of the great scientists of his age, had the misfortune to live in France during the *French Revolution* and was guillotined 1794. It takes centuries to produce men or women with a head for science like Lavoisier had it; but it took only the sharp edge of a guillotine to take it off in less than a second. The Swedish physician Carl Linnaeus, Carolus Linnaeus with his Latinized name, devoted his life to the taxonomy of botany and zoology. A giant of eighteenth-century taxonomy, he redefined and changed it in many ways that still are very much reflected in modern taxonomy. He introduced his classification system with his *Systema Naturae* (*Systems of Nature*) published in 1735. Linnaeus divided nature into three - animals, minerals, and vegetables - subdividing animals and vegetables further into class, genus, and species. Animals, for example, were grouped into six classes: mammals, birds, amphibians, fish, insects, and a catch-all-other class with the Latin name *vermes.*[18] Linnaeus was not the first to use a binomial system classifying a species with two names. Gaspard Bauhin for example had started using such a scheme more than a century earlier. However, due to the additional amount of botanical and zoological data that was available in Linnaeus time, he could base his system on much firmer ground. In a sense, it was much clearer in his time how different characteristics of organisms related to each other, what the commonalities were that identified a species, and which ones were more primitive or advanced than others. The two names that classify a lifeform in Linnaeus binomial system are the generic name, the genus name of a lifeform, and the specific name, the species name of a lifeform. The convention is that the genus name starts with a capital letter while species names are written in lower case, always. As an example, we can look at ourselves, *Homo sapiens*, with *Homo* (human) being our genus name and *sapiens* (wise) being our species name. Two years after introducing his classification scheme for animals, Linnaeus published *Genera Plantarum* (*Genera of Plants*) in which he applied it to botany, classifying over 18,000 types of plants. Today's plant and animal nomenclature traces its origin back to the taxonomy principles Linnaeus published in his *Species Plantarum* (*The Species of Plants*) in 1753. As scientific as Linnaeus proceeded with the classification of plant and animal life, one would have expected him to include humans in a similar way. Far from it, Linnaeus classified humans by emotional state, which when it came to non-Europeans pretty much reflected the racist biases of his time. Another constraint on his classification of humans seems to have been his belief in *Divine Creation*.

Others during his lifetime and in the generation coming after him added to or improved Linnaeus system. They include the Swedish naturalist and physician Peter Artedi (1705 – 1735) or with his Latinized name Petrus Arctaedius, a friend of Linnaeus; the French botanist Antoine Laurent de Jussieu (1748 – 1836); and the French naturalist and zoologist Jean Léopold Nicolas Frédéric, Baron Cuvier (1769 – 1832), better known as Georges Cuvier or Baron Cuvier. Artedi is frequently called the father of ichthyology, the branch of biology devoted to the study of fish, a title he earned with his *Ichthyologia,* *Sive Opera Omnia de Piscibus* (*Ichthyologia, the Complete Work about Fish*) which was posthumously edited and published by his friend Linnaeus in 1738. Jussieu, an eminent botanist, modified some of Linnaeus classifications and was also influenced by John Ray. He was the first to provide a classification of flowering plants in his *Genera Plantarum* (*Plant Genera*) in 1789. Cuvier contributed to taxonomy in both areas, zoology and botany, and is mostly known today for his comparative anatomical studies. His major work in this area, *Le Règne animal distribué d'après son organisation* (*The Animal Kingdom Arranged after its Organization*) was published in 1817.[19] Cuvier's approach to taxonomy subdivided the animal kingdom into *Vertebrata*, *Mollusca*, *Articulata*, and *Radiata*. The *Vertebrata*, or vertebrates, included mammals, birds, reptiles, and fish; the *Mollusca* included among others snails, clams, oysters, but also cuttlefish; *Articulata* are all animals with segmented bodies such as insects or spiders, and *Radiata* includes animals with radially symmetric body plans such as starfish or polyps. Curvier was also an early fossil hunter and so eventually had to confront the question of evolution. His spiritual worldview led him to become one of the major proponents of *Catastrophism*, the presumption that life on Earth was periodically wiped out, only to be completely replaced with new life by another *Divine Creation* event. He put this view forward in his 1812 publication of *Recherches sur les Ossements Fossiles de quadrupeds* (*Researches on Quadruped Fossil Bones*). As we have seen in the last chapter, in the mid-nineteenth century *Catastrophism* started to lose ground to the theory of *Uniformitarianism*. One can only wonder what the response of a man of Cuvier's scholarly stature would have been to the concept of evolution, which would eventually put *Catastrophism* out of business for good.

Among the significant contributors to the development of taxonomy at the time was also Comte de Buffon whom we already met in the last chapter. His encyclopedic work, *Histoire Naturelle, Générale et Particulière* (*General and Particular Natural History*) which appeared in thirty-six volumes between 1749 and 1788 familiarized much of educated Europe with the natural world as it was then seen and discussed by scholars in the eighteenth century. In it, Buffon also considered openly for the first time the similarities between humans and apes, which of course offended some, such as the faculty of theology at the university in Paris, the Sorbonne. While that did not intimidate him, he ultimately rejected the possibility of a common descent of humans and apes. Buffon was also the first to propose a species concept that was based on the ability to procreate. It states that two animals should be considered as belonging to separate species if they cannot produce fertile offspring.

Muscle physiology, the nervous system as well as the study of tissue at the cellular level – a branch of biology and medicine referred to as histology - was led by scholars such as the Italian physician Luigi Galvani (1737 – 1798), the Swiss Albrecht von Haller (1708 – 1777), and the French Marie François Xavier Bichat (1771 – 1802). Galvani's discovery of animal electricity made him the founder of bioelectricity. Haller was an eminent physiologist exploring the human nervous system and the first one to refer to the nervous system as "irritable". Bichat created a tissue classification system for the organs of the human body distinguishing eventually twenty-one tissue types, which he described in his *Anatomie générale* (*General Anatomy*) published in 1801. Because of his groundbreaking work in this area, Bichat became known as the father of histology.

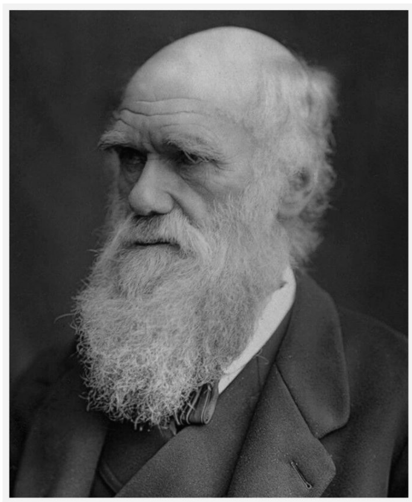
From Biological Evolution to Genetics

During the eighteenth century the nascent science of chemistry started to have a real impact on medicine and biology. Progress in medicine and biology remained closely linked, however, the groundwork for biology to start establishing itself as its own discipline was laid back then. The nineteenth century brought even more rapid advances in chemistry but also new physical instrumentation never available before. This enabled biology and medicine to make substantial leaps forward. Along the way, new research areas evolved that were now distinctly perceived as belonging to biology. Throughout the history of biological and medical thought a number of contentious topics that were at the very heart of biology kept surfacing and were intensely debated. In the eighteenth century, these debates crested in different ways. Some of them were adjudicated by the beginning of the nineteenth century and others were now much more scientifically interrogated than was the case before. Among them was the idea of spontaneous creation, which we already encountered. The other major controversy was the debate between *Ovists* and *Spermatists*, at the root of which lay the idea of preformation. *Ovists* held the view that the intact organism lived in the egg, "ovus" in Latin, and *Spermatists* thought that the completely formed organism is contained within the male sperm. Then, there was of course the debate about the belief of *Divine Creation* with its unrelenting assertion that all life as it exists had been divinely created at one point in time. The idea of spontaneous creation would be laid to rest starting with the Italian biologist and priest Lazarro Spallanzani (1729 – 1799). His carefully conducted and exhaustive experiments led him to conclude that spontaneous creation did not exist. However, it would take almost another century until the French biologist Louis Pasteur (1822 – 1895) finally disproved the doctrine of spontaneous generation for good. The debates between *Ovists* and *Spermatists* as well as the controversy about the doctrine of preformation would be settled in the nineteenth century through a two-pronged approach. Via the development of evolutionary concepts and discoveries in cellular biology on the one side and by the beginnings of the modern science of genetics on the other side. What makes this so interesting is that the theory of the evolution of life was about to be conceived without knowing the causes for the underlying fundamental forces driving it in the first place. It was known for some time that a large number of obviously extinct species must have once populated Earth, as was evident from their fossil relics. With an increasing number of new fossils being dug up it became ever clearer that there must have been a very great many species once living on Earth, none of which had survived. Why so many species had become extinct and how new species evolved was a complete mystery. As we have seen earlier, the explanation based on *Catastrophism* still had its forceful defenders who, motivated by their religious convictions, continued to argue for it to reconcile the Book of Genesis with the facts at hand. However, this became increasingly untenable. So how do new species arise and why do species become extinct? Or, to ask the ultimate question, why are there species at all, how did life come to be in the first place? Evolution would provide an answer for why there are species at all, and it would allow us to understand what drives the evolution of species and where all life on Earth comes from. However, there is no answer yet as to how life started in the first place. Today we know so much more about how life could have started on Earth, but the final answer still eludes us.

Jean-Baptiste Lamarck
1744 – 1829

The name of the French naturalist Jean-Baptiste Lamarck is probably familiar to many and more likely than not, most of us will associate his name with a quickly discarded evolutionary theory. However, the latter is not quite true. The theory of evolution that Lamarck formulated in his *Philosophie Zoologique* (*Zoological Philosophy, or Exposition with Regard to the Natural History of Animals*) published in 1809, introduced the concept of inheritance of acquired characteristics. According to Lamarck, two major forces are driving the evolution of species. First there is the tendency of organisms to become ever more complex over time; and second, there are powerful environmental forces requiring life to adapt. Consequently, he argued that animals would adapt in such a way that the use or disuse of any specific organ or body part would drive the further evolution of that organ or body part. This part of Lamarck's evolutionary theory has often been ridiculed. To mind comes the famous example of giraffes having long necks because they continuously stretch to reach the juiciest leaves high in the top of trees that no other ground based animal could get to. However, if one looks at the use or disuse of body parts over the long-term evolution of a population, Lamarck had actually a good explanation. Just take as an example the ancestors of whales. They once walked the land on four legs before returning to the ocean to evolve eventually into fully-fledged marine animals again, just as their distant ancestors once were. As marine animals, they had no use for legs anymore and therefore today's whales have no hind-legs and only retain the evolutionary remnants of a pelvis. This clearly was an example of body parts lost because of disuse. Lamarck's approach also recognized the importance of adaptation to local environments; and it provided an explanation for evolutionary dead-ends in the case of highly specialized adaptations.

Lamarck was not the only one in his time to think about evolution in terms of transmutation of species or transformism. Lamarck introduced the term transformism, derived from the French word *transformisme*, but it was a concept widely discussed well before his time. This prominently includes the English physician and natural philosopher Erasmus Darwin (1731 – 1802), Charles Darwin's grandfather. Incidentally, some of the ideas that often are associated with Lamarck pre-date his work as evidenced by the publication of Erasmus Darwin's *Zoonomia* in 1794, parts of which advocated the inheritance of acquired characteristics. Once Lamarck's publications had spread from the European continent to the British Isles, it was clear to many of Erasmus Darwin's compatriots that he had earlier articulated similar views and so in some places he became hailed as the English Lamarck. In a sense, the late eighteenth century was pregnant with what we call now Lamarckian ideas. Therefore, when many today joke about Lamarck's views on evolution and some even ridicule them, they should consider that he really was only the most prominent and most articulate natural philosopher at the time to consider concepts of transmutation or transformism. What was difficult to accept for those how came after Lamarck was that he pictured evolution happening on the time scale of an individual.



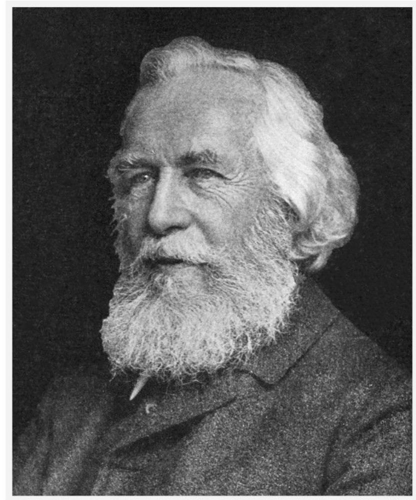
Charles Darwin
1809 – 1882

Crucially, he also proclaimed adaptations acquired by an individual organism to be heritable. With new insights from genetics, our understanding of what is heritable and what not keeps changing. So interestingly, after being ridiculed for about two hundred years, it is exactly this viewpoint, which in the second half of the last century has received more attention. As we will see later, some adaptations of an individual acquired through its lifetime may indeed be passed on to the next generation but in a manner that Lamarck could not have imagined. Part of Lamarck's views on evolution were clearly ahead of his time while others, even though scientifically argued, were wrong and still others were not scientific at all but vestiges of a pre-scientific past that Lamarck was still beholden to. Charles Darwin would take some inspiration from Lamarck, in the sense that Lamarck tried to explain evolution through laws of nature. Maybe it is more appropriate to say that Darwin took inspiration from his grandfather's views on the inheritance of acquired characteristics with which he must have been quite familiar before he ever got to studying Lamarck's ideas in the first place. Whichever is the case, Darwin would indeed allow for a role of Lamarck's use and disuse of organ and body parts in his own evolutionary theory. However, he did so reluctantly and then only as a secondary evolutionary mechanism and pretty much restricted to explaining for example the loss of organs or other body parts such as the one we mentioned before in whales. Throughout his decade-long struggle to formulate a coherent theory of evolution, Darwin would remain adamant in distancing himself from Lamarck's theory of transformism. Specifically, when to his chagrin even some of his closest admirers and supporters sometimes made the mistake of confounding Darwin's theory of evolution with Lamarck's earlier effort.

Charles Darwin's *The Origin of Species*, published in 1859, represents a watershed not just for biology or paleontology but for our own history as a species. There is a before and after in the sense that before, a concept like *Divine Creation* could presume religious belief to be the equivalent of scientific knowledge. After Darwin, this would not be possible anymore and religion would no longer be in a position to prescribe, at least not to most of us, what scientific facts have to look like in order to comply with religious orthodoxy. As we have seen in the last chapter, Darwin was very much familiar with the developments in the field of geology in his time, specifically with the theory of *Uniformitarianism*, of which his friend and mentor Charles Lyell was a major proponent. It was the vast geological times scales that provided the canvas and the required depth in time for an evolutionary process as proposed by Darwin to play out. Early in his career, Darwin may well have seen himself as more of an aspiring geologist than anything else. He was quite proud of his geological exploits during the five-year voyage of the HMS Beagle around the world. How he initially weighed them and their impact on others as compared to his just beginning work on species upon returning from his five-year adventure speaks to that. It took more than two decades of intense research and experimentation before Darwin would publish his unique views on species evolution. The full title of Darwin's first work on species evolution reads: *On the Origin of Species by Means of Natural Selection or the Preservation of Favored Races in the Struggle for Life.* It spells out what in Darwin's view provided the answer to the question of why species do rise and fall: natural selection. Only in later revised editions towards the end of his life would Darwin use the phrase "survival of the fittest". A change that friends and colleagues encouraged him to adopt from his fellow British naturalist and philosopher Herbert Spencer (1820 – 1903) who had coined the phrase in his 1864 work *Principles of Biology.* At the time he made the change, survival of the fittest had not yet acquired the negative sociobiological connotations it carries today; it is doubtful if Darwin would have made the change if he could have known what this phrase would eventually come to signify in a negative sense. To some extent, Darwin's concept of natural selection reflects underlying thought currents in his time, particularly regarding population dynamics as reflected in the writings of his fellow countryman Thomas R. Malthus (1766 – 1834). Darwin observed that parents of any given species produce more offspring than what is required to sustain the species. In the struggle for life, the competition for resources to sustain themselves, only those best equipped for their respective environments will survive. Some may be stronger or faster than others, have better eyesight or hearing, or are better camouflaged to protect them from predators. However, how those qualities emerge that make some better equipped than others in order to come out on top in this struggle for life seems to be more due to statistical variance of these qualities within a given population, thereby making success in the struggle for life somewhat of a chance event. Much of that would have been lines of thought in which many in Victorian England could follow Darwin. They must have been familiar with the struggle for life that they could see happening in front of their doorsteps and in their city streets under the selective pressure of early capitalism, in the midst of the epochal transition from an agrarian to an industrial society.

However, many of them would struggle with the propositions Darwin put forward in *The Descent of Man, And Selection in Relation to Sex*, which appeared in 1871. Clearly, the idea that Man had evolved from earlier lifeforms and shared a common ancestry with apes was not palatable for most holding cherished religious views. This was also not easy to accept for many holding more liberal views. Darwin's concept of evolution makes Man truly a part of the animal world, hard to swallow for those continuing to believe in Man as a *Divine Creation*. More importantly, Darwin put forward another mechanism in addition to natural selection: sexual selection. The latter is not the struggle for existence in relation to other organic beings or as it relates to external conditions but for the right to procreate. Only those able and successful to mate will procreate and have offspring. Darwin was not alone in developing concepts of evolution based on natural selection. We now know that the English naturalist Alfred Russel Wallace (1823 – 1913) came up with the main ideas of evolution independently of Darwin. Darwin always acknowledged that natural selection was as much a brainchild of Wallace as of himself. The difference between Darwin and Wallace is that Darwin was first to provide his thoughts on the concept of evolution in a much more comprehensive and detailed account to a broader public; and that he continued to expand on it. While Wallace would be one of the first to replace in his views natural selection with survival of the fittest and asked Darwin to do the same, he himself would eventually take a very different view on the origin of humankind from

Darwin. To Wallace, some kind of creator remained indispensable as he saw humankind as created differently from the rest of nature. Darwin would have none of that. Darwin and Wallace were certainly not the only ones at the time thinking about evolution. In a sense, the late eighteenth and early nineteenth century had already conceived various concepts of evolution but some of them were stillborn such as *Catastrophism* or seemed to be even more far-fetched than what Darwin proposed; like the ideas on evolution that the French naturalist Jean-Baptiste Lamarck had advocated. Biological evolution was an idea whose time had come, and it was part of the broader transition of human societies into modernity. Darwin's friends and colleagues whose support was critical for the acceptance of evolutionary thoughts in Victorian Britain are a good example for that.[20]
A critical factor in this transition were the insights scientists gained into the development of life itself. These finally laid to rest the dispute between *Ovists* and *Spermatists*, and with it the doctrine of preformation. In cell biology we see the discovery of cell division and chromosomes: mitosis, resulting in two cells identical to the parent cell with the same number of chromosomes; and meiosis, resulting in four daughter cells that have half of the parent cell chromosomes.[21] It also became clear that chromosomes somehow were related to heredity and they started to be referred to as vectors of heredity. Among the biologists and physicians associated with these discoveries are the Germans Walter Flemming (1843 – 1905) and Theodor Boveri (1862 – 1915), the Belgian Édouard van Beneden (1846 – 1910), and the American Walter Sutton (1877 – 1916). The Swiss anatomist Rudolf Albert von Kölliker (1817 – 1905) and his erstwhile student, the German anatomist Karl Gegenbaur (1826 – 1903), would come to understand that the egg of a vertebrate is a cell in itself and that all cells in any organism derived from an egg are the result of cell divisions materializing from and starting with the egg.
With the Baltic-German-Russian Karl Ernst von Baer (1792 – 1876) we see the first appearance of what is called the *Biogenetic Law*. This pronouncement contends that in the phases of embryonic growth the general character of an organism appears before the more specialized one. This is sometimes more loosely described as the embryonic development of higher animals passing through stages that resemble the embryonic development phases of lower animals. In 1817, the Baltic-German biologist Heinz Christian Pander (1794 – 1865) described the early development of chicken embryos in terms of what today are known as the primary germ-layers, the cell layers that form during the embryonic development: the endoderm, the ectoderm, and the mesoderm.[22] Baer built on this discovery of his friend, although he thought there were four germ-layers, and laid down the foundations of what is now referred to as comparative embryology. The German physician Robert Remak (1815 – 1865) confirmed that there were three primary germ-layers in a mammalian egg and conjectured that the ectoderm would give rise to the skin and the nervous system; the musculature would develop out of the mesoderm; and the digestive track be derived from the endoderm. We know today that the list of body parts evolving from these layers is longer but the basic insight remains, all parts of vertebrate organisms trace back to the three germ-layers. With increasing insights into the embryonic development of organisms, it became clear that resemblances in embryonic development of different species could help reveal how different animal groups are evolutionary connected. Darwin certainly understood that it may be possible to see homologies at the embryonic developmental stages that would not be visible anymore in the adult organism.



Ernst Haeckel
1834 – 1919

In the late nineteenth century, this research was referred to as evolutionary morphology because it explores relationships between evolutionary structures in embryonic development. Today we call it evolutionary developmental biology or often just Evo-Devo. The foremost biologists associated with this development were the Germans Fritz Müller (1821 – 1897) and Ernst Haeckel, the Polish-Russian Aleksander Kowalewski (1840 – 1901), and the British Francis M. Balfour (1851 – 1882). All of them were ardent supporters of Darwin's concept of evolution. In 1866, Kowalewski found the same basic three germ-layers also in invertebrates showing that they shared a common lineage with vertebrates sometime in the distant past. Müller and Balfour used embryonic homologies to search for shared ancestry among animals, both being convinced that natural selection applied as much to the early stages of development, the embryonic phases, as it did to the adult stages of animals. Haeckel embraced the evolutionary concepts developed by Darwin early on. The idea that an embryo goes through phases resembling successive stages in the evolution of the animal's ancestor was not original to Haeckel. It had been around since the early nineteenth century. Haeckel popularized the concept and tried to reconcile it with Darwin's evolution. The origination and development of an organism, usually from fertilization until an organism reaches its mature form, is referred to as *ontogenesis*.[23] *Phylogenesis* refers to the development and diversification of a species, the evolutionary tree of life is essentially a *phylogenetic* tree depicting the evolutionary relationships between species.[24] Haeckel's proposition, his version of the *Biogenetic Law*, was that *ontogenesis* equals *phylogenesis*. To illustrate this, if this would indeed be the case, a bird embryo would have to recapitulate the phylogeny of fish and amphibian before entering the reptilian and bird stages of its embryonic development.

We know today that Haeckel was not correct though some of his ideas were retained by modern evolutionary developmental biology. While embryos do reflect the course of evolution, they do so in a much more complicated and more convolved way than Haeckel thought. The *Biogenetic Law* Haeckel believed in also very much influenced his vision of the tree of life which reflected a steady linear progression from bottom to top. Haeckel was not the first to depict a tree of life just like that but his trees of life imprinted on others very successfully. Fig. 3.2 shows an example of such a tree published in his 1866 work *Generelle Morphologie der Organismen* (*General Morphology of Organisms*). The tree also shows a new taxonomic group which Haeckel introduced that same year, the kingdom of protista.[25] Under the name protista, Haeckel grouped a diverse collection of organisms almost all of them microscopic and single-celled organisms with a cell nucleus and an assortment of specialized functional cell parts generally referred to as organelles. For many of us, likely familiar examples of protista are amoebas and single-celled algae. Today, protista are not a taxonomic category anymore as biology has developed a much more differentiated view since then, but the category name is still widely used. Most of the former protista fall into the kingdom of *eukarya*, single-celled organisms with a nucleus.

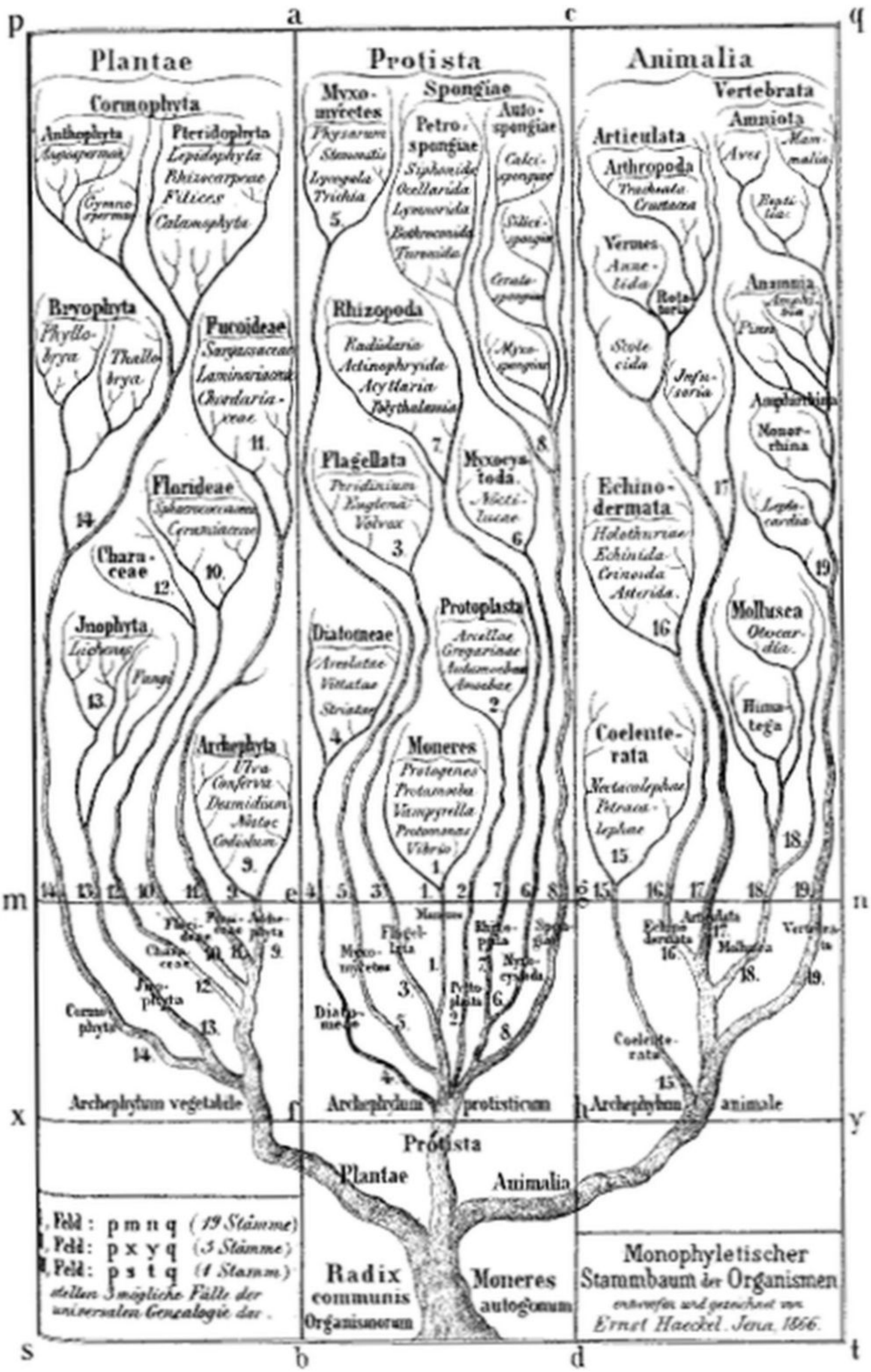


Figure 3.2: One of Ernst Haeckel's many trees of life. The example shown here was published in Haeckel's *Generelle Morphologie der Organismen* (*General Morphology of Organisms*) in 1866.

We will look into eukarya in more detail in the next section.[26] In the nineteenth century, cytology, cell research in today's context, evolved to become a distinct branch of biology.[27] Progress in both, animal as well as plant cytology, was greatly aided by continued improvements in microscopy. The cell nucleus was observed already by Leeuwenhoek in 1682, first described by the Austrian microscopist Franz Bauer (1758 – 1840) in 1802, and first recognized for what it is and properly named in 1831 by the Scottish botanist Robert Brown (1773 – 1858).[28] In 1839, the German biologist Theodor Schwann (1810 – 1882) published his seminal work *Mikroskopische Untersuchungen über die Übereinstimmung in der Struktur und dem Wachstum der Tiere und Pflanzen* (*Microscopical Researches into the Accordance in the Structure and Growth of Animals and Plants*). In it he concluded that all livings things are composed of cells and cell products. The other founder of nineteenth-century cell theory was Schwann's compatriot, the botanist Matthias Schleiden (1804 – 1881). He realized the significance of the discovery of the cell nucleus by Brown. His 1842 publication of the textbook *Grundzüge der wissenschaftlichen Botanik* (*Principles of Scientific Botany*) illustrates the progress that had been made in botanical research which up till then often had been a distant second when compared to zoology; not anymore. The microscope was critical to the combined efforts of Schwann and Schleiden. They benefited in this respect significantly from having one of the leading optics manufacturers in their backyard. The firm Carl Zeiss, founded by Carl Zeiss (1816 – 1888), one of the best optical instrument makers of the time, was very responsive to their needs for improved microscopes.

The discovery of nucleic acids, the building blocks of life, although that was not clear at the time, was published in 1871 by the Swiss physician Friedrich Miescher (1844 – 1895). At the end of the nineteenth century, the five organic compounds that represent the letters used to encode the blueprints of life in deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) molecules were isolated by the German biochemist Albrecht Kossel (1853 – 1927). It would still be a long way from there to actually understanding the structure of DNA molecules and eventually determining the sequence of nucleotide base pairs that make up DNA. The last decade of the nineteenth century saw the birth of virology with the discovery of the first virus. In 1892, the Russian botanist Dmitri Iosifovich Ivanovsky (1864 – 1920) identified the tobacco mosaic virus, a discovery the Dutch microbiologist Martinus Willem Beijerinck (1851 – 1931) confirmed in 1898. It was Beijerinck, who then coined the designation virus for the newly discovered infectious agent.[29]

While scientists were busy learning the bio-chemical language of life they were oblivious to the groundbreaking work of the Austrian monk Gregor Mendel (1822 – 1884) in his monastery in Brünn, today's Brno in Moravia, Czechia. Mendel used several characteristics of garden pea plants to understand how these were passed on over successive plant generations. This led him to the discovery of what he called recessive and dominant traits. A concept that is still very much with us as such recessive and dominant traits in human biology are responsible for our eye or hair colors, just to give two examples. In today's language, we would express Mendel's findings by saying that if both parents had the dominant gene the offspring would express the dominant gene; if one parent carried the recessive and the other the dominant gene, the dominant gene would be expressed; the recessive gene would only be expressed in offspring if both parents carried it.

Gregor Mendel
1822 – 1884

Mendel's work remained undiscovered for quite some time, mostly because he was not very good at publicizing his results, after all, Mendel was a monk; and it cannot have helped that his monastery was located in a provincial capital of the Austro-Hungarian Empire. The fact that he unwittingly used self-fertilizing plants prevented him from consistently reproducing his results and likely also made him delay publication. To avoid this would have required crossing of two species, or hybridization, instead of self-fertilization. In 1866, Mendel did eventually publish his results in *Verhandlungen des naturforschenden Vereines in Brünn* (*Transactions of the Brünn Natural History Society*) but this publication never reached the audience for whom it would have been a revelation. Most importantly, Darwin never knew about it and it would take until the early twentieth century for the true significance of Mendel's work to be rediscovered. The dark side of the rise of evolutionary concepts in biology at the end of the eighteenth and the beginning of the nineteenth century is their misinterpretation and misrepresentation to support or establish racial prejudices. We can be assured that racism is not an invention of the nineteenth century, it raised its ugly head certainly long before that. However, the modern concept of race does have its roots in the nineteenth century. And so does eugenics, as the quest for improving the genetic stock of human populations came to be known; its name deriving from the Ancient Greek for "well-born". Modern science is unequivocal in its assertion that there is no scientific basis for race, but its demons are still very much with us. Modern-day race debates entering scientific discussions precedes Darwin by about forty years and initially started with the English surgeon William Lawrence (1783 – 1867) and the publication of his *Lectures on Physiology, Zoology and the Natural History of Man* in 1819. What Lawrence put forward in this work was fairly modern as he insisted any concept of race would have to be studied through a zoological approach. This is exactly what modern science has done and how it found no grounds that would support the concept of race. However, in Lawrence's time others were quick to introduce race concepts to support and propagate their racially biased views not only in science but throughout all aspects of society.

It needs to be remembered that these were still times when it was not considered outrageous to suggest that women were inferior to men because they had smaller brains; which is the same as to say that men of shorter built are inferior because they also have on average smaller brains. We know today that this is all nonsense but at the time, this precept was used to deny women equal rights because they were seen as less intelligent. Eugenics is also linked to a man who may have had nothing but good intentions, the English statistician Francis Galton (1822 – 1911), Charles Darwin's half-cousin. However, he was greasing the wheels for others who would come later and would have no compunction to eliminate the "undesirable" properties in human populations, a language that Galton introduced. The plagues of eugenics and racial biases would become the scourge of twentieth-century societies. In their names millions of people would be killed just because they were seen as different, inferior, and pictured as vermin, not human. We still are visited by those plagues and have not been able to eliminate them for good. As it is, their coinage seems to be again becoming respectable currency in some societies, a development we all must watch closely and nip in the bud.

During the nineteenth century, like all the other natural sciences, biology evolved into a professional science leaving the era of the gentleman scientist behind. It was still mostly men, although women increasingly succeeded in securing admission into universities. However, it would take several generations before women would be able to join their male colleagues in the pursuit of science on a level playing field. There were of course some famous exceptions but even they had to contend with continued prejudices. Just as for physics and chemistry, the reductionist approach to biology began to require ever more expensive and complex instrumentation and resources that most individuals could not afford. More importantly, being part of the ongoing discussions within the scientific community almost made it a requirement to be a member of a university or a large research institution to participate fully. As the fields of molecular biology and genetics would come into their own in the twentieth century, this became ever more important. Although the reductionist approach in biology clearly gained prominence and eventually became dominating, the systems approach has remained a vital part of biology. With the systems approach in biology, it is not so much the resources one can command but the ideas one can come up with that help to make sense of the bigger picture in biology. This makes biology unique among the natural sciences as similar broad, encompassing, and fundamentally conceptual discussions about the foundations of a science do not exist for example for physics and chemistry, where they are perceived as beyond the scope of what these sciences can provide answers to. Not so for biology.

Along with the other natural sciences, twentieth-century biology profited enormously from the availability of ever better instrumentation. This has enabled biology over the last century to develop a detailed functional understanding of organisms from the organ level down to the cellular and molecular level. In the course of this pursuit, biology, just like the other natural sciences, increasingly differentiated into specialized fields. In the case of biology, there are now arguably more than thirty specialized fields or branches. Some of them, including such fields as cell biology, biochemistry, ecology, epidemiology, evolutionary biology, and microbiology, already had their start in the nineteenth century. Other prominent ones only came to fore over the past one hundred years including astrobiology, biochemistry, genetics, marine biology, and neuroscience among others. Medical research had already made giant steps in the nineteenth century and this progress would accelerate in the twentieth century. Advances in medicine and in biology had always been closely linked and for most of history, biologists had been trained physicians. This was the default situation and that would only start to change towards the end of the nineteenth century. However, even as biology became an independent study course at universities, all life sciences remained closely linked. The rise of research universities and with it of biological and medical research departments provided the basis for the development of modern medicine as it is practiced today. Many diseases or injuries that were practical death sentences in the past, maybe as recent as a couple of generations back, can be cured today. Would anyone really desire to live in earlier times?

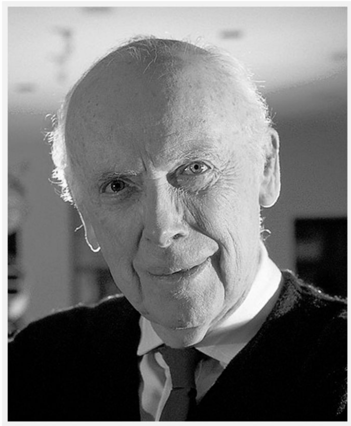
Thomas Hunt Morgan
1866 – 1945

Only a century ago, many medical treatments were often inefficient at best and not infrequently made things worse and not better because the underlying reasons for many ailments afflicting human health were just not known. Give it another century and maybe future generations will say something similar, at least to a degree, with respect to our time and age. Miescher first isolated the DNA molecule, the blue print for life, in the nineteenth century. The English biologist William Bateson (1861 – 1926) was the first to refer to the study of heredity as genetics and the Danish botanist Wilhelm Johannsen (1857 – 1927) was first to use the word gene to describe them as the building blocks that comprise the units of hereditary information.[30] Johannsen also introduced the words genotype and phenotype with the genotype referring to an organism's collection of genes and the phenotype being the observable expression of this genotype in the physical characteristics of the organism. During the first half of the last century, scientists were meticulously studying how genes and heredity related. Much of the accomplishments of genetics over the last century would not have been possible without the widespread use of genetic model organisms. This method was pioneered by the American biologist Thomas Hunt Morgan who used fruit flies to study how mutations were passed on between generations of this short-lived model organism. His results not only convinced him that Mendel's work on heredity was correct but more significantly, that it was the chromosomes which carried the hereditary information. In addition to earning him a Nobel Prize, Morgan's genetic work on fruit flies established these insects as the model organisms for geneticists. Like many of his colleagues at the time, Morgan was initially skeptical of the Darwinian theory of evolution. However, his very experimental work with fruit flies gave him first-hand insight into how important mutations were for the evolution of organisms. From there it was a small step to realize that mutations combined with Darwinian selection could drive the evolution of organisms. The summary of his findings and thoughts regarding the role of Darwinian evolution and of genes and their mutations published in *A Critique of the Theory of Evolution* in 1916, practically made Thomas Hunt Morgan the father of modern genetics.

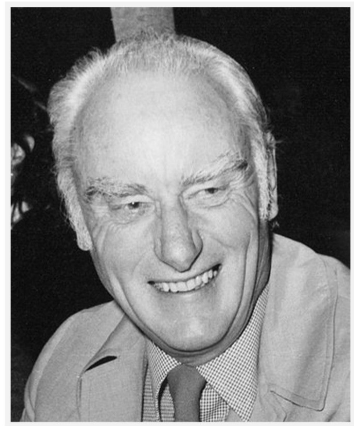
In 1895, the German physicist Wilhelm Conrad Röntgen (1845 – 1923) had discovered X-rays, electromagnetic radiation just like visible light-waves.[31] Because of their much shorter wavelength, X-rays can penetrate materials that are opaque to visible light and can image objects at much higher resolution. The development of X-ray imaging enabled biologists to analyze much smaller objects than what the resolution limits of optical microscopes allowed, including biomolecules such as DNA. In 1953, this finally led to the discovery of the double-helix structure of the DNA molecule by four scientists: the American biologist James Watson and his British colleague Francis Crick, the New Zealand born British physicist Maurice Wilkins, and the British chemist and X-ray crystallographer Rosalind Franklin.

This discovery marks the very beginning of modern molecular genetics. Eventually, it would lead to the complete decoding of the human genome during the first years of this century. However, this discovery of the DNA structure was also tainted. The three men - Watson, Crick, and Wilkins - who in 1962 were awarded the Nobel Prize for Medicine for this discovery had used the data of their colleague Franklin without acknowledging it. Franklin's X-ray imaging data provided clear proof of the DNA's double helix structure. Without this experimental proof, the model the three men came up with, mostly conceived by Watson and Crick, would have been just another theoretical conjecture. It would have been a hypothesis, a model such as the one proposed a little earlier by the famous American chemist Linus Pauling (1901 – 1994) who had suggested a triple helix structure for the DNA molecule. Watson, Crick, and Wilkins, who between themselves were incapable of coming up with experimental proof of the double helix structure had no ethical reservations using their colleagues data without properly recognizing her contribution. Recognizing Franklin's contribution, the experimental discovery that the DNA backbone made of nucleotides, sugars, and phosphates had a double helix structure, would have in no way diminished their own achievement in figuring out how the rungs of the ladder were made up of pairwise combinations of the bases adenine (A), cytosine (C), guanine (G), and thymine (T). If a scientist today would use a colleague's data in such a way without acknowledging his or her contribution it would without doubt be rightly considered unethical behavior, a severe violation of professional conduct, and the career of the scientist engaging in this behavior would be much damaged or even be over for good. Not so in 1950's Oxford where it was seemingly still okay to do just that. What made it likely even easier for the three men to exclude their colleague was that Franklin was a woman and women were just not part of the all-boys science network in places like Oxford and therefore, seemingly could be taken advantage of with impunity. The Nobel Committee awarding the Nobel Prize for Medicine could have rectified this injustice by giving the three men credit for working out the chemical structure of the DNA ladder rungs and recognizing Franklin for the experimental discovery of DNA's double helix structure. Tragically, Franklin had passed away by the time the prize was awarded for this discovery in 1962 and dead scientists are not honored with Nobel Prizes. We do not know if the Nobel Committee would have done the right thing and included Franklin in the prize if she had still been alive in 1962, we can only hope so. More often than not it is the victors who write our history and for a long time Franklin's contribution has been belittled. However, more recently, Franklin is given the credit she deserves as one of the co-discoverers of the DNA structure. It is up to us to remember that Rosalind Franklin was a distinguished experimental scientist who discovered the double helix structure that the three men were incapable of discovering by themselves experimentally. Even more so, Rosalind Franklin was on her way to understand the chemistry of the DNA ladder structure herself.

Rosalind Franklin
1920 – 1958



James Watson
1928 -



Francis Crick
1916 – 2014

Maurice Wilkins
1916 – 2004

She was no less a scientist than her three male colleagues and certainly not the clueless and incompetent lab assistant some sought to portray her later. The discovery of the four molecular letters, A-C-G-T gave scientists the alphabet and some grammatical rules for how the code of life is written into DNA. It may not sound too difficult a challenge to decode a text written in only four letters. However, deciphering the code of life requires the translation of a rather lengthy text, some three thousand million base pairs, without the benefit of a Rosetta stone at hand to make sense of it in the first place.[32] It is hard to overstate the magnitude of what modern genetics has achieved in accomplishing this feat. While we are still at the very beginning, the little we can read in the book of life for now has already a dramatic impact on our understanding of the evolution of life and our abilities to edit parts of the code of life itself. According to ancient Greek mythology, Prometheus created humans out of clay, stole fire from the Olympian Gods to give it to humans, and thereby started civilization. As humanity acquires the capability to engineer life, we may be at the threshold of another Promethean era. The quest to understand the origins of life and more than that, what life actually is, goes far beyond biology. Not surprisingly, specifically the latter has drawn in scientists from physics and chemistry, grappling to define life in their own language. Erwin Schrödinger is one of the earliest examples of that with a series of lectures that were published in 1946 under the title *What is Life?*. The thoughts he outlined in this work explore what defines life from a physicist's perspective and the questions he asked in it back then are still very much with us, refined and expanded by what scientists have learned since then.

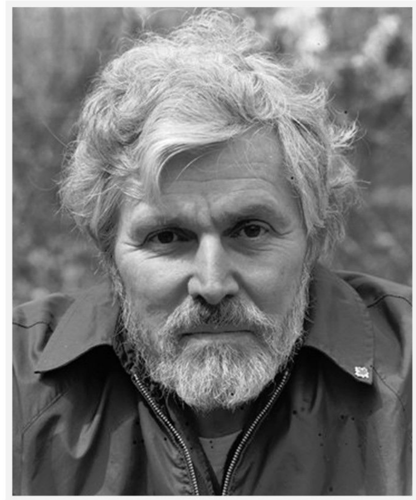
The progress neuroscience has seen over the past few decades has provided us with a profoundly changed understanding of the human brain. At the same time, we are getting ever closer to understanding the fundamentals of life itself and we have started to develop tools that will eventually allow humans to engineer their genetic evolution. Added to that, we are also increasingly discovering the biological origin of our perception of self, of who we are as humans and as individuals. The ancient mind-body problem resided until modern times very much within the domain of philosophy. Until quite recently, we knew too little about the human brain and how it works to consider the biological foundations of how we behave, or how we think, or how the individual concept of self may be rooted in biology. Even though there is still so much we do not know about the brain, neuroscience has now put us into a position where we can start asking quite sensible questions regarding the mind-body problem from a biological perspective. As it turns out, we may be much less in control of our own self than we believe. Not that this would abrogate our cherished notion of free will completely, but it does certainly dent it. Our understanding of consciousness is still almost non-existent. We only are able to describe it as to its impact on what we may or may not be doing or thinking but in no way can we say anything meaningful as to how it comes about. Maybe we never will be able to do so because consciousness trying to analyze itself is a circuitous problem, we may not be able to get around. However, there are certainly things that we now know about the brain and its development which should and at some point will have wider ranging consequences.

As we increasingly learn how the wiring of our brains is conditioned throughout our childhood and into early adulthood, there are instances where it really becomes questionable who is at fault for something that happens as a consequence of a brain not developing in the way it should or being faulty so-to-say from the start. The combined efforts of neuroscience and genetics are helping us to decipher how seemingly our genetic make-up determines to a large extent how our brains function. Like for example, how some of us become addicted easily while others do not, some are much more aggressive while other are not, or some gain weight much more easily than others while not consuming that much more. Society will eventually have to revisit what seemed to be a settled discussion on the topic of nature vs. nurture. We can expect that this will have wide-ranging implications for how we organize our societies from healthcare to education to criminal justice. That is, unless we decide to continue to treat the symptoms instead of addressing the causes for what afflicts human societies.

The system level aspect of biological understanding is just as important as the reductionist research approach that gave us microbiology, genetics, and much of modern medicine, just to pick a few examples. How exactly does system level biology benefit us? That is, beyond the sheer joy of understanding and obtaining knowledge that for some, including yours truly, would already be motivation enough. Well, as we increasingly discover, life does not exist in isolation. Humans would not be here today if they had not been part of the deep and entangled web of life that is our living Earth. For a long time, it really looked like as if humans had it all for themselves and we could extract resources and destroy life around us with impunity. We now know we cannot. To understand what we can and cannot do or what we must and must not do to ensure that our living Earth continues to support the broad diversity of lifeforms we can still find today is critically important. It is with respect to these questions where system level biology can help us find answers in the sense that it considers the totality of life and its environment. It is only in very recent times that words such as ecology, environmental impact, or sustainability have become common currency in policy discussions and decisions. Not everybody agrees with this development and there are still too many who do not understand yet that we cannot continue to exploit living and nonliving resources on Earth with impunity without endangering our very own survival as a species.

A simple consequence of the mutual interdependence of life on Earth is that one cannot fully understand life by studying it in isolation. We all interact with each other and with our environments in all aspects of our lives; and so do all other living creatures on this planet, animals and plants alike. Behavioral biology is an interdisciplinary field that explores how behavior and biology influence each other. Even though we do not understand yet deeply or detailed enough how our brains really function, we already know that our brains biology may have something to do with how we are likely to behave and react with respect to various conditions or stimulants. The latter is something that more narrowly is referred to as sociobiology. More broadly, how we perceive events and respond to them in turn influences how we react physiologically, for example by increasing hormone or enzyme secretions resulting in muscle contraction in preparation for flight or in sexual arousal in expectation of something more pleasant. Without the insights of behavioral biology much of the animal worlds more complex behaviors would be quite difficult or even impossible to understand. Today, behavioral biology encompasses the study of human and animal behavior in their natural and social contexts; plant life is not a part of this effort yet. However, that may simply be a limitation in our current perception of plants. For much of plant life the internal clock ticks just slower than for animal life and everything seems to happen on a much more leisurely time scale than in the animal world; and then most of the plant world does not move much but mostly stays in one place for a whole lifetime. But does that mean plants have no behavioral reactions to their environments or their behavior does not influence their environments? Quite certainly not. If biology has taught us anything, we can expect that plants will keep surprising us. Plants, at least in our current perception, may not have a flourishing social life comparable to humans and many animals. However, taking it for granted that they have no social life at all is likely a bad bet.

Evolutionary biology has benefited enormously from the advances in genetics over the past couple of decades. Evolutionary developmental biology, or Evo-Devo as it is often called today, has provided us with a significantly more accurate and in-depth understanding of the evolution of life on Earth than anyone in Darwin's time could have imagined. The ability to trace the genealogy of individual species back to their respective last common ancestors has revealed a more more detailed picture of life on Earth than anything that was available to Darwin. As genetics developed the tool sets for Evo-Devo to accomplish this, paleontologists continued their meticulous study of past life on Earth. The evidence from paleontology and genetics about species living on Earth, past and present, often corroborating, sometimes complimentary, paints a rich tapestry of life, and of death. As we know today, the vast majority of all species that ever lived on Earth are extinct today. While there have been some marathon champions in evolution who managed to survive for hundreds of millions of years, most species do not last that long. We humans are a very recent species addition to life on Earth. How much do we really know about life on Earth? That is a difficult question to answer. The brief walk we just took through our understanding of life up to the modern age should instill some humility. On many occasions, it looked like a new explanation, insight, or discovery finally seemed to pull back the curtain to reveal the complete canvas that life on our planet has painted since it first originated. However, every time it became clear all too quickly that we were still only scratching the surface. Does the combined scientific knowledge that is part of our modern human culture give us a much greater and more detailed understanding of life on Earth than humans had at any time before? Yes, most certainly.



Carl R. Woese
1928 – 2012

However, to what extent can we consider our knowledge complete in the sense that we know almost everything that there is to know about life on Earth? The honest answer is we just do not know. But if we let history and human nature be our guides, we can only conclude that there is so much more to learn that we do not yet know about life on Earth. It is meaningless to try quantifying what we know as a percentage of what complete knowledge about life on Earth would truly mean. What should make us very cautious is for example the simple fact that until about forty years ago we had not even discovered one of the three major groups of life on Earth. On trees of life such as the one shown in fig. 3.2, the lineage of animals was always much better understood than the part of the tree that dealt with single-celled organisms such as bacteria; restricted of course at any given time by what was known about animals in the first place. The simple reason for this is that plants and animals are much easier to characterize as species by their visible characteristics as discerned from living and dissected specimen. Categorizing bacteria into species is a much more difficult challenge. Bacteria do come in a few different forms such as round, elongated, or spiral but how does that enable one to group two different bacteria into one species? Here is where the American microbiologist Carl R. Woese came in. Using protein comparison to establish genetic relationships between organisms had been suggested previously but no one had ever applied this method to determine species relationships among bacteria. Carl Woese's insight was that he could use the protein analysis methods available in the 1970's, which compared to modern technologies was very cumbersome and time consuming, to do just that. Comparing proteins with high accuracy and in a sufficient number to get reliable data for a clearer picture of what constitutes a bacterial species took years. His great discovery came with the inclusion of some bacteria that were a bit unusual as they derive their energy from different chemical reactions than photosynthesizing bacteria. In the late 1970's, Carl Woese knew that these special bacteria that were of the heat-loving or methane-producing kind actually were nothing like bacteria at all but a new lifeform that we know today under the name archaea. Archaea together with bacteria and eukarya became known as the so-called three domains of life. How those domains of life resemble a tree of life like the one drawn up by Ernst Haeckel shown in fig. 3.2 has been a controversial discussion ever since. Specifically so as more and more evidence emerged that made it plain clear that the picture of evolution that had evolved early in the twentieth century from the merger of population genetics based on Mendelian heredity with Darwin's concept of evolution was incomplete.

This early twentieth-century modern synthesis of these two perspectives on evolution is often called Neo-Darwinism. There is still much discussion as to how this modern synthesis in the form of Neo-Darwinism and the Evo-Devo view of evolution centered on developmental biology can be integrated into a new synthesis but that has not happened yet. Linear heredity in the Neo-Darwinian view where genetic material with the occasional mutation is passed on from one generation to the next to be subjected to the forces of natural selection is only part of the story. Evo-Devo approaches evolution from a developmental perspective, focusing on the understanding of the developmental processes of an organism and how they impact the phenotype. With knowledge that was not available to Darwin, today's view of how evolution worked and continues to work has become more differentiated since Darwin penned his *Origin of Species* a little more than 150 years ago; and it is also quite different from what evolutionary orthodoxy still held true for much of the second half of the last century. However, a new synthesis cannot just be about merging Neo-Darwinism and Evo-Devo. In the second half of the last century things were changing and today our picture of evolution, certainly still far from complete, is a more differentiated one and includes a number of new concepts such as co-evolution, endosymbiosis, convergent evolution, and epigenetics; and most importantly, the discovery of widespread horizontal gene transfer within and across species.

It was already clear to Darwin, specifically from his studies of the close relationships between flowering plants and insects, that species evolution was in some way interdependent. However, such topics did not become part of evolutionary biology research until the middle of the last century at which time the term co-evolution was introduced. The concept of co-evolution describes evolutionary pathways where two or more species reciprocally affect each other's evolution. In its narrow sense, co-evolution is strictly reciprocal with two species exerting selective pressure on one another under which certain traits nature selects against then evolve. In a broader sense, co-evolution is not confined to species pairs but also includes multiple species having significant effects on the evolution of any particular traits within a small ecological community. In both cases, co-evolution is most likely to take place where species interact with each other in environments that are differentiated by for example geographically well-defined habitats with assured long-term continuity thereby enabling close and stable ecological relationships. These can be of a varying nature including predator-prey or host-parasite relationships as well as relationships that are either competitive between species or to the mutual benefit of species. Classic examples of co-evolution can be found where some plants and insects have formed almost exclusive relationships that assure the pollination of the plant and provide food or protection benefits or even both to the insect. Similar exclusive arrangements have also evolved among animals as for example between some ants and aphids. The latter are small sap-sucking insects, which the ants herd and protect in exchange for the aphids providing sweet nutrition to the ants. In prey-predator relationships, predator and prey are tied in an arms race where the prey evolves better protection or protection strategies while the predator grows larger teeth or claws and comes up with new ways to outsmart the prey. For the co-evolution of parasite-host relationships, we only have to look at ourselves where for example both head- and pubic lice have co-evolved with humans.

Endosymbiosis, sometime also called symbiogenesis, is an evolutionary concept that has provided an answer to what for a long time remained a riddle to biologists. How did eukaryotic cells evolve to have a nucleus and organelles within their cell membranes? One of the most colorful and intellectually independent biologists of the last century, the American evolutionary biologist Lynn Margulis, discovered the solution to this riddle. Before Margulis, most evolutionary biologists had only ever considered and advocated vertical gene transfer as genes could only be inherited and not acquired otherwise.

Lynn Margulis
1938 – 2011

As Margulis acknowledged herself, the creation of new species through symbiosis, was not really an original idea but something that she herself and the broader biology community had remained unaware of. The German botanist Andreas Schimper (1856 – 1901) observed in 1883 that in plant cells the organelles called chloroplasts, tiny factories converting sunlight into usable energy for a plant, reproduce through cell division; just as cells do themselves. Schimper's published discovery, *Über die Entwicklung der Chlorophyllkörner und Farbkörper* (*On the development of chlorophyll granules and colored bodies*), came to the attention of the Russian biologist Konstantin S. Mereschkowsky (1855 – 1921). Mereschkowsky built on Schimper's work and published his theory of symbiogenesis, *Teoriya dvukh plazm kak osnova simbiogenezisa, novogo ucheniya o proiskhozhdenii organizmov* (*The Theory of Two Plasms as the Basis of Symbiogenesis, a New Study of the Origins of Organisms*) in 1909. Mereschkowsky contended that cells that are more complex evolve from the symbiotic relationship between less complex ones. Because organelles such as the chloroplasts do not behave like organs or organelles but like independent organisms, they must either be that or symbionts, organisms living in symbiosis with their host. The American biologist Ivan E. Wallin (1883 – 1969) developed similar ideas in his 1927 publication of *Symbioticism and Origins of Species*, suggesting that eukaryotic cells were composed of microorganisms. Margulis put forward her findings and ideas on endosymbiosis in her 1970 publication of *Origin of Eukaryotic Cells*. At the time, they were rejected by most biologists, just as the works of Schimper, Mereschkowsky, and Wallin had been put aside before. However, this time would be different. Endosymbiosis as proposed by Margulis became the accepted explanation of how eukaryotes acquired their nucleus and organelles. Up to Margulis work, the community of evolutionary biologists was only prepared to accept vertical gene transfer as the mechanism driving evolution. The horizontal gene transfer suggested by endosymbiosis represents a paradigm shift in how we have to come to think about evolution. As the German born American Ernst Mayr (1904 – 2005), one of the leading biologists of the twentieth century, put it, evolution through horizontal gene transfer may well be the most important event in the history of life on Earth. Endosymbiosis changed how we view evolution and consequently how we symbolically represent it in schematics such as trees of life. The linear progression of evolution represented in trees of life of the past is gone and while we do not know what a good replacement for it would look like, a bush with a healthy number of twigs crossing over horizontally in the lower parts of it may be a good way to start thinking about it. There is a lesson here: picturing our ideas is very helpful as long as that does not start to constrain our thinking, as the linear progression of the old trees of life had been doing for quite some time.

As it turned out, the specific version of horizontal gene transfer that takes place in endosymbiosis is just one of several mechanisms through which organisms can take an evolutionary shortcut. In the late 1920's the British physician Frederick Griffith (1879 – 1941), while researching several strains of bacteria causing pneumonia, had observed that these bacteria were seemingly capable of transferring cellular material between themselves [33]. This somehow resulted in non-virulent living bacteria transforming themselves into virulent bacterial strains through absorption of cellular material from dead virulent bacteria. Nothing like that was supposed to happen but it did. It took time and the effort of many brilliant scientists to understand that the transfer of genetic information between bacteria is not the exception but practically the rule.[33] What Griffith had discovered was one of three mechanism we know of today by which bacteria can exchange genetic information. The material that Griffith's living bacteria had acquired from the dead bacteria of the virulent strain was DNA material. It would take another twenty-five years after Griffith's unbeknownst observation of horizontal gene transfer between bacteria through transformation, before the DNA molecule would eventually be discovered. The transformation process Griffith observed is now known as a form of genetic recombination in which a DNA fragment liberated from the busted cell of a dead bacterium enters into the cell of another bacterium where it is swapped with a piece of DNA from the recipients cell. Critically, it was found that this transfer is not limited to within a bacterial species but can jump between bacterial species. When this was first discovered, it must have felt like an enormous surprise. Back then, the species concept was still perceived as something that could not be bridged by anything. However, since then our understanding of the species concept has very much changed and it now seems to be widely recognized that it has only very limited applicability for bacteria. In addition to the mechanism of transformation, bacteria can also accomplish horizontal gene transfer through the mechanisms called conjugation and transduction.

Bacterial conjugation is the process by which one bacterium transfers genetic material to another through direct contact. This is why it is sometimes regarded as the bacterial equivalent to sex, which however is incorrect. In the process of conjugation, the genetic material is introduced into the recipient bacterium as a so-called plasmid, a short circular strand of DNA. The plasmids in a bacterial cell are physically separated from the bacterial chromosomal DNA and can replicate independently. This is critical as it allows the exchange of DNA material between different bacterial species and is therefore key to the very fast adaptation of bacterial strains to environmental changes. Plasmid DNA that benefits the recipient bacterium may eventually become included into the chromosomal DNA of that bacterium. If bacteria had to wait for mutations in their chromosomal DNA to eventually provide them with protection against the antibiotics we develop, humans would have won this arms race long ago as likely the respective bacteria would be wiped out before they had time to adapt. However, by exchanging genetic information through horizontal gene transfer such as bacterial conjugation, the bacteria harmful to humans that we are fighting with antibiotics can adapt very quickly. The results are the so-called superbugs that have acquired resistance to the most powerful antibiotics we have in our arsenals. Bacteria likely developed this ingenious method for quick adaptation long ago to fight off naturally occurring antibiotics. It seems unlikely that we humans will be able to win this war that bacteria have been successfully fighting for more than three billion years unless we come up with some new game-changing ideas.

The third method of horizontal gene transfer after transformation and conjugation is the transfer of a DNA fragment from one bacterium to another by a bacteriophage; it is called transduction. A bacteriophage is a virus that parasitizes a bacterium by infecting it to reproduce itself. The bacteriophage genome seems to be able to achieve this in two different ways: the viral genome directs a bacterium to replicate the virus in the process of which the bacterial chromosome will be broken up; or by the bacteriophage splicing its genome into the chromosomal DNA of the bacterium.[34] In the first case, newly created viruses inside the bacterium can accidentally enclose bacterial DNA fragments, either from the chromosomal DNA or from plasmids. As it dies, the bacterium releases the newly created viruses, some of which can carry bacterial DNA and transfer it into a new host bacterium. In the second case, the viral genome spliced into the chromosomal DNA of the bacterium stays dormant, a state in which it is referred to as a prophage. As the bacterial DNA replicates during cell reproduction so does the virus genome. As long as the prophage stays dormant, the virus genome reproduction will continue in this way, albeit being subject to any other genetic modification of the bacterial DNA that may also affect its viral genome. This cycle stops once the prophage genome is expressed again, or more directly, until the prophage genome information is used to produce new viruses. This happens by cutting out the respective prophage portion that the bacteriophage originally spliced into the bacterial chromosomal DNA to reproduce new viruses. However, this process is not perfect and the portion cut out from the bacterial chromosomal DNA may also include some bacterial DNA, or the portion that contained the prophage genome itself may have been altered over many bacterial reproduction cycles. This second version of the bacteriophage life cycle is more specific in what kind of DNA fragments it can pick up. Just as in the first version of the bacteriophage reproduction cycle described above, once the viruses newly produced in the host cell use the prophage genome to execute their release from the host, the bacterial cell dies.

Given the large amount of genetic information acquired through horizontal gene transfer, the species concept seems to be a difficult one to apply to single-celled organisms. More importantly, we now know that horizontal gene transfer also affects multicellular lifeforms including animals. This certainly is something that would not have been imaginable when Darwin formulated his concept of evolution. However, biologists have for example found that some insects acquired bacterial gene sequences through horizontal gene transfer. While initially much of the transferred genes were thought to be non-functional, this has turned out not to be correct either. We know that the bacteria that evolved into the organelles, once they were absorbed through endosymbiosis by the ancestors of plant and animal cells, transferred large pieces of their genomes to be integrated into the DNA of the host cells. This process left the organelles with a much smaller genome than their bacterial ancestors had. Bacteria are seemingly able to generously swap genes and our human bodies are hosts to a large number of them. The still frequently quoted ratio of ten to one for bacteria vs. human cells in our bodies is wrong. The typical human body contains about the same number of human and bacteria cells but bacteria cells are quite a bit smaller than human cells. A so-called standard person weighing seventy kilograms has about thirty-eight million million bacterial cells in her or his body and a roughly equal number of human cells. But because bacteria are so much smaller than human cells, their total mass is quite a bit less. Recent studies tell us for our seventy-kilogram reference person that he or she carries a total bacteria weight of around two hundred grams [34]. This large presence of gene swapping bacteria in our bodies of course begs the question: do bacteria in humans on occasion also swap genes with human cells? The answer is a resounding “Yes”. Humans have evolved in a world of bacteria and we should therefore not be surprised to find a quite large amount of bacterial DNA in the human genome. Bacteria and the distant ancestors of our human lineage likely have been swapping genes since animals first evolved [35,36].

When human genome analysis first became feasible, bacterial sequences were routinely removed from analysis results because they were assumed to be due to contamination of the human genome. However, as it turned out, not all of the bacterial DNA detected alongside human DNA was contamination, some of it was actually part of the human genome itself. Bacterial DNA does not account for all of the foreign DNA introduced via horizontal gene transfer into our human genome. The human genome also contains foreign DNA introduced into it by viruses and so does the DNA of other animals. And there is more, as some DNA fragments in the human genome are for example being attributed to protists, the name still in use to refer to large single-celled organisms, or fungi. Our human genome is much more like a mosaic with quite a large amount of genetic material which we acquired from other species and did not inherit from our direct line of animal ancestors. That seems to be the case for most other lifeforms on Earth at varying levels. This does not sound like an evolution as Darwin conceived it or as the broad consensus of scientists saw it until only a few decades ago. But that does not mean that Darwinian evolution is wrong, as some would like to have it who still believe that Earth was created only a few thousand years ago. In no way does it subtract from the seminal insights and contributions Darwin made to our understanding of evolution.

However, it tells us that our understanding was not complete. We humans tend to quickly turn some of our great insights and discoveries into orthodoxy and tend to forget that learning never ends. Newton was not wrong because Einstein gave us a new way to look at space-time and gravity, helping us to understand the true nature of our universe as we discovered it since Newton had developed his view of gravitation some two hundred and thirty years earlier. In the same vein, Darwin was not wrong to see evolution constrained by species boundaries because in his time and long after that, there just was no evidence to suggest different. We are all beholden to our times and the knowledge and facts that are available to us to build on; and so were Newton and Darwin. Some day, we will likely say the same about Einstein. Evolution is more complex and intricate than we may have believed. We only just have started to understand parts of the intricate web of life that has evolved on Earth through history, presenting us with the world we live in today. This includes our recent understanding of what we now refer to as convergent evolution.

Convergent evolution is a concept that emerged in the late twentieth century. Simply put, convergent evolution acknowledges the fact that nature can drive separate species to come up with similar evolutionary adaptations to similar evolutionary constraints, independently from each other. Or, stated more formally, convergent evolution reflects the observation that organisms evolve essentially similar structures that serve analogous functions without their evolutionary ancestors being closely related. Of course, as we know, all life on Earth is related, how close is just a matter of degree. However, once two species have split as for example reptiles and mammals did at one point, both of them evolved animals with wings as we can still see today with birds and bats; the former, whose distant ancestors are the dinosaurs, clearly of reptilian progeny, and the latter of course members of the mammalian class of animals. Convergent evolution need not be restricted to specific functional parts of organisms but can also extend to whole organisms. The most instructive case is the evolution of marsupials in Australia. Until a couple of hundred years ago, life in Australia had been isolated from the rest of the world for many millions of years. After the demise of the dinosaurs, as placental mammals evolved in the rest of the world to fill available ecological niches so did marsupial mammals in Australia. Marsupials thrived in other places as well, notably South America but there they faced competition from placental mammals much earlier than in Australia. One can virtually put together a list of evolutionary solutions that produced similar organisms among placental and marsupial mammals to occupy similar ecological niches. Now, all of these examples show convergent evolution working over very long time spans. However, more recently, it has also become clear that convergent evolution works on much shorter time scales, actually so short that it can be studied in the field.

One of the great debates in evolutionary biology came to fore only in the last several decades of the twentieth century. The fundamental question that this debate centered on was: if we could roll back everything to 4,000 million years ago, so-to-say reset the clock and have evolution start again with a clean slate, would we see a similar outcome? Now this is a very big question to ask. Some evolutionary biologists answered it by insisting on a strict causality. Everything needs to happen the same way in such an evolutionary replay to produce a similar outcome. If in our evolutionary history some event $\boldsymbol{B}$ depended on another event $\boldsymbol{A}$ preceding it, then event $\boldsymbol{A}$ must also happen in an evolutionary replay exactly the same way to cause event $\boldsymbol{B}$. This is a very narrow view and for some time it had the advantage that there was no real way to test the repeatability of evolution.

However, that has changed as we now have evidence for convergent evolution taking place on a humanly observable time scale. Today we have many examples of replaying the evolutionary tape where convergent evolutionary responses have produced similar outcomes when driven by similar evolutionary constraints. New species generation can be studied in nature and has been beautifully demonstrated with lizard populations on small islands, to give just one example of several that are known today. Clearly, we will never be able to do a complete replay of evolution. However, from the evidence we already have it is indisputable now, that the strict causality view requiring everything to happen the very same exact way to provide the same overall outcome is not correct. Given the complexity of life's evolution over the past $\sim$4,000 million years with unknowable millions of species rising and falling, we should not be surprised that strict causality approaches to understand how all of this unfolded, have only a very limited applicability.

There are many interesting aspects of convergent evolution to explore, but here we will only pick one. From the earliest life on Earth, we know of today, it took evolution almost 4,000 million years until the arrival of our species; the only species on this planet, as far as we know, capable of pursuits such as trying to understand the universe, its own history, and its own self. It is a natural question to ask: does it always take so long for such a species to evolve on planets that can sustain life similar to Earth? More poignantly, is the solution always a bipedal organism with two free hands to make tools, sensory organs facing to the front in the direction of walking, and a brain comparable to ours? Convergent evolution would suggest a cautious and conditional "Yes" to these questions.

Clearly, other animals than mammals could have evolved a convergent evolutionary solution along the line that humanoids represent today, the dinosauroid that never came to be is frequently mentioned in this context. The second part to the question is difficult to answer as we only have one example of evolution that we can study. Could humans have evolved much earlier? Maybe, maybe not. We really do not have any good answer to that question yet and we may never have. As we have seen earlier, as far as the evolution of stars and planets is concerned, there is a good level of predictability as to how things will play out over time. We know for example what is required before a star can fire up its nuclear furnace, and how long that furnace will be burning hydrogen before it will switch to helium; and we know how long the star will last. For life and the evolution of life, we have no such predictability. Our universe is home to what feels like an infinity of stars and most likely also to countless habitable planets. However, we know nothing about life and the evolution of life on any such planet and most likely we never will or at least not for hundreds of millions of years; provided humans should last that long. Therefore, we should look at life on Earth as a one-of-a-kind experiment, something that we must work hard to preserve and protect, whatever it takes.

A much more recent twist in our understanding of the evolution of life on Earth focuses on a very particular problem that already Charles Darwin was struggling with. A key aspect of Darwinian evolution is that it plays out over very long time spans as random mutations may or may not lead to advantages that result in permanent evolutionary changes. Darwinian evolution works on populations through natural selection. Most mutations do not result in a significant benefit. Worse than producing no biological change at all, many mutations are outright harmful. Darwinian evolution is therefore a process that reshapes life on Earth over very long time periods. Without an understanding of Earth's true geological time scales, which only began to emerge a few decades before Darwin, pioneered by men like James Hutton and then Darwin's mentor Charles Lyell, Darwin's concept of evolution would not have made sense.

Evolution as Darwin saw it required the enormous geological time scale to work its magic. Evidence for this slow and gradual evolution of life was supposed to include transitional lifeforms, so-called missing links, connecting later lifeforms to their evolutionary ancestors. Eventually, as the fossil record for life on Earth accumulated over the past 150 years, this started to become a problem. Scientists carefully scrutinize this accumulating evolutionary fossil record in search for those transitional fossils that according to Darwin's interpretation of evolution should provide the evidence for this gradual evolutionary change. However, most often, the fossil record does not provide it. Even if one makes room for the fact that the fossil record of evolution is fragmentary at best, there are just too many instances where new species seem to arise out of nothing as other species vanish forever without any transitional forms preserved. This is particularly so when paleontologists look at the geological formations that separate what was before and what came after one of the several mass extinctions that life on Earth has experienced.

This conundrum has been perplexing paleontologists for quite some time now. In a way, the mold of purely Darwinian evolution based on the concept of only vertical gene transfer had already been broken when it became clear that horizontal gene transfer was instrumental in the evolution of eukaryotic cells. Instead of random mutations driving eukaryotic cell evolution, the evolutionary mechanism at work was the acquisition of complete bacterial genomes and their subsequent transcription into the genome of the host cell. Instead of only searching for answers in the fossil record, as is the purview of paleontologists, their evolutionary biologist colleagues began to explore in their labs what could be behind the missing "missing links". As it looks like, the answers they are finding point in a quite peculiar direction, Lamarck may have been on to something after all. It is an understatement to say that Lamarck's work on evolution, predating Darwin by about fifty years, was not well received. Lamarck for once had the bad fortune of not being part of the scientific establishment of his native country, France. However, even worse, in Baron Cuvier, whom we also encountered earlier, he had a formidable opponent who seemingly did everything to ruin Lamarck's scientific reputation and in doing so also deprived him of the means to sustain himself. Lamarck died as a pauper and to this day, Lamarck's reputation suffers from an interpretation of his thoughts on evolution that practically make him look stupid and ignorant. However, we are now at a point where biologists tell us that acquired characteristics, or characteristics changed through environmental factors, may indeed result in those modifications being passed on to the next generation. Even though not quite as dramatic as Lamarck may have imagined it, regardless, here it is, inheritance of acquired characteristics.

The name for this process is epigenetics and more importantly hereditary epigenetics.[35] Epigenetics is the study of changes in organisms caused by gene expression rather than by an alteration of the DNA, the genetic code itself. Epigenetic changes in general can only affect the living organism without being passed on to the next generation but the more narrow definition of epigenetics requires these changes to be hereditary. Crucially, epigenetic changes can be driven by environmental factors stressing an organism that can alter gene activity during the organism's lifetime and can be passed on to successive generations, eventually resulting in a permanent change to the DNA, the organism's genome. Changes to how an organism's genes can be expressed without changing the DNA sequence can occur in a number of ways: through DNA methylation, through histone modification or through the methylation of specific RNA molecules.

DNA methylation is the process by which enzymes attach methyl groups, CH_3- in chemistry notation, to DNA sites, frequently near the beginning of a gene sequence. However, the beginning of a gene sequence is the same place where proteins attach to activate a gene. If a methyl group blocks this site, then proteins cannot attach, and this usually means that the gene remains turned off. Thereby methylation of these sites can change the activity of the respective DNA segment without changing the DNA sequence. Histones are the proteins that serve as the support structures of the DNA that wraps around them to fold unto itself for efficient packaging into a cell; together the histone proteins and the DNA as they are folded up in the cell nucleus constitute what is called the chromatin, which in turn forms the chromosomes. In a similar way to the methylation of DNA sites themselves, histone modification occurs through attachment of methyl molecules or other chemical groups to the histone proteins. The result is a change in how the DNA can fold unto itself and consequently how RNA molecules can or cannot attach themselves to specific DNA strands affected by this histone modification. The third pathway for epigenetic changes acts on specific RNA molecules that can modify the shape of the chromatin structure. This happens through the methylation of these small RNA molecules in a similar way to the DNA methylation described above. The record of these epigenitc changes brought about by DNA methylation, histone modification, or methylation of specific RNA molecules is referred to as the epigenome. If there are no such changes then the genome as recorded in the DNA sequence is identical to the epigenome. As epigenetic changes accumulate over the lifetime of an organism, the epigenome starts to differ from the genome. How genes are expressed can therefore differ between two organisms that start out with the very same genome, just as identical twins do. It was indeed the quest to understand how identical twins with the same genome could differ in gene expression, such as one twin developing cancer while the other does not, that led scientists towards the discovery of epigenetics.

The impact of epigenetic changes can be very significant, specifically if they act on what are called regulator genes, which control the blueprints of how cells are being constructed and differentiated. For a long time it had been assumed that for sexually reproducing lifeforms all of the epigenetic changes accumulated over the lifetimes of individuals would be erased, actually erased two times – once during the formation of the unfertilized egg and the sperm and a second time upon conception. However, it now seems to be clear that some epigenetic markers can be passed on in what is referred to as transgenerational epigenetic inheritance. This is where we come back to Lamarck as clearly the inheritance of acquired characteristics is what he had suggested some two hundred years ago. In all fairness, it needs to be said that Lamarck back then knew of course nothing about epigenetic changes, gene expression of the epigenome versus the genome, or transgenerational epigenetic inheritance. However, neither did Darwin know anything about genetic inheritance or the genome. While still difficult to swallow for many at the time and for some still today, from a plausibility standpoint Darwin's concept was not as much a stretch as Lamarck's ideas were then.

We are still very much at the beginning of understanding the role of epigenetics in the evolution of life. In all likelihood, even with the integration of epigenetics into the modern synthesis or a new synthesis of Neo-Darwinism and Evo-Devo, many open questions remain and much work is still to be done towards a more complete understanding of the evolution of life. However, epigenetics may have far more important near-term consequences for humanity. The extent to which environments can trigger epigenetic changes may have great implications for human evolution and for life on Earth in the twenty-first century industrialized world that we have built. We already may have changed our evolutionary path through changing our environments over millennia. Given the increased pollution we are exposed to, we may not just be poisoning ourselves, but we may also alter the trajectory of human evolution in real time. With all the different kinds of stress that we are subject to in our modern world and the changes in behavior they may trigger, we have to ask ourselves seriously what this will do to human evolution. There is already an ongoing discourse about what we should and should not do with our evolving genetic engineering capabilities. As genetic engineering and technology mature, this debate will concern many more of us. What kind of human genome modifications our societies should allow and which ones should be banned will shape the future not just of one society but the future of our species. However, as important as it is to consider the merits and dangers of purposefully tampering with our human genome, we need to pay significantly more attention to the silent and involuntary epigenetic changes through which our industrial societies clandestinely alter the human genome.

There are a number of other concepts scientists have discovered that help explain evolutionary processes and results. Frequently and not surprisingly, these concepts emerged as we have tried over much of the past hundred years or so to better understand our own evolutionary history in the broader context of evolutionary population dynamics. We cannot cover everything here and certainly not at the depth that a more profound understanding would require but we will introduce two of them here briefly, as we will refer to them later: genetic drift and gene surfing. In small populations, a genetic mutation that may be quickly lost in a much larger one, may actually become widespread to the extent that one hundred percent of the population carries the mutation within only a few generations. This effect, called genetic drift, is believed to have been in play when small groups of early humans migrated out of Africa. Every time small populations, human or otherwise, set out to migrate towards a new territory, genetic drift gets a chance to work. This must have been the case for early humans in Africa as they migrated and spread all over Africa long before any modern human left the continent; partially, the higher genetic diversity of modern African populations may still reflect this.

Where population densities are low and genetic drift is high, another effect can come into play, which is called gene surfing [37]. In the broadest sense, gene surfing describes an amplification of gene variation along the frontier of a population expansion. As new mutations appear on the wave front of an expanding population migration, they can increase in frequency as they continue to ride the migration wave expansion in space and time. Where they do this successfully, the tools of genetics allow us to trace back the trajectories of surfing genes and thereby trace the migratory patterns of population movements, including our human ancestors. The effects of both, genetic drift and gene surfing are decreases in local genetic variability. Short-term, these effects allow population to adapt quickly but long-term, they can deplete the gene pool if small populations stay isolated for too long.

In this section, we recaptured roughly the outline of the last two hundred years of biological thought, the time span over which our modern understanding of biology evolved. Essentially parallel to discovering the reality of the wider cosmos around us and acquiring a much more intimate knowledge of Earth, our species also began to understand the evolution of life on our planet and life itself. From the time humans first entertained biological thoughts until about the late eighteenth century, the veils of ignorance clouding our understanding of life were only slowly lifted. Our pace of learning has very much accelerated over the past two centuries and much of what was once a mystery to us is so no more. However, the mystery of life itself and our fascination with the story of life remain, even as we continue to understand it ever more clearly. The evolutionary concepts developed over the last two centuries, most of them quite recently, when joined together make for a much richer tapestry of the evolution of life on Earth than Darwin may ever have imagined. We should not be surprised that in a few decades the story of the evolution of life in all likelihood will be even more colorful than it already is today. Any brief walk through the concepts of evolution such as the one we just took is naturally constrained and it needs to be made clear that a deeper understanding of what underlies the evolutionary concepts briefly reviewed here will require a much more thorough study. Maybe this brief introduction will stimulate some to do just that. At this point however, we are ready to start our journey through the history of life on Earth.

The Evolution of Life

To tell the story of the evolution of life on Earth at the level of detail this most wonderful of stories really demands is nigh impossible in only a few pages. So, our journey will unfortunately be more like one of those guided tours that pretend to show tourists the richness of a countries culture and heritage in a couple of days. My sincere hope is that the reader will see enough to explore this story on her or his own as there are a number of great books that do tell different parts of that story in a much more detailed and marvelous way; some of them are listed in the bibliography section.

Before we start our journey, we need to increase our familiarity with the language of biology and add a bit more to our biology vocabulary. As discussed in the previous section, depictions of trees of life do not really convey the true relationships between lifeforms on Earth as they have evolved because they do not account for horizontal gene transfer. Even though, they still provide a good view of the ancestral lineages of species as long as one keeps in mind the likelihood of substantial lateral gene exchange, specifically in the distant past, closer to the symbolic root of the tree. Fig. 3.3 shows a schematic example of a frequently referred to tree of life, classifying life on Earth into three domains of life. This specific kind of schematic goes back to Carl Woese and his classification of life into the three domains of archaea, bacteria, and eukarya. With the caveats mentioned above, this concept is still quite useful. Many much more complicated pictures of webs of life have been drawn up that try to account for horizontal gene transfer but the simple tree shown in fig. 3.3 is easier to understand and just as useful as long as one is aware of its limitations. The tree of life shown here is the so-called universal phylogenetic tree of life, derived by analyzing similarities and differences in the genetic and physical makeup of organisms. For making the genetic comparisons in this tree of life, ribosomal RNA is used. The ribosome, one of the cell organelles we will discuss in a few paragraphs, is an ancient part of all cellular organisms. Throughout evolution, the function of ribosomes and ribosomal RNA has remained highly preserved and the variation of ribosomal RNA sequences allows us to trace evolutionary relationships on the species level. This method was developed first in Carl Woese's laboratories in the late 1970's and since then has been much improved by today's modern techniques that produce significantly better results in much less time. The phylogenetic tree in fig. 3.3 reveals the genetic relationship between species and between the three domains of life. Seemingly, eukarya, the domain humans belong to, are more closely related to archaea than to bacteria. All three branches are rooted in what is referred to as the Last Universal Common Ancestor, or LUCA. What that LUCA exactly is we do not know. It may not even make sense to have a single LUCA at the root of the tree.

Earth is a planet of bacteria and archaea, they are far older than any other lifeforms and their biomass outweighs the combined biomass of all other life on Earth. We just have to look at ourselves to see how much, or better how little, we are in charge. For every human cell in our body there is at least one microorganism living in us and with us. We just very easily overlook them because they are so small and their combined weight in an average human body is small, roughly around 0.3% percent of a typical person's body weight. In places where they are concentrated in our bodies and perform a specific function, the medical profession has come to view them functionally as an organ, like in our intestines.

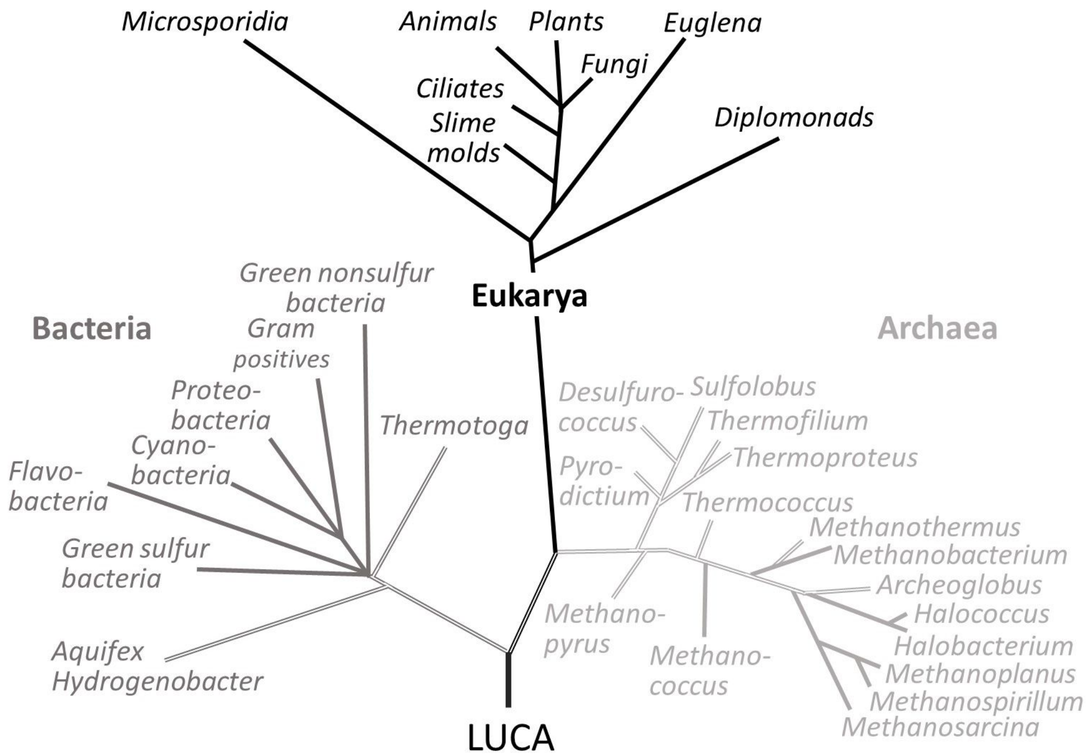


Figure 3.3: A simplified sketch of the universal phylogenetic tree of life with its three main branches, the domains of life, after Ref. [38]. Double lines lead to organisms that are heat loving. The acronym LUCA stands for the Last Universal Common Ancestor.

There are of course bacteria that are harmful to humans and that is what must be feeding the antibacterial frenzy flickering over advertising screens. However, most bacteria are not harmful to us and many bacteria live in symbiosis with us. This should not be surprising because we and other higher lifeforms have evolved on a planet of bacteria and a symbiotic relationship with most bacteria is a consequence of this co-evolution. The way our human body works today, it could not do so without its resident bacteria. It is as simple as that: without bacteria, we would not be here in the first place. Many of the kingdoms of life shown under the three domains of life are likely not familiar to most of us. On the side of the eukarya we can find some more familiar names identifying the kingdoms of animals, plants, or fungi; mushrooms being prominent members of the latter. Most of us would correctly guess that microsporidia must be related to fungi as indeed these large single-celled organisms are. However, there are also several names likely unfamiliar to many; and even if their names sound familiar, like for example slime molds, they do not readily connect to our daily experiences. We will get to them later. On the bacteria side we already encountered the cyanobacteria in the previous chapter, and we will meet them again. Names in the archaea domain are likely even less familiar, if any of them are at all. What may be confusing is that there are archaea that carry the word bacteria in their name. That of course has historical reasons as it was not always clear which branch belonged exactly where; and the domain of archaea is sometimes still referred to as archaebacteria, the original name given to it by Carl Woese.

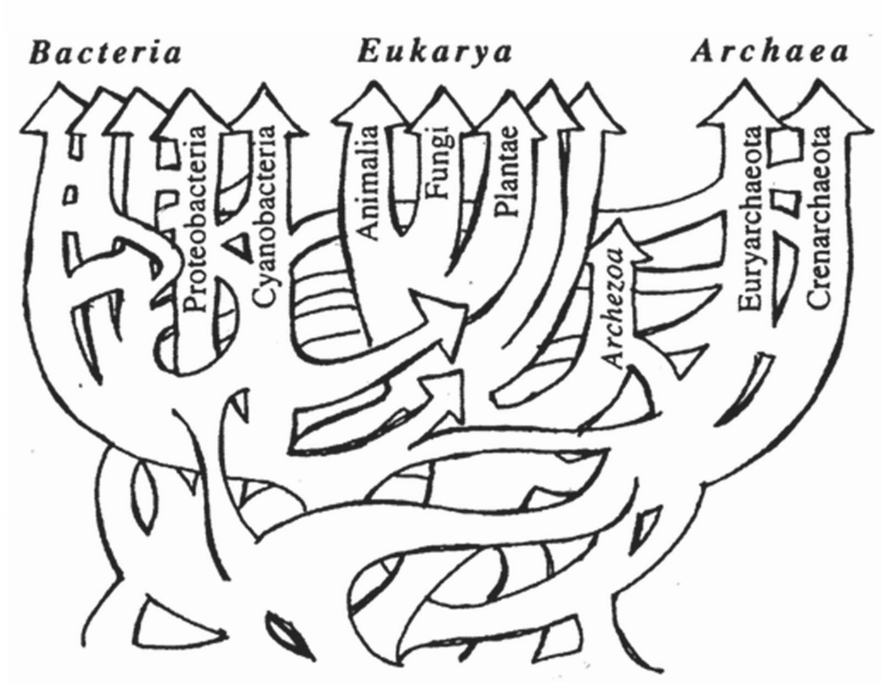


Figure 3.4: An example of a tree of life reflecting horizontal gene transfer after Ref. [39].

On the other hand, bacteria are sometimes referred to as eubacteria, as the true bacteria. The whole picture of how the respective branches in the three domains relate to each other, specifically in the bacteria and archaea domains is still evolving. Here, genetics is making a major difference as it provides the tools that can reveal how different lifeforms really relate to each other. We just have seen in the last section the great extent to which bacteria thrive on horizontal gene transfer; no less is true for archaea. The tree of life version shown in fig. 3.4 gives an idea as to how horizontal gene transfer limits the usefulness of thinking in terms of a tree of life concept; more so for archaea and bacteria than for eukarya [39]. Some of the names in the universal phylogenetic tree of life in fig. 3.3 also give a clear indication as to what the energy sources are which lifeforms rely on, specifically in the archaea domain where we find branches that use methane or sulfur as their primary source of energy; as do also some bacteria. It may not be a bad prediction that within a few generations our picture of the bacteria and archaea domains will be much richer and more differentiated as the vast majority of bacteria and archaea species are still very much unknown to us and we are only at the beginning of discovering how sturdy some of these lifeforms really are. As we have seen, using the species concept with bacteria can be treacherous and this is also the case with archaea. The species concept really only gets on more solid ground with the higher animal lifeforms including humans but even there we have learned by now that interbreeding between what were supposed to be different human species did happen; not only once, but multiple times. When using the species concept, we have to be mindful of this. However imperfect, it still is the language of biology and nothing more precise has replaced it yet. Precision in nomenclature is not what understanding life is about and we can still have very fruitful discussions using the current vocabulary of biology.

We just need to acknowledge that at times we are talking about species only in approximations. The universal phylogenetic tree of life in fig. 3.3 contains one more information. Some of the tree branches are not drawn as solid lines but as double lines, indicating branches on this tree of life that lead to heat-loving organisms. These organisms are only found among bacteria and archaea but not in the domain of eukarya; and the archaea branch includes a significantly higher number of them than the bacteria. The two domains of archaea and bacteria are single-celled lifeforms which are jointly referred to as prokaryotes. The name prokaryote is a combination of the Greek words for before, *pro*, and kernel or nut, *karyon*, indicating that a prokaryote cell does not contain a membrane-bound nucleus. Fig. 3.5 shows a simple schematic of a prokaryote cell. All cells have a cell membrane. It is the outermost surface in animal cells, separating the cell from the external environment. Some cells have an additional cell wall as their outermost layer; this includes most bacteria and archaea; among eukaryotic cells, fungi, algae, and plants have them. The function of the semi-permeable cell membrane is similar to that of our skin. It controls what gets in and out of the cell and is the means by which a cell interacts with its environment. Prokaryotic cells can also have multiple cell membranes separated by a small space between them. The function of the cell wall present in non-animal cells is mostly to provide strength and rigidity and to prevent large molecules from entering the cell. Ribosomes, their name combining the words for ribonucleic-acid and the Ancient Greek word for body, *soma*, are complexes of RNA and protein. Their function in the cell is the assembly of proteins. They get their instructions from the DNA, which they translate into the processes that then build the required proteins. The single large circular strand of DNA in the cytoplasm of a prokaryotic cell contains most of the genes needed for cell growth, survival, and reproduction.[27] The cytoplasm in prokaryotic cells is a gel-like fluid substance in which all of the other cellular components are suspended. For a long time it was thought that unlike eukaryote cells, prokaryote cells do not have a cytoskeleton. However, this has turned out not to be true and we now know that prokaryote cells also have the structural filaments referred to as cytoskeleton. In a sense, the cytoskeleton is the highway system along which the inner soft part of the cell can move around and it also helps a prokaryote cell to divide and maintain its shape.

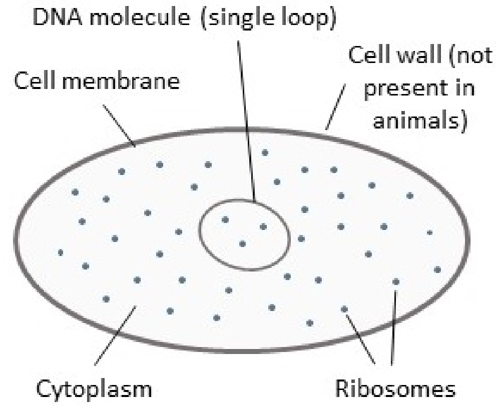


Figure 3.5: The basic structure of a prokaryote cell.

In the previous sections, we have seen some of the history behind the nomenclature that biology uses for the classification of life. Tab. 3.1 illustrates the eight-level hierarchy that biological taxonomy employs to classify lifeforms, using the example of modern humans in this case. All of us are likely familiar with the term species. Species is the lowest classification level that biologists or botanists use in this taxonomic scheme. Most of us have heard the term *Homo sapiens*, used to describe modern humans. The species designation in *Homo sapiens* is the *sapiens* part and *Homo* is the genus specification.[36] The genus classification is the next higher up from the species classification.

Table 3.1: The biological classification scheme of life as applied to modern humans.

Taxonomic Rank	"Us"	Description
Domain	Eukarya	A cell with a nucleus containing DNA (different from Procaryota)
Kingdom	Animalia	Eukaryota that are not plant, fungus, or protista*
Phylum	Cordata	Animals with a backbone containing a nerve cord
Class	Mammalia	Cordata with fur or hair, producing milk, and giving birth to live young
Order	Primates	Mammals with collar bones and graping fingers
Family	Hominidae	Primates including humans, chimpanzees, gorillas, and orangutans
Genus	Homo	Hominids with upright posture and large brains
Species	Sapiens	Modern Human

**Note: Catch-all for any eukaryotic organism that is not an animal, plant, or fungus.*

Genus refers to a group of species that have similar body blueprints, which in our case means the group *Homo* that includes all human species. Next up in the hierarchy is the family classification. We fall into the family of hominidae that includes humans, chimpanzees, gorillas, and orangutans. The next higher level is the order classification, which for us is primates. In addition to humans and the great apes in the hominidae classification, the order of primates also includes the lesser apes such as gibbons, old and new world monkeys, but also lemurs. Above the order category sits the class category, which for humans would be mammals. The next higher classification level is the phylum which relates for example to broadly shared body plans such as vertebrates, or cordata in Latin, to use the correct term. Cordata are essentially all animals that have a backbone, or a precursor of it, containing a nerve cord. With that, we are already at the kingdom classification level where we modern humans fall into the animal kingdom. The highest classification level is that of the domain which has only the three members, archaea, bacteria, and eukarya. Archaea and bacteria are grouped together as prokaryotic single-celled microorganism whose cells have no nucleus. In contrast, eukarya includes all single-celled life that has a cell nucleus and membrane-bound organelles as well as multi-cellular organisms such as us. This brings us already to the end of our brief digression into the topic of biological taxonomy. With this basic vocabulary in hand, we are equipped to start our excursion into the history of life on Earth.

The Origin of Life Question

Figure 3.6: Fossil stromatolite from 3,400-million-year-old rocks in the Strelley Pool Chert in the Trendall Locality, a Geoheritage reserve in the Pilbara region of Western Australia (a) and modern stromatolites growing in the close by Shark Bay area (b). Photographs courtesy Geological Survey of Western Australia [40].

Now we are ready to start our journey through the evolution of life on Earth where our itinerary will include only rough time estimates, pointing out the significant markers in the development of life on Earth that we can relate to without getting lost in detail. After all, the objective here is to gain a perspective of the bigger picture of the evolution of life on Earth. As we have learned in the previous chapter, the oldest minerals on Earth we have found so far date to roughly 4,400 million years ago. The analysis of these zircon minerals has provided solid evidence that they must have been interacting with water, resulting in the conclusion that Earth at that time already had a stable hydrosphere.[37] Life never seems to be far behind once water is present. So life could have emerged on Earth as early as 4,400 million years ago and probably no later than maybe 3,700 million years ago when there seems to be first potential trace evidence for life. The first fossil evidence we have for the single-celled organisms we call prokaryotes, archaea or bacteria, is more recent, maybe some 3,400 million years old. Potential indications of life before 3,700 million years ago are still being evaluated but paleontologists cannot substantiate yet a date closer to the earliest possibility for life supported by the presence of water. The challenge is that in order to find fossils of a certain age, one must find rocks of that age. The oldest rocks found so far are dated to about 4,280 million years ago.[38] Finding fossil traces of life older than the oldest rock formations we know is quite impossible. In the 1980's, a structure thought to contain fossilized cyanobacterial remains was found in the 3,500-million-year-old Apex Chert Formation in Australia. For a long time, the organic origin of those structures continued to be disputed. Eventually, in 2011, with the help of improved imaging and analytic techniques, they were conclusively shown to be of inorganic origin. However, around the same time new fossil traces of early life were discovered in the Strelley Pool Chert formation of Western Australia. Fig. 3.6a shows an example of the kind of stromatolite fossils geologists and paleontologists find in the Strelley Pool Chert of the Trendall Locality, a Geoheritage Reserve in the Pilbara region of Western Australia. Stromatolites, frequently described as mounds or pillows, have their origin in microbial mats built by cyanobacteria, layer upon layer.[39] As successive new layers grow on the buried layers, the latter mineralize over time as the cyanobacteria in the top layer continue to photosynthesize while washed over by the sea in the shallow pools where the stromatolites grow. It is such stromatolite fossils as shown in fig. 3.6a that preserved some of the oldest evidence for life on our planet including about 3,400-million-year-old cellular lifeforms with a sulfur-based metabolism [41]. A few places around the world are still home to living stromatolites. The photograph in fig. 3.6b shows one such place, Shark Bay, also in Western Australia and not far away from the Pilbara region where the stromatolite fossil in fig. 3.6a was found. With respect to stromatolites, Western Australia is unique. Its geology has not only preserved some of the very oldest stromatolite fossils we know of but at the same time it is still home to living stromatolite-building microbial communities today.

A few years after the discovery of the 3,400-million-year-old sulfur-based cellular lifeforms in Western Australia, stromatolite-like structures were found in 3,700-million-year-old rocks from the Isua Supracrustal Belt in southwest Greenland. If confirmed, they would indicate the presence of such early microbial life in shallow marine environments. Even earlier traces of life may have been found in Canada. A form of carbon derived from an organic precursor, referred to as biogenic graphite, has been found in rocks from the Saglek Black Formation in Labrador, Canada. We do not know yet, which lifeforms left the traces of biogenic graphite, but they would have lived around 3,950 billion years ago. Recently, the discovery of microscopic filaments and tubes in rocks found along the eastern shore of Hudson Bay in northern Quebec may point to bacterial life being present on Earth as far back as 4,300 million years ago. The evidence supporting this interpretation is still in dispute. However, if proven correct, microbial live may have existed only a few hundred million years after the formation of Earth. The environment supporting such early life could have been that of a nearby vent in the seafloor with the surrounding water heated by volcanic activity. Clearly, as the date for the oldest potential life on our planet pushes up against the age of Earth itself, we can expect that evidence for such early life will get ever more difficult to find. Even if we find tangible indirect evidence indicating the presence of life, it may probably be next to impossible to identify the kind of life that could have left such traces.

Finding ever-older evidence for life may help us to resolve the biggest mystery of all. How did life first come to be on Earth, and when and how did the transformation from lifeless matter to the first lifeforms occur? Life did certainly not start with the oldest fossil finds we know of today. There must have been other lifeforms preceding those that eventually evolved into the lifeforms that currently are the oldest records of life we have. We do not know how many different such lifeforms there were and how many of them were in the path of the evolution that left us the fossil record we have. Given the propensity of life to flourish even under the most challenging circumstances, it would be a mistake to rule out that other primitive lifeforms may have existed that we never will know about. It could well have been that at some point in Earth's early history, life got started but became extinct, maybe that could have happened not just one time but multiple times throughout something like the first 2,000 million years of life on Earth. There may not be an actual continuity of lifeforms represented in the earliest fossil trace records of life. For all we know, there could have been multiple beginnings and multiple failures. Life had almost 2,000 million years to emerge between the presence of first water on Earth and the beginning of the so-called Great Oxygenation Event some 2,500 million years ago. Cyanobacteria were instrumental in creating the environmental conditions that could support the evolution of more complex lifeforms. When cyanobacteria started to oxygenate our world, they already had been around for more than 1,000 million years. A lot could have happened between that appearance of first water on Earth maybe as early as a couple of hundred million years after Earth formed and the fundamental changes that started after some 2,000 million years into Earth history, enabling the evolutionary path that eventually produced us. Many lifeforms could have evolved, only to disappear forever without a trace left behind. We may never fully know the richness of life our planet has witnessed, but we know that one of those many path that life took eventually would lead to the evolution of higher lifeforms some 540 million years ago. To discover how life first appeared on Earth we must understand the transition from whence we are talking about physics and chemistry to where it becomes biology - how life can arise from lifeless matter. How do we define lifeforms in broad and general terms to begin with? A widely used definition is for example:

> *"Lifeforms are open systems far removed from the thermodynamic equilibrium, subject to a Darwinian evolution process with intrinsically faulty information transfer"*

This definition combines physical and biological aspects. Let us inspect it a little closer. The definition of an open system is that it has external interactions which can include energy, mass, or information transfer in and out of the system. The exchange of energy, mass, and information between an open system and its environment can take different forms. For example, energy exchange can also be an information exchange. As systems replace lost energy, they can at the same time acquire information. Different wavelength of light or varying environmental temperatures can for example provide sensory inputs at different levels. Arguably, for lifeforms it would seem that mass, energy, and information in the most general terms are the relevant parameters to interact with their environments so defining them as open systems in this sense seems reasonable. Thermal equilibrium is a physics definition. Of course, lifeforms have to be systems far away from thermal equilibrium as they must not only gain mass, energy, and information from their environment but must also transform mass, energy, and information into forms they can use in their struggle for survival. That certainly is an everyday experience we can all relate to and this is no different for more primitive lifeforms such as bacteria than it is for us. The last part in the above definition of lifeforms is probably familiar to most of us; it reflects on how the evolutionary processes which we discussed earlier act on species populations under natural selection. This definition of lifeforms does not really refer to populations but that is self-evident as the whole evolution concept acts on populations and makes no sense if one only looks at individuals. Definitions such as the one above are helpful in exploring the various aspects of life that are part of that definition. However, they do not really help us much in providing a satisfactory system definition of life. Rather, they describe what we seek to understand and help us to develop fundamental studies of lifeforms in systematic ways. We briefly outlined here only one approach that science takes to define life. There are many others but none really satisfactory or helpful to understand how lifeless matter can transform into what we call life. Maybe this is one of the instances where we cannot define something with the accuracy we would like because we fundamentally do not understand it yet. It seems to be the better approach to focus on what we can understand and hope that eventually enough pieces of such incremental understanding will accumulate into the equivalent of a critical mass of knowledge. At which point we may cross the threshold that allows us to start understanding the fundamental aspects underlying the transition from lifeless matter to life on Earth.

The critical components of the organic life we know are proteins and their constituent amino acids plus the molecules that can actually synthesize them, the RNA molecules. In order to understand how lifeless matter can transform into life we need to understand how biomolecules such as RNA evolved in the first place. How could Earth transition from a world without RNA to a world with RNA-based life? What are the physical and chemical precursors, or the physical and chemical evolutions, which needed to precede biological evolution? From what we know today, we cannot answer a number of major questions yet, all we have so far are hypothesis at best. Where do the molecules forming the basis of Earth's life, the amino acids, come from? How did the building blocks required for life assemble from amino acids and what are the processes that underlie such transformations? How did the transition from building blocks of life to life itself happen, how did the capability of self-replicating protein-synthesis evolve? Darwin never published anything on the origin of life itself, but he shared his thoughts in some of his many letters to close friends. In 1871, Darwin said in one such letter:[40]

> *"...it is often said that all the conditions for the first production of a living being are now present, which could ever have been present. But if (and oh what a big if) we could conceive in some warm little pond with all sort of ammonia and phosphoric salts,—light, heat, electricity present, that a protein compound was chemically formed, ready to undergo still more complex changes, at the present such matter would be instantly devoured, or absorbed, which would not have been the case before living creatures were formed ..."*

Some eighty years after Darwin shared this thought with a friend and some seventy years after Darwin died, this is in a way exactly the scenario that scientists started to explore. In 1953, the American chemists Stanley Miller (1930 – 2007) and Harold Urey (1893 – 1981) carried out a landmark experiment that came to be known as the Miller-Urey experiment. Their objective was to understand how the basic molecules needed for life to emerge could be created in an artificial early Earth atmosphere. At the time, the assumption was that the early Earth atmosphere was composed of water, methane, ammonia, and hydrogen. The Miller-Urey experiment demonstrated that an electrical discharge in a gaseous mixture of those components could indeed produce a variety of organic molecules including about half the amino acids required to synthesize proteins. However, today we know that the early Earth atmosphere was likely quite different from what Miller and Urey assumed back then. Earth's early atmosphere is now thought to have been a mixture of nitrogen, carbon dioxide, water and some sulfur dioxide. In such an atmosphere, it is difficult to synthesize the amino acid molecules created in a Miller-Urey type of experiment.

Several alternative scenarios have been proposed as to how Earth may have acquired the complex organic molecules necessary for life to start. We know for example that meteorites do contain organic materials. Therefore, it is conceivable that some of the critical organic components could have been contributed by meteor impact instead of being synthesized in a primordial soup or gas such as in the Miller-Urey experiment. Those organic materials contributed by meteor impacts could over time have enriched hydrothermally heated shallow waters. Combined with abundant UV light this would have supported a complex chemistry. Essentially the primordial soup, the warm little pond Darwin had speculated about back in 1871.

Intriguingly, as already mentioned in our earlier discussion on stellar evolution, more recently scientists have discovered that our universe may be seeded with organic molecules resembling amino acids to start with. Seemingly, amino acids are already present in stellar nurseries, the dust clouds from which stars are born. It is quite impossible that any such amino acids from the stellar nursery that seeded our solar system would have survived the formation of Earth itself, which as we learned earlier was initially in a molten state. However, amino acids from the stellar nursery forming our solar system may well have been preserved in the primordial debris of our universe that still orbits our Sun out there in the asteroid belt located between Mars and Jupiter. It is the fragments of these countless asteroids which sometimes hit Earth as meteorites. No wonder that some of them do contain amino acids.

In many ways the deep ocean floors have been as unreachable for us as the stars, but this started to change over the second half of the twentieth century. Exploring the deep-sea floor is still a huge technical challenge but we have made a start. The little that we have seen already has changed our minds about life on Earth. This includes the discovery of hydrothermal vents, so-called black smokers, on the bottom of the deep sea surrounded by lifeforms we have never seen before. Hydrothermal vents are nothing new but the ones we were familiar with were all closer to the surface where the life around them still benefits from some sunlight coming through and with it from photosynthesis. Water absorbs light and with increasing depth, less and less sunlight makes it through until eventually at a depth of approximately two hundred meters photosynthesis is not possible anymore. Therefore, life on the deep-sea floors cannot use photosynthesis or rely on the oxygen it produces to sustain itself. Instead, lifeforms on the deep-sea floors around hydrothermal vents harvest the energy they need to sustain themselves from sulfide minerals in a chemical way, feeding on the planets energy instead of depending on the Sun. Maybe some kind of hydrothermal vents were also the sources of the first amino acids on Earth;

of course, we do not know that, but we cannot exclude it either. That we only learned about these new lifeforms so recently should caution us. There is still a lot about life that we do not know, and we have no good idea as to how many other surprises may still lurk in the depth of our Earth's oceans.

The more we look for potential sources of amino acids the more we seem to find. While at first the availability of amino acids looked like a major hurdle towards understanding how life may first have evolved we now have learned that there are multiple pathways in which Earth could have been seeded with amino acids. This does not tell us why and how exactly the roughly twenty amino acids that are the foundations of life on Earth do form in the first place. However, with several probable sources for such amino acids to seed Earth we are getting an increasingly more detailed perspective on the potential starting point of the chemical evolution that could have preceded the formation of RNA on Earth.

It is a very big step akin to the proverbial spark of life, which takes us from chemical reactions producing the organic building blocks of life to a self-reproducing biochemistry. Scientists bridge this gap with something they refer to as collective auto-catalytic networks.[41] What they are envisioning is an RNA world where a set of simple RNA molecules act as carriers of information and as their own mutual catalysts allowing the set of RNA molecules to self-reproduce. That may sound like a circuitous approach, but it is actually not when one considers that the environment must play a role in any such reaction. It may be that a set of simple RNA molecules could not reproduce in one type of environment while in another environment that same set of RNA molecules could reproduce in an auto-catalytic reaction. Maybe some RNA molecules in contact with a surface containing certain minerals were more likely to reproduce via auto-catalytic reactions than RNA molecules without such help. Once that happens, we could look at a runaway proliferation of RNA, or put differently we could look at first primitive life. While much of this is still speculation today, it does outline one potential pathway for life to emerge and there may be others.

If this somehow sounds like the survival of the fittest RNA molecules, this is precisely what it is and that brings us to the hypothesis of the RNA world. The conjecture behind this hypothesis is that before the first single-celled organism evolved there was a time during which the RNA molecule that eventually ended up inside a first cell evolved from simpler RNA-precursors. There are now a good number of laboratory experiments demonstrating that natural selection also works for self-replicating molecules such as RNA. Evidently, RNA molecules that through some kind of chemical modification, which a number of events could cause, acquire an increased ability for faster self-replication can quickly out-compete other RNA molecule "species". The RNA molecules we find in cells today are rather complex so the assumption is that the first auto-catalyzing RNA-precursor molecules where quite a bit simpler. One of the big questions is if there could have been more than one RNA molecule version successfully self-replicating in such a chemical auto-catalytic evolution from simpler RNA-precursor molecules. The answer is likely yes, this may have happened, but eventually there must have been a winner-takes-all outcome in the end, or otherwise we would not live in the world we see around us today. Without such a winner-takes-all outcome, we should be finding RNA molecules that do not share the ancestors of the RNA molecules from which all life we know today

seemingly originates. An interesting but unanswerable question is if with a different RNA molecule to start with, evolution may have taken a very different course in expressing its solutions for life to fill the respective environments that Earth provided.
Regardless, even if the answer to this hypothetical would be yes, it still would have been life, just a different one. However, there are implication for how life elsewhere in our universe could evolve. If life gets its start through something like an RNA world where natural selection is at work, then there could certainly be different outcomes. However, the range of such outcomes would still be constrained. To start with, one could ask questions regarding the probabilities for potential non-carbon-based equivalents to our RNA world existing in our universe. Is there a possibility for a different chemistry that could lead to a similar self-replication process? Like one for example based on silicon, which seems to be a favored option in science fiction movies. We may not be able to answer this question definitively, but we should be able to assign probabilities for other than carbon-based life in the universe. Would it not be good to know that most likely other life in the vast expanses of the universe is also carbon based, facing similar difficulties as we do as a species to become a space-based species?
At some point in time, something must have happened that resulted in RNA finding itself enclosed in a protective membrane. This event must have provided an evolutionary advantage or else life - if we call the RNA world that preceded this event the earliest life - would have been unlikely to evolve much further. In which case, Earth could have been stuck in an RNA world for much longer or worse, cellular life may never have evolved in the first place. There is still so much that we do not know yet about early life. Darwin must have felt this keenly as we can read from his correspondence on this topic to his trusted friends. Even today, his warm little pond scenario is no less likely than it was in Darwin's time, the major difference really is that we can explore such scenarios now with a very different tool set and understanding than would have been available to scientists in Darwin's time. We can take hope from the fact that as compared to what we knew only a few generations ago about how life came to be, we have learned an enormous amount in a short time. There is every reason to believe that this will continue and that at some point in time we will know how live evolved, we may even know at some point how to create life but that and the potential implications of it pose very different questions.
To understand fully how life on Earth first appeared we may need to learn how life can evolve on planetary systems more generally. Astrobiology, the study of life in the universe, is a science discipline that has evolved out of this effort over the past decades. Such exploration should help us to learn more about the prerequisite physical and chemical evolution that potentially can create favorable conditions for first life on other planetary bodies. Even today, some planetary bodies in our solar system may support conditions for first life to emerge. Maybe some day we will also be able to take our quest to understand the origins of life beyond the boundaries of our solar system. Anything we will learn on this scientific journey could help us understand how first life on Earth got started. For now, we have to accept that there is still a very big gap in our understanding of the earliest life on Earth. As we move on, we will continue our journey with the first fossilized life on Earth that we know of today.

The First 3,000 Million Years of Life on Earth

Let us return to the fossil evidence and what it can tell us about Earth's early life. More or less, even though sometimes with sizable gaps, it looks like we can trace the evolution of life from these earliest fossils to later life. It seems we can do this to the extent that we do not have to ask the question if life evolved more than one time. However, this changes once we consider that suitable conditions for life existed much earlier, long before the ~3,700-million-year-old first fossil evidence for life. As we have seen, Earth had oceans maybe as early as 200 million years after the planet had formed. We also know that amino acids would most likely have been available, so life could have moved on to proteins and from there to nucleic acids much earlier. However, more likely than not, the late heavy bombardment could have wiped out any early life Earth may have had. As we saw in the previous chapter, the massive pounding of Earth by meteorites and asteroids we call LHB started around 4,100 million years ago and lasted some 300 million years until about 3,800 million years ago. The speculation is that this LHB period was so extreme that the heat it produced evaporated Earth's oceans and our planet only regained its liquid surface water once this pounding subsided and the evaporated water could condense to form new oceans. If that is the case, then practically the evolution of the first life we know happened in roughly 100 million years, between the LHB period ending around 3,800 million years ago and when we find earliest fossil evidence for life dated to 3,700 million years ago. Turning this argument around, if life could have evolved in some 100 million years, given the right conditions, there should have been enough time, some 200 million years, for life to have started before LHB may have put an end to it.

One of the critical questions for early life that are still unanswered is how prokaryote cells acquired their membranes. If we take the above scenario, sometime between 3,800 to 3,700 years ago, prokaryotes with cell membranes must have made their appearance. How they did that we do not know but maybe in this 100-million-year time window Earth's geology could have preserved some evidence that we have not found yet, some kind of biomarkers that provide an indication of when that happened and how. The first cell membranes must have been quite different from the complex cell membranes we find today with their sophisticated control mechanisms that regulate what can get in and out of a cell. The early membranes of these protocells must have been porous and maybe the first were not continuous membranes at all but more like the viral envelopes found with some viruses today.

By some 3,700 million years ago, when we potentially have first fossil evidence for life, scientists believe that the two oldest domains of life, bacteria and archaea had already split. The cell membranes of today's bacteria and archaea seem to indicate the existence of a protocell LUCA with a simpler membrane structure from which the differences in archaea and bacteria cell membranes originate. Scientists are actually working to reverse engineer such a unified cell structure of a protocell LUCA in their laboratories. Therefore, some day we may be able to recreate how nature evolved cell membranes and how DNA first became enveloped into what would evolve into the membranes that provided the protected environment for complex cell chemistries to evolve; cell chemistries on which all of Earth's life relies on. Microbial life has evolved cell chemistries that seemingly can adapt to almost any imaginable external energy sources. We know microorganisms that live off methane, sulfur, and of course of light, but there are also bacteria that can digest metals, even toxic ones. However, the energy source that most of Earth's life relies on is oxygen and the story of oxygen starts with cyanobacteria.

Cyanobacteria, more commonly known as blue-green algae, are the only prokaryote organisms that can produce free oxygen through photosynthesis. Before cyanobacteria began to transform Earth's atmosphere there was no free oxygen in it. We know this because of early rock formations that contain many elements such as iron and uranium in their reduced states, which is in their pure metallic form and not oxidized. In rocks younger than 3,000 million years, we start finding those metals in their oxidized states. This means that there cannot have been free oxygen available before 3,000 million years ago. The oldest undisputed cyanobacteria fossils are dated to 2,100 million years ago and there is some evidence that dates their presence back to as early as 2,700 million years ago. Scientists believe that cyanobacteria are quite certainly much more ancient, likely as old as the first stromatolites that we discussed earlier that we start to find between 3,400 to 3,700 million years ago. Just as today's stromatolites in places like Shark Bay in Western Australia are built by successive layers of cyanobacteria this is also assumed to have been the case back then with the first stromatolites we know of. However, these first stromatolites were likely not built by the same kind of photosynthesizing and oxygen producing cyanobacteria we find later.

Fig. 3.7 shows the high and low estimates for Earth atmosphere's oxygen content according to Ref. [42]. There is still much that we do not know about how and why the oxygen level of the atmosphere has varied in Earth's distant past. By reading the stromatolite fossil record and using conservative dates, we can bracket the start of oxygen production by cyanobacteria between 3,400 million years as the earliest date and about 2,700 million years as the latest date. This leaves us with a 700-million-year-wide window of uncertainty. To get a sense of just how big this is, we only need to remember that 700 million years is about as far back from today as we find the first so-called higher lifeforms in the fossil record. Or, put differently, 700 million years is more than enough to replay the whole evolution of higher lifeforms as we find it in the fossil record today.

Admittedly, we do not exactly know when cyanobacteria acquired their capability to generate oxygen through photosynthesis. Before they did so, Earth was a planet without free oxygen. However, that certainly does not mean that it was a lifeless planet as cyanobacteria must have lived off another energy source before they began to photosynthesize. As we can see in the tree of life (fig. 3.3), some archaea and bacteria rely on sulfur and methane for sustenance. More recently, we also learned from new lifeforms discovered around deep-sea hydrothermal vents that besides burning oxygen there are other ways for organisms to gain the energy they need to sustain themselves.

When cyanobacteria first released oxygen into the early oceans, all of it would have been readily absorbed by decomposing organisms or would quickly react with iron suspended in the water to form iron oxide. Rusted layers of iron, once deposited on the ocean floors, became the massive rock formations called BIF we already encountered in the previous chapter. They are the sources of most of our planets commercial iron ore. BIF's have been built through successive layer depositions of iron rich materials alternating with the deposition of other sedimentary materials producing bands of magnetite (Fe_3O_4) or hematite (Fe_2O_3) alternating with layers of compressed silica such as chert.[42]

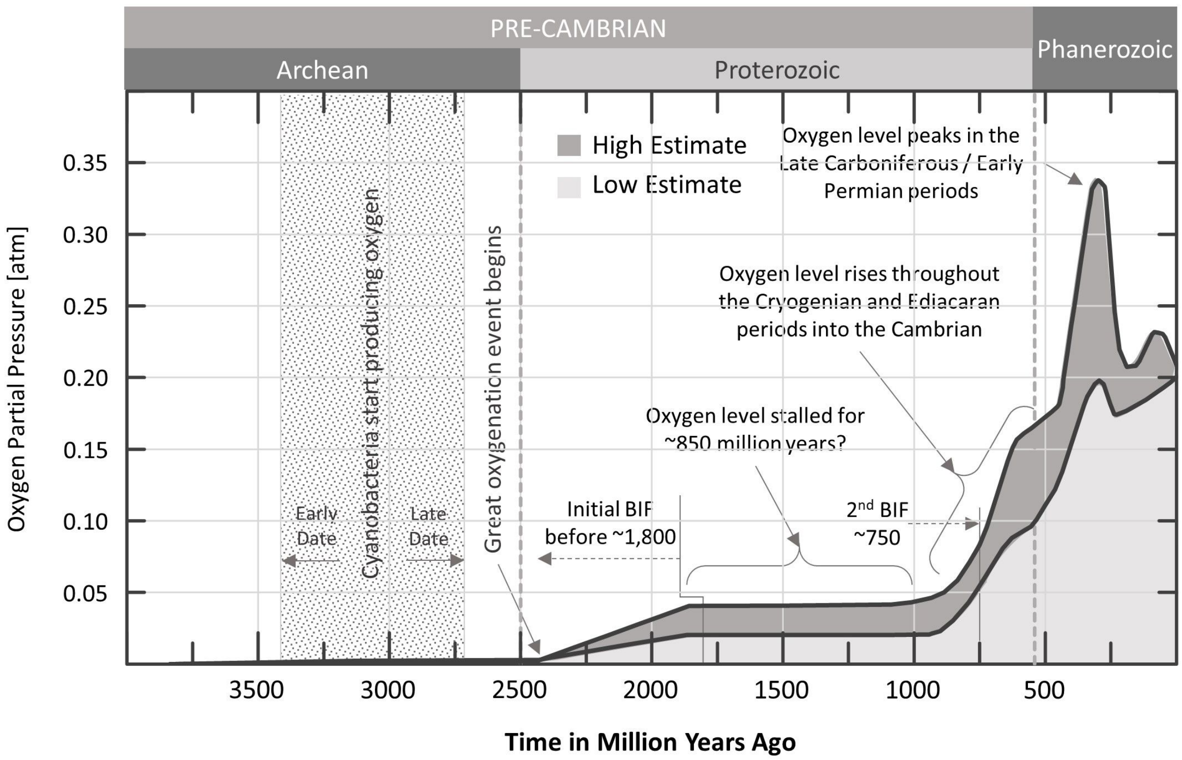


Figure 3.7: High and low estimates for the oxygen partial pressure in Earth's atmosphere for the past 3,800 million years after Ref. [42]. The initial or first BIF phase could have started as early as ∼3,400 and no later than ∼2,500 million years ago.

Once the vast amounts of suspended iron were consumed and deposited on the sea floors, oxygen could start to accumulate in the atmosphere. We are not sure yet what color the oceans had before cyanobacteria started to release oxygen, but that would likely have depended on the kind of color pigments used by photosynthesizing bacteria at the time. Scientists believe that iron oxide containing colloidal nanoparticles were a byproduct of the BIF deposition. These suspended nanoparticles, essentially hematite, seemingly turned Earth's oceans red, creating what is sometimes called *Red Earth*.

When oxygen finally began to accumulate in the atmosphere around 2,450 million years ago, it marked the beginning of the Great Oxygenation Event. At the start of the GOE, Earth's atmosphere was rich in methane produced by constant volcanic activity; and of course, it did not have an ozone layer. The lack of an ozone layer meant that Earth and early life on it was not protected from UV radiation as it is today. Because of that, life is thought to have been constrained to the shallow oceans, where the water was deep enough to shield organisms from UV radiation; and at the same time shallow enough for light to penetrate and photosynthesis to occur. The free oxygen in the atmosphere reacted with methane to produce carbon dioxide. Methane is a far more potent greenhouse gas than carbon dioxide and scientists believe that because of methane conversion to carbon dioxide, Earth's climate started to cool. Eventually this cooling period resulted in the Huronian glaciation, the longest episode of glaciation in Earth history. As we have seen in the last chapter, this led to the entire planet being frozen over in what is referred to as the first *Snowball Earth*, which is estimated to have lasted about 100 million years from 2,400 to 2,300 million years ago. As oxygen combines with methane, not only carbon dioxide is formed but also water which must have made for a more humid atmosphere initially but also produced more precipitation. In combination with a cooling climate, this would have increased snowfalls. Finally, with large enough amounts of oxygen released into the atmosphere the build-up of the ozone layer must have started. Once completed, the ozone layer would block biologically harmful UV radiation from reaching Earth surface.

The initial phase of BIF deposition may have started as early as when cyanobacteria first began to produce oxygen as a by-product of their photosynthesis. Why BIFs are composed of distinctive bands where magnetite or hematite layers alternate with layers of iron-poor shales or cherts is not clear yet; but it is reasonable to assume that the bands reflect oxygen fluctuations. Some kind of a seasonality in oxygen level, the origins of which remain elusive for now, must have been behind it. Unfortunately, we also do not know the periods of this seasonality, as we are not sure about the exact time duration between depositions of alternating bands in these banded iron formations. A less scientific and more poetic picture compares this periodic oxygen variation to the equivalent of a breathing Earth. The initial phase of banded iron formation lasted up until around 1,800 million years ago and may have started as early as when the first cyanobacteria began producing oxygen. Some rocks as old as 3,700 million years ago, have been interpreted as banded iron formations. If that is indeed the case, they could be as old as the earliest fossil evidence we may have for microbial communities, the ∼3,700-million-year-old stromatolite-like structures discovered in Greenland a few years ago. But maybe there was another mechanism that could have produced banded iron formations that we just have not discovered yet and that could have given us these very early banded iron formation rocks; only time will tell.

A second much shorter period of banded iron formation occurred around 750 million years ago when the oxygen level rose again significantly. Between the end of the initial banded iron formation phase and before the onset of this brief second phase there is a very long time span, roughly from 1,900 to 850 million years ago, where oxygen levels on Earth remained below less than one quarter of modern levels. Interestingly, this long time span of consistently low oxygen levels almost coincides with what frequently is referred to as the *Boring Billion*, a stretch of time from ∼1,800 to 800 million years ago named by paleontologists for the dearth of interesting fossil finds or other evidence for life that are seemingly characteristic for this *Boring Billion*. This is likely not a coincidence. However, one has to be careful not to read too much into this record as we only have very few data points for this roughly 1,000-million-year time span. If yours truly were a betting man, he would put his money on future generations of researchers unearthing much more interesting data that will allow them a vastly improved understanding of this long stretch of time; compared to the little we can make sense of today from the few data we have for those 1,000 million years.

Once the oxygen level started to rise again it continued to increase throughout the Cryogenian and the Ediacaran periods where we find the first solid evidence for more complex lifeforms. The oxygen level kept rising right into the Cambrian period and would do so until the Late Carboniferous and Early Permian epochs where at its peak oxygen made up roughly thirty-five percent of Earth's atmosphere in volume. It is the time in Earth history where we encounter lifeforms that could only sustain their large sizes because of the high level of oxygen content in the atmosphere. Insects would grow to the size of small birds something they clearly could not achieve in today's world with a roughly twenty-one percent oxygen level.

Fig. 3.7 shows that the amount of oxygen in Earth's atmosphere can change significantly over time but the time scale on which this happens can be so vast that it is hard for us to grasp the reality of it. Unfortunately, we lack critical information that could help us better understand the implications of a varying oxygen level. This includes something as basic as the weight of Earth atmosphere. Today, at sea level, we measure the pressure of the column of air on top of it as one atmosphere, which is how this unit of measure is defined. However, Earth atmosphere in the past can have been less dense or much denser than it is today. A lesser dense atmosphere means that the weight of the column of air above sea level is less and the pressure measured at sea level would be lower than one atmosphere; a denser atmosphere would result in a heavier column of air and therefore produce a higher-pressure reading at sea level.

That early in Earth history its atmosphere may have been much denser was a hypothesis considered to potentially resolve the conundrum of an early weak Sun and the early temperature record on Earth; something we discussed in the previous chapter. A denser atmosphere may well have amplified any potential greenhouse effect related to a higher carbon dioxide concentration. Changes in atmospheric pressure with altitude pretty much define where life is possible and where not; what matters is the amount of oxygen available. A less dense atmosphere even if it had a higher oxygen concentration may still result in a smaller total amount of oxygen available as indicated by the volume percentages. Where we have a fossil record, it helps us to get a sense of what the absolute amount of oxygen available to life could have been as compared to today. In that respect the gigantic insects of the Late Carboniferous, for example, tell a convincing story.

By creating their energy through photosynthesis, cyanobacteria were able to generate up to sixteen times more energy than their single-celled competitors remaining anaerobic. This likely enabled them to multiply much faster than their anaerobic competitors could do. Not only that, free oxygen, the byproduct of photosynthesis, also decimated their anaerobic single-celled competitors, until then the dominant lifeforms on Earth. Thus, cyanobacteria likely caused one of the most devastating mass extinction events on Earth. At the same time, cyanobacteria are also responsible for creating the conditions for large oxygen-consuming higher lifeforms such as us to be able to exist on Earth. The metabolism of such large higher lifeforms has a very high energy demand which could not be sustained without oxygen. Half of the oxygen produced on Earth today comes from phytoplankton, a diverse set of microscopic organisms to which cyanobacteria belong that live in watery environments, both salty and fresh.[43] The diatoms we discussed in the previous chapter in the context of Earth's climate, accounting for roughly one-fifth of natures oxygen, also belong to that group. There are very good reasons to consider cyanobacteria as one of the most important organisms, if not the most important, in the history of life on planet Earth.

The next milestone in life's evolution is the first eukaryote cell fossil record, dated to around 2,700 million years ago. As we saw above, organisms that fall into the domain of eukarya all have a cell nucleus that contains the DNA carrying their genetic information.[44] Without nuclear DNA, prokaryotes reproduced by cell division only but with eukaryotes having nuclear DNA, sexual reproduction could now enter the picture. However, for quite some time eukaryotes seem to have kept reproducing asexually through the process called mitosis, the division of a mother cell into two identical daughter cells.[21] It is thought that around 1,200 million years ago eukaryote cells started to reproduce sexually or to use the technical term, through meiosis.[21] How did eukaryotes acquire a cell nucleus? This question is not answered today but there seem to be three potential explanations. The first one is that prokaryotic cells simply produced another membrane, which surrounded the single loop DNA of bacteria. This would be an evolution along the lines of traditional vertical gene transfer. The two other possibilities involve a horizontal gene transfer. Prokaryotic cells engulfing other bacteria in a symbiotic embrace may have as well produced the nucleus. Scientists are quite certain that the organelles became part of eukaryotic cells via endosymbiosis and the same could have happened in the case of the cell nucleus. The third possibility is that the eukaryotic cell acquired its nucleus through the symbiotic absorption of a virus.

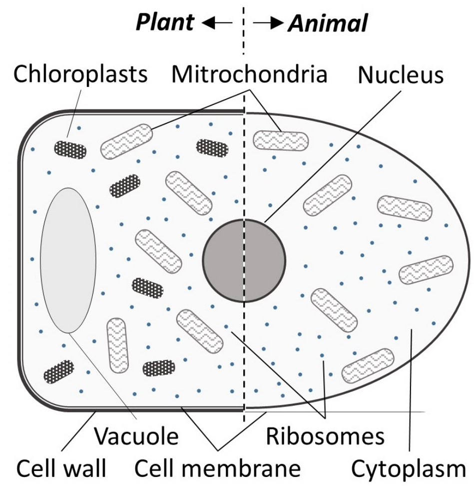


Figure 3.8: The basic structures of eukaryote plant and animal cells.

We know very little about where viruses fit into the phylogenetic tree of life as shown in fig. 3.3. Viruses existing today cannot replicate on their own and are very small and simple. As discussed earlier, the way these viruses reproduce is to transfer their genetic information into a host cell and have it produce new copies of the virus. However, that may not always have been the case. Viruses are likely descended from very early primitive organisms and the speculation is that on the phylogenetic tree of life they share a common ancestor with bacteria. However, while bacteria became more complex, viruses went the other way, reducing their complexity to the point where they only carry the most essential genetic information and use other lifeforms to reproduce. That certainly has been a very successful strategy.

For quite a long time and even sometimes still today, viruses have been described as somewhere between lifeless matter and living cells. The motivation for this is the fact that viruses cannot replicate by themselves, so they cannot continue their existence without other lifeforms. However, the same is true for practically all other lifeforms. Humans could certainly not survive without "using" other lifeforms. If an organism needs these other lifeforms to reproduce or just to stay alive in order to procreate, that is only a difference in degree. Viruses are lifeforms just as we are and because they have been around for much longer than us, we can say that they have been a lot more successful than we have yet demonstrated. We may have to wait a little longer but eventually scientists will find out how the eukaryotic cell acquired its nucleus. If it then should turn out, that the cell nucleus originated through the acquisition of a virus that may change our perception of viruses from something that we usually fear, into something that is also part of us. Which it already is, because as we saw earlier, our very own human DNA contains some small amount of viral DNA. However, viruses playing an integral part in creating the cell nucleus of eukaryotic cells, which form the basis of all higher life on our planet, would raise their role in life on Earth to a completely new level. Interestingly, our perception of viruses has already started to change in a different way, as they may provide a unique approach to fight cancer. Viruses can inject their genetic material into cells, as this is how they reproduce. One of the most recent developments in cancer treatments takes advantage of this by using a neutered virus to deliver genetic material into a cell that helps our immune system recognize cancer cells it otherwise could not detect as such and destroy them.

The domain of eukarya includes all higher plant and animal live. Not surprisingly, eukaryote cells are larger than prokaryote cells, otherwise they would not have been able to accommodate their many tenants they swallowed over time in the symbiotic embraces discussed earlier. The whole process must have started out with larger cells swallowing smaller ones and taking them in as permanent residents not unlike in a "big fish eats small fish" scenario. Sizes of prokaryotic cells range between fractions of 1/10,000 to 1/200 of a millimeter; for comparison, on average adult human head hair has a diameter of ∼1/13 of a millimeter. Typical eukaryote cells are quite a bit larger than prokaroyte cells and range in size between fractions of 1/100 to 1/10 of a millimeter. Fig. 3.8 gives a schematic of the basic structure of eukaryote cells where the left half of the sketched cell resembles a plant cell and the right half an animal cell. Plant cells tend to have distinct edges and frequently look rectangular and square shaped which is due to the thick cell walls they evolved from the ones they inherited from their prokaryote ancestors. In contrast, animal cells have only kept the cell membrane, which provides them more flexibility to change shape. The half-cells sketched here give a simplified picture of the eukaryotic cell as both plant and animal cells include more organelle types than shown. They also have much more detailed internal structures than would make sense to reproduce and discuss here. Of the organelles shown, there are also typically many more than drawn here, expect of course for the nucleus where there is always only one.

One such numerous tenant are the mitochondria.[45] They are the powerhouses of eukaryote cells and in essence the functional equivalent to our digestive tracts as they take in nutrients, process them and deliver the energy rich molecules the cell needs to sustain itself. Eukaryote cells have taken in many of these symbiotic bacteria turned mitochondria with most mammalian cells containing anywhere from about 1,000 to 10,000 mitochondria, depending on the specific cell type; mature egg cells can contain hundreds of thousands of them. Mitochondria multiply by simple division, splitting in two just as bacterial cells do. Their numbers, just like those of other cell organelles that kept their capability to self-replicate, are regulated by the specific needs of different eukaryote host cells and the environments they live in.

Both, prokaryotes and eukaryotes, did give rise to multicellular organisms. Prokaryotes did this in the form of colonies such as the bacterial mats that over time for example built up the structures we call stromatolites. However, only eukaryotes developed complex multicellular organisms. The oldest eukaryote multicellular fossil we have today goes back some 2,000 million years, some 700 million years after the first presence of eukaryote cells in the fossil record. Therefore, multicellular eukaryotes could have evolved much earlier than the 2,000 million years ago the fossil record shows us today. How much earlier we cannot say, as we have no idea how long such processes do take in nature. It could be very early, shortly after the first single-celled eukaryote lifeforms arrived. The first eukaryote cells show up in the fossil record about a billion years after we can detect the presence of the first prokaryote cells. Naturally, the same question arises: how long does it take prokaryote cells from swallowing one another in endosymbiotic embraces to evolve into eukaryotic cells with a cell nucleus and organelles? Does it really take about 1,000 million years or have we just not found the fossil evidence yet for eukaryote cells being around much earlier than what we know today? Which came first in the evolution of eukaryotes, the acquisition of the cell nucleus or the organelles? There are still many questions here where we have no good answers.

In addition, scientists keep finding more and more evidence that the inner workings of the supposedly simple bacteria are significantly more complex than we thought. It now looks like prokaryotes also contain organelle-like structures. They were harder to find because prokaryotes are so much smaller than eukaryotes, sometimes even smaller than some of the organelles we find in eukaryotes. The origins of organelle-like units in prokaryotes are not clear yet at all. Finding an answer as to how some of these structures in prokaryotes evolved could help us understand when eukaryotes really diverged from prokaryotes. How different was the first eukaryote ancestor already from a prokaryote before the first organelle tenant was acquired through endosymbiosis? We do not know, and we also still do not know what came first in eukaryote cell evolution, organelles or the cell nucleus. The latter could have happened via one of the three possibilities outlined above or maybe eukaryote cells acquired their cell nucleus in a way that we have not even imaged today. The when and how of the story of early life is still far from complete and much of the evolution of prokaryotes, their branching off into eukaryotes, and the arrival of the first complex multicellular organisms, remains a mystery.

For a long time paleontologists have struggled to understand how life progressed from the early single-celled lifeforms to what could be recognized as the first signs of animal life in the late Precambrian. There are still many gaps in our understanding of this transition, if one can even speak of transitions given that we look at a roughly 1,400-million-year gap between the earliest multicellular fossil and the earliest evidence we have for animal life. Scientists hypothesize today, that this gap may have been bridged by lifeforms called protista, or protists in English. The oldest organic microfossils we have today go back 1,800 million years, right to the beginning of the Statherian period of the Proterozoic eon. For such old organic microfossils, we have no good indication as to how they relate to any of the taxonomic groups we know; because of that, scientists refer to them also as acritarchs.[46] Most of these acritarchs are single-celled organisms with sizes ranging between a few thousands of a millimeter on the small side and up to a millimeter on the large side. Scientists assume that the majority of acritarchs were eukaryotes forming a phylogenetic group of organisms not necessarily closely related. Among others, they include for example phytoplankton such as the diatoms we encountered in the previous chapter as well as many protists.

We can see protists in Ernst Haeckel's tree of life in fig. 3.2 identified as a separate phylum next to animals and plants. This became problematic because members in the protista group turned out to be more dissimilar than was early on recognized. A closer look at fig. 3.2 also shows that animals were thought to diverge from protists but plants were believed not to be related to them. This as we now know erroneous view also gave the protists their name as it derives from protozoa, identifying what were thought to be "pre-animals".[47] Likely, the most well-known protist is the amoeba, which many of us may have heard about or seen in high-school biology, the giant amoeba.

An amoeba is a single-celled organism whose structure allows the protoplasma contained inside its cells to alter the shape of the cell. Which means that amoebas can move and use their pseudopodia, which are protrusions of the protoplama extending the cell in a desired direction, to move or to engulf prey.[48] Once the prey is engulfed all around by pseudopodia, it is brought inside the cell through constriction of the cell membrane. Many different kinds of amoeba have been found, some of them were found to belong to the ancestry of animals and some of them are part of the ancestry of plants; and others do not seem to belong into the two at all but are their own branch in the evolution of eukarya. Among the latter are the euglenia, diplomonads, ciliates, and slime molds shown in the modern phylogenetic tree of life in fig. 3.3. The challenge we face with protists is that they are all soft-bodied, so they rarely leave any kind of fossilized records.

Paleontologists face a similar difficulty when they try to narrow down when the first plant or animal life emerged. As we just saw, protists were at least around since 1,800 million years ago and likely longer than that, but plant and animal life only appears much later in the fossil record; more than 1,000 million years later. The oldest animal fossil records now suggests that primitive sponge-like creatures may have existed around 650 million years ago as indicated by fossil finds in South Africa; more recently a sponge fossil from southern China was dated to around 600 million years ago. This puts the first animal fossil towards the end of the Cryogenian period, just before we enter the last period of the Precambrian super-eon, which is the Ediacaran, lasting from 635 to 541 million years ago. Animals must have existed before then and may have existed even much earlier, but we may never find evidence of them in the fossil record.

The problem is that those early animal life forms were soft-bodied and therefore they did not have the same chance as hard-bodied animals with shells or skeletons to be preserved in the fossil record. Compared to a half-century ago we are already in a greatly improved situation because back then it really looked like there was no animal life predating the Cambrian period. With the continued discovery of new fossils, the boundary for the evidence of first animal and plant life may well shift further back in time. For now, there remains a large gap in our record of life between protists showing up and first animals and plants joining them. Currently that gap spans roughly 1,150 million years. That seems to be a very long time for protists to slowly crawl around in the seas until magically animals and plants appear. We need to remember, there was no life on land yet at all and for a long time, Earth's land surfaces would remain as barren as one can imagine.

The Evolution of Plant and Animal Life

The enormity of change Earth has gone through since the evolution of higher life began in the late Precambrian defies comprehension on any human-type scale. It is only very recent that humans as a species have accomplished to affect Earth's environments such as we do today. The consequences of human induced global warming and the respective changes that it drives in various environments are expected to be quite dramatic for life on Earth. While small compared to the environmental changes that Earth has experienced in its distant past, they will nonetheless very likely put the very essence of human civilizations at risk. Some of us already experience these consequences firsthand, as for example Pacific island homes are swallowed up by the rising ocean. However, we do need to keep in perspective how minuscule these changes are that may now endanger our civilizations future when compared to the much bigger challenges all life evolving since the end of the Precambrian had to cope with.

We have already seen in the last chapter how much global temperatures and carbon dioxide levels varied over the last 541 million years, sometimes being much higher and then much lower than today's levels (see fig. 2.11 and fig. 2.12). Since the first cratons formed on Earth's surface billions of years ago to provide the nuclei for today's continents, Earth's continental map has also dramatically changed. As the continents accreted they would at times all join together to form single landmasses, the supercontinents of Earth's geological past, only to drift apart again. This continental dance continues to this day and eventually a new supercontinent is expected to form. The challenge for us is that our lifespans are so short that it is tempting to perceive Earth as static, making us believe that it will remain the same forever because during the time human civilizations evolved it has changed so little. This is why we can today relate to the writings of ancient travelers when they describe the landscapes, mountains, rivers, and lakes they have seen in their times. Very little has changed on Earth since humans emerged from pre-history with the first written records dating to some five thousand years ago. If humanity should still be around in a few hundred million years, a feat other species have accomplished in the past, nothing in such an ancient travel guide would make sense anymore because everything will be completely different. Mountains will have eroded away, rivers and lakes dried up and new ones formed, continents will have moved, our world's geography, its seas and landscapes, will resemble nothing like what surrounds us today.

The human and chimpanzee lines split only a few million years ago from a common ancestor and only about 200,000 to 300,000 years have passed since evolution witnessed the first modern humans entering the stage. Modern humans and their close cousins like the Neanderthals did experience significant environmental changes forcing them to adapt; the ice ages come to mind. However, humans only left pre-history after the beginning of the Holocene epoch, which brought more favorable and stable environmental conditions that may have been instrumental in the formation of the first human civilizations a few thousand years ago. Figuratively speaking, the evolution of human civilizations took place during a balmy spring day at the shores of a calm lake. For hundreds of millions of years before that, life had to stir through the tempests that challenge its evolution on geological time scales; humans have not experienced anything even remotely close to that yet. Just like the oxygen level changed significantly since the Precambrian, so did sea levels.

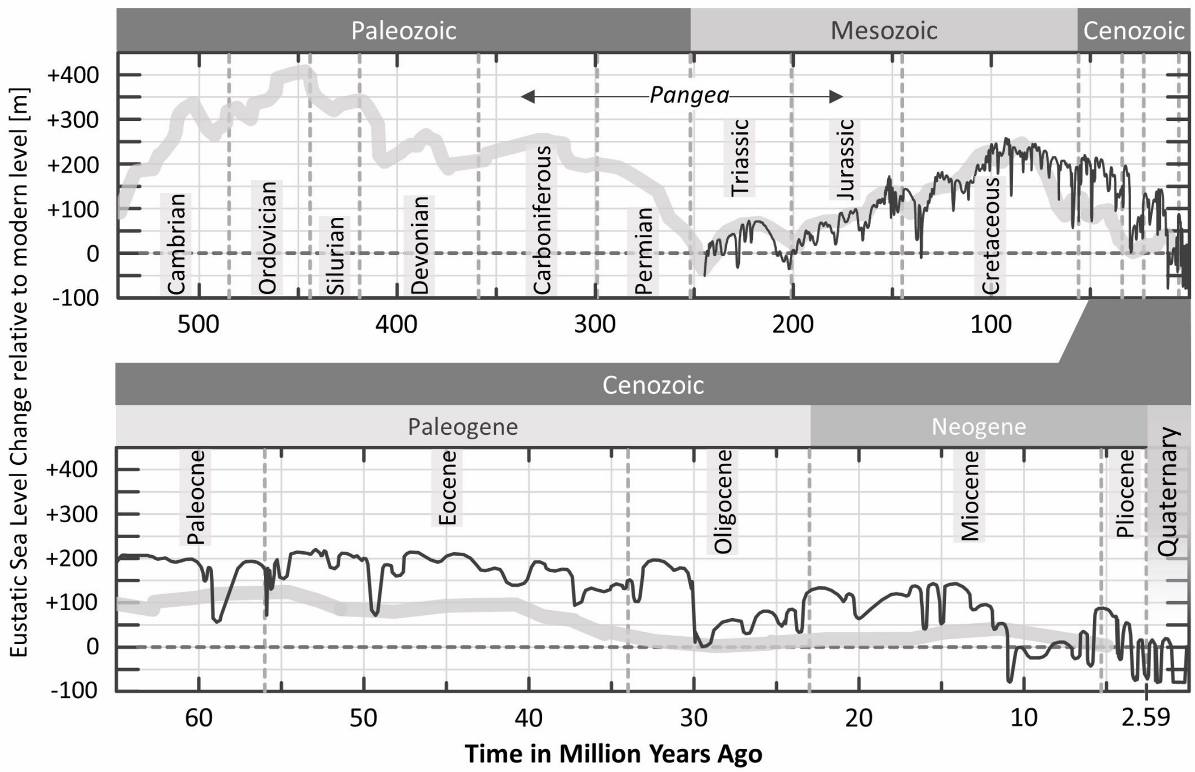


Figure 3.9: Sea level changes over the past 541 million years. The lower part of the figure breaks out the data for the last 65 million years. The two sea level data curves are according to Refs. [43,44].

And changes in sea levels can be as critical in shaping the evolution of life as changes in oxygen level. At the end of the Precambrian, all life was still marine based with no plant or animal life on land; the only life that there may have been on land was likely bacterial life, tied to the water's edge or to the rims of hot springs. Some of the latter may not have looked too different from the bubbling mud ponds, hot springs, and geysers that we can see today for example in Yellowstone National Park. Shallow oceans were the nurseries for the evolution of higher life as they often expanded into so-called marine transgressions, forming large inland seas only to retreat again at some later time. This ebb and flow of shallow seas provided unique conditions for life to experiment and take a foothold. Fig. 3.9 and fig. 3.10 show the eustatic sea level changes as compared to the modern-day sea level over the past 541 million years and the past one million years, respectively.[49] Eustatic sea level changes describe global variations in the masses or volumes of the oceans and are different from relative local sea level changes caused by lands rising or sinking. The main factors behind such eustatic sea level variations are changes in ice sheet volumes, global temperatures, or plate tectonics driven changes to the shape or geographic extent of the ocean basins. When ice sheets melt, their water run-off adds to the oceans and increases sea levels; when ice sheets grow, they remove water from the oceans. The latter happens by ice sheets retaining the humidity transported to them from the evaporation of ocean surfaces in the form of snow precipitation.

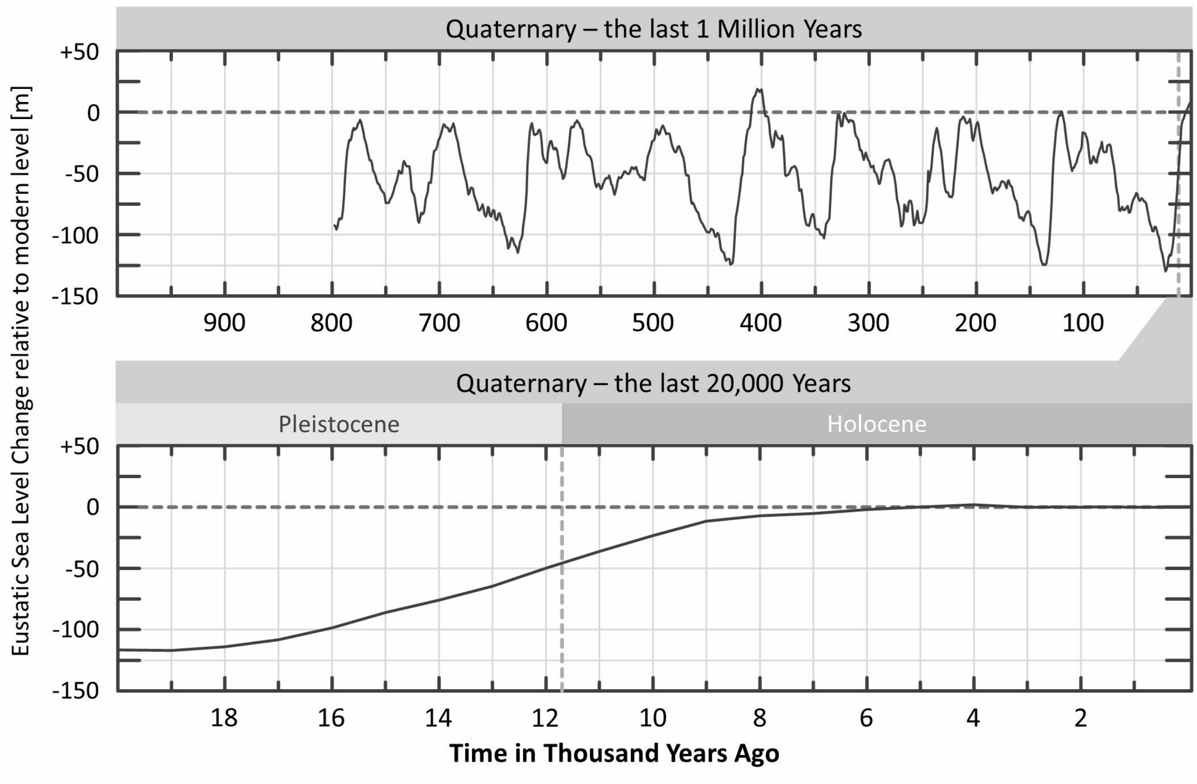


Figure 3.10: Sea level changes over the past one million years. The lower part of the figure breaks out the data for the last 20,000 years. The data curves are according to Refs. [45,46].

As snowfalls continue, the increasing pressure of the accumulating snow converts buried layers of snow into ice, thereby slowly building up an ice sheet. The density of water varies with temperature and therefore greenhouse environments will cause an increase of ocean volumes while a reduction in greenhouse gases will cause ocean volumes to decrease. In the previous chapter, we have seen melt and growth of ice sheets throughout Earth history mirrored by respectively higher or lower relative amounts of heavy O-18 oxygen and light O-16 oxygen in the oceans. The records of this isotope ratio can be obtained through the analysis of foraminifera shells, provided they do exist for the time period of interest, which in many cases they do. As for the impact of greenhouse gases on ocean volumes, scientists can reconstruct it through the available historical carbon dioxide records. Therefore, it is not surprising that the eustatic sea level changes shown in fig. 3.9 and fig. 3.10 resemble to a good extent the carbon dioxide records from fig. 2.11 and fig. 2.12. This is particularly evident when comparing the data for the last one million years for carbon dioxide levels to the respective eustatic sea level changes. Plate tectonic driven changes impact eustatic sea level on a very long time scale. One such example is the formation of the Himalayan Mountains and the Tibetean Plateau due to the subcontinental plate containing India diving under Asia since it started colliding with Asia about 50 million years ago. This mountain-building process has given us an eustatic sea level roughly 70 meters lower than what it would be without it. Supercontinents also

contribute to sea level changes as sea levels fall when they form and rise again once they start to break apart. Fig. 3.9 indicates this cycle for the supercontinent Pangea, from its formation about 330 million years ago to the beginning of its break-up around 175 million years ago. In addition to those long-term patterns in sea level change driven by plate tectonics, Milanković cycles govern short-term sea level variations. During the glaciations periods of Milanković cycles sea levels decrease and during the interglacial periods, as the ice sheets melt, sea levels increase.
We are currently still living in an ice age, albeit in the warm phase of it, a so-called interglacial period. This interglacial period, starting with the end of the last glaciation period in the late Pleistocene is an unusually warm one and it provided the environment for humans to emerge from pre-history. The start of this warm phase of our current ice age is reflected by the steady sea level increase, the onset of which can be seen in the lower part of fig. 3.10 around twenty-thousand years ago. This steady rise in sea levels started to slow a few thousand years ago and since then, sea levels have changed little. However, with human induced global warming and climate change introduced over the past couple of hundred years and very much accelerating since a few decades, sea levels have started to rise again. How much sea levels will ultimately rise over the next one hundred years or so, is still to a good extent under our control. However, if we do nothing or even worse, if we increase the amount of greenhouse gases in the atmosphere while we continue to level rain forests, sea levels will rise a lot more. Within a few generations, they may rise to levels not seen since five million years ago; or put differently, the last time they were at such levels was at a time when it is believed that the human and chimpanzee lines just had split.
As can be seen in fig. 3.9 and fig. 3.10, over the past 541 million years eustatic sea levels have been much higher, hundreds of meters higher; and also quite a bit lower, more than one hundred meters lower than they are today. Because of climate change, eustatic sea levels are expected to be ∼0.5 to ∼1.2 meters higher in 2100 than they were in 2000; depending on how much we can reduce greenhouse emissions from now on. If we are successful, we may not see much more than one meter of sea level rise. However, if we continue business as usual, all bets are off and the longer-term outlook beyond 2100 will become more dire. If the Greenland ice sheet were to melt completely, eustatic sea levels would rise ∼5-7 meters; melting of the West Antarctic ice sheet would add another ∼5 meters; and if the East Antarctic ice sheet should melt, a further 60-meter increase can be expected. With the water from both polar ice sheets added to the oceans, eustatic sea levels will rise by roughly seventy meters.
On the scale of the eustatic sea level changes that happened over the past 541 million years, a 70-meter rise in sea level is still comparatively small. Life on Earth will certainly change but it will survive as it did survive a number of even more dramatic sea level changes in Earth's geological past. We can count on life to persevere, maybe with a different cast of characters replacing the current actors on the stage of life, just as it has happened several times in the past; but survive it will, of that we can be quite certain. However, what we cannot be certain off is how our human civilizations will fare through such a challenge. Will our societies be able to resolve conflicts resulting from rising sea levels peacefully? Given our ongoing difficulties to manage what are comparatively speaking minor human migrations, how will we cope with much bigger such challenges?

To master those, humanity will need true leadership representing the broad interest of all its peoples as a whole but such leadership is nowhere in sight. Unfortunately, the short-term greedy interests of the few are still much more powerfully represented in shaping the path of nations and their approach to climate change than the needs of the billions and their descendants who stand to lose most. For now, we need to shelf this topic, but we will have to revisit it later in a broader context.

As was pointed out earlier in the previous chapter, the Precambrian received its name because for a long time, actually until quite recently, there had been no evidence for life before the Cambrian period. Therefore, the Precambrian was thought to be completely devoid of life. Now we know that this is not correct. A few pages earlier, we have just gone through some of the early records of life left by ancient archaea and bacteria. As we found there, the evidence we have, certainly traces life back to the early Archean eon. Maybe one day we find that the records of life go even further back to some 4,300 million years ago into the Hadean eon. Later, eukarya joined archaea and bacteria, first as single-celled and later as multicellular organisms; and later still, larger single-celled organisms, the protists, appeared. Only long after that, we eventually find the first evidence for plant and animal life.

In Darwin's time and up until quite recently, life seemed to first appear in the Cambrian period and we knew nothing about the ancient history of protists. Darwin readily acknowledged that he had no explanation as to how the rich life that showed up at the start of the Cambrian period had evolved; in his time there was simply no fossil trail to trace life back further. The dearth of fossil records for earlier times remained the greatest challenge for paleontologists for a long time after Darwin.

Fossil records are almost exclusively the preserved skeletal features of lifeforms, in rare cases soft materials such as plants may be preserved under special conditions. The soft parts of animals usually are not preserved and for animals that have no skeletons to start with, chances are slim that we ever find a trace of them. So, our record of life based on fossil evidence is highly biased. The first glimpse of plant and animal life in the Precambrian came from a fossil find made in 1957. This fossil, entering the record as *Charnia masoni*, was found in the Charnwood Forest of the English Midlands, hence its name.[50] It had already been discovered a year earlier by a teenage student, who reported it to her teachers. However, the teachers refused to believe her because according to orthodoxy there could be no fossils in rocks from the Precambrian; so another year went by until the examination of this fossil by someone more open-minded.

Because of its fern-like appearance, it is believed that *Charnia* was not unlike today's sea pens. Just like them, it seems to have spent its life anchored to the seafloor, feeding on nutrients drifting in the water. *Charnia masoni* was actually not the first discovery of Ediacaran fossils, as they came to be known. The Ediacaran received its name from the Ediacaran Hills that are part of the Flinders Ranges of South Australia. It was there, where the first Ediacaran fossils had been found in 1947. However, since the paleontological orthodoxy at the time was that no fossils could be found in rocks older than the Cambrian period they were not recognized as Ediacaran fossils for a number of years. As many more such pre-Cambrian fossils turned up, they revealed a fascinating menagerie of pre-Cambrian life for the Ediacaran, the last period of the Precambrian. Ediacaran life has puzzled scientists ever since, because it is so different from what came after it.

This late Precambrian world is sometimes referred to as the Garden of Ediacara, as it seems, a sentiment based on the assumption that life in the Ediacaran was different from animal life today, notably without much predation. Reinforcing this view are a number of observations: Ediacaran lifeforms seem to lack protection from predation such as shells; many of them do not have any feature potentially resembling a mouth; and they seem not to have moved around in any hurry. In fact, for lifeforms such as the disk-shaped flat *Dickinsonia* genus, just to take one example, we do have no idea how it actually fed. The best guess currently seems to be that *Dickinsonia* hoovered over microbial mats until those started to rot and they could somehow absorb them. A leisurely way to feed indeed, that could only be pursued if such an animal, practically representing a served dinner plate on the sea floor, was not predated on itself. Hence one view, the assumptions about feeding, reaffirms the other ones, no predation and no fast movement and therefore no legs. Sounds all quite logical but in a sense, it also sounds all too convenient. We have to assume that our views are highly biased by a likely very much skewed fossil record. Or could it be true that the Garden of Ediacara was a vegetarian's paradise?

Can we seriously believe that symbiotic relationships dominated life on Earth from when it first emerged until the Ediacaran period at the end of the Precambrian? That before the Cambrian period predation was not a factor driving evolution until at some time suddenly it became one, in what is practically a blink of an eye on evolutionary or geological time scales? That is likely a stretch and things in life are never this black and white. For a long time, so it seems, archaea and bacteria just ran their little sulfur- or methane-based metabolic engines which at some later point in time were joined by photosynthesis processes, one of them generating oxygen. But what about eukaryotes?

Thanks to Lynn Margulis, we are quite sure that eukaryote cells containing a nucleus and all the little organelles that drive and energize all cell functions, as they do in the various cells in our bodies, are the result of endosymbiosis. To reiterate in a somewhat cartoon-like description: at some point in time eukaryote cells originated by one bacteria swallowing another one, but instead of digesting the swallowed bacteria, the host allowed its prey to become a tenant and evolve into an organelle. In fact, all the organelles in eukaryote cells are thought to have this very same kind of origin; and as discussed above, that may even be the case for the cell nucleus. However, complicating the picture, we more recently learned that prokaryotes also seem to have organelle-like structures, albeit much smaller in size than in eukaryotes.

The concept of endosymbiosis is one of those paradigm-shifts that radically altered our perception of how life evolved. Before this discovery, we were looking at a tree of life concept where lifeforms evolve from one another through linear descent in a given branch via selection on genetic mutations happening at an astonishingly constant rate. Now we see lifeforms acquiring genetic materials from other lifeforms located on a different branch of what we previously envisioned as the tree of life and make it their own. Instead of the simple vertical lineages we thought trees of life should reflect, scientists are getting now a much more nuanced perspective of evolution. The picture that emerges is more like a bush of life where at least in its early phases lateral gene transfer may have been as important as vertical gene transfer. And in some cases, without lateral gene transfer, life as we know it would not exist unless nature would have found another way to evolve the equivalent of eukaryote cells without endosymbiosis.

The tree of life picture that we have been familiar with since the time of Darwin is problematic in other ways too. We know that the vast majority of species that ever lived on planet Earth is extinct today. Many of them may have belonged to branches of the tree of life that eventually turned out to be evolutionary dead ends. Life's history on Earth must have included a good number of such evolutionary dead ends, but just how many, we will likely never know. But instead of a tree branching into the sky with a lush crown we may have to consider that the proper picture for life is a pruned tree or better bush where there are fewer twigs in the crown than somewhere further down; much like the pruned tree cones one tends to find in Europe's Baroque or Rococo gardens. In fairness, our knowledge of life, while vastly increased as compared to only a few generations ago, is in all likelihood still very incomplete.

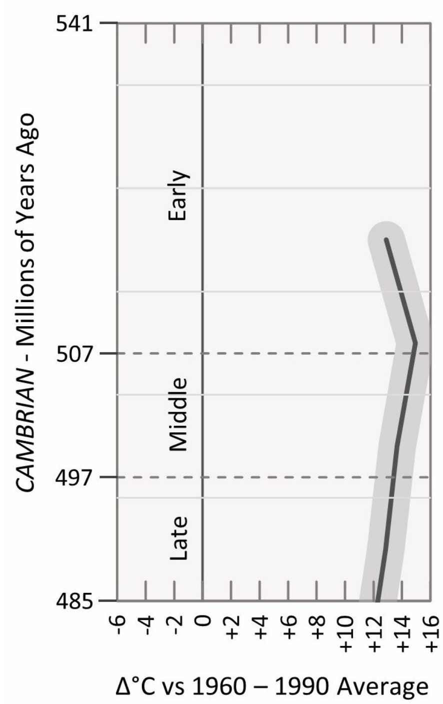
To answer the question of when predation really started, we need to be more specific about what we mean by predation. Most would agree that a characteristic of animals is the fact that, unlike the majority of plants, they do not generate their own food from the energy of light and nutrients in the soil. Instead, many animals live off plants by grazing, and others live off the grazers with some of them doing occasionally both. Staying in this picture, the question we are really asking is: when did animals first "discover" that by eating other animals they could gain much more energy per weight measure from animal protein than from plant matter? That tends to turn our attention to the arrival of animals having sensory organs and the equipment required to become carnivores. That may include eyes to pursue, teeth to kill, bite or chew, and claws, or as is often the case with arthropods, many legs that can shred smaller prey into tiny morsels ready for the intake. The other side of this would be to look for evolutionary developments to counter such predation, leading to animals that have sensory organs to see their predators, shells to protect them, or other means to escape, defend, or hide and evade being eaten in any possible way. Inevitably, that indeed does lead us to the conclusion that up until the end of the Ediacaran, Earth may have been a world without predation. However, this picture is likely too simplistic as some of the assumption leading to it are questionable. First, as we follow the animal and plant branches to their common ancestors in the bush of life it becomes increasingly difficult to determine which is which. While that is no surprise, it messes up our simple picture.

Few will likely find it difficult to distinguish plants from animals; unless they are shown one of the more challenging specimen. However, that was not always so, as there have been organisms that in many ways had characteristics of both, plants and animals, but they are extinct today. The further we go back in time, the more difficult it can be to differentiate what is a plant from what is an animal. One of the more solid supports we have is that plants must have preceded animal as animals live off plants. However, even that may be questionable. Not all plants are just living off the energy of the Sun and the nutrients in the soil. A number of them actually are predators digesting anything from small insects to the eventual unlucky small rodents that cannot get out of the treacherous traps those plants set for them and in which these plants slowly digest their prey; sometimes assisted by animals in a symbiotic relationship. Life seemingly excels in using the tiniest of advantages to survive and spread itself across Earth. Should that not make us wonder why it would have taken so long for life to find out that animal protein is much more nourishing than plant matter? That it should have taken more than 1,000 million years between the arrival of the first protists and the end of the Precambrian for life to discover the nutritious advantages of animal proteins seems unlikely. We do not know when some bacteria started feasting on others and by some chance event nature discovered that is was more advantageous to engage in endosymbiosis than to digest fellow bacteria as food. It just does not seem probable that the very first encounter where one bacterium swallowed another immediately led to endosymbiosis. Early single-celled lifeforms did not have claws or teeth, but they were certainly capable of swallowing each other if the opportunity arose. It seems to be much more plausible that bacterial predation was part and parcel of life already early on and that uncountable numbers of bacteria were digested before endosymbiosis eventually kicked in. After all, hundreds of millions of years did pass between the appearance of the first bacteria and the arrival of the first eukarya. The evidence that paleontologists have discovered so far shows that there is roughly a 1,000 million year time span between those two events making it almost a certainty that predation among some bacteria must have been part of early life. For all we know, Earth is still home to a good number of predatory bacterial species today.

Predation as a driver of natural selection most likely did not arise all of a sudden in the Ediacaran period, it must have been around for much longer. However, the quality of predation changed over time. The acritarch fossils dating to some 1,800 years ago may indicate the arrival of a different kind of predation as some such organisms bear marks of micro-predation. Many acritarchs were protists and they could have been among the first to gobble on their fellow animals. Another such turning point may have been the later Ediacaran when animals became relatively speaking larger while still remaining quite small as compared to the animals that we start to see with the arrival of the Cambrian. Only with the arrival of large animals did the quality of predation eventually begin to reflect the famous dictum of a “”Nature, red in tooth and claw as the British poet Alfred, Lord Tennyson (1809 – 1892) put it in one of his verses.

Ediacaran life evolved body plans with bilateral and radial symmetries. Some of the bilateral body plans came with head or tail ends, but many had none. Often, Ediacaran fossils reveal a distinct left-right asymmetry in their bilateral bodies. Seemingly, the oceans were full of all kinds of jellyfish predating on planktonic matter as it drifted through the waters. Most other life seems not to have moved around much, many species not changing places at all but sitting anchored on the sea floor while filtering out tiny food morsels wafting by with the ocean currents. The boundary between the Ediacaran and Cambrian marks a major transition for life on Earth. For reasons we do not understand yet, much of the Ediacaran life seems not to survive into the Cambrian. It is as if a page had been turned when the history of life before the Cambrian was left behind in the Precambrian and new lifeforms began to burst onto the stage of life with the start of the Cambrian period. While this is what it looks like, paleontologists have become careful in marking such clear boundaries for life. Frequently, it is more the lack of knowledge and the limited information available to us, in this case about Ediacaran life, that shapes our view and reality may have been quite different. We can only hope that with new fossil finds new insights will become available. This has happened many times before and every time this has helped us to gain a better understanding of what happened as Earth transitioned from the Ediacaran to the Cambrian period.

The Cambrian: 541 - 485 Million Years Ago



Cambrian: Global temperature vs modern average.

The supercontinent Pannotia, which had formed in the early Ediacaran period, started to break up towards the end of the Ediacaran. This process of fragmentation continued into the Cambrian but much of the landmass remained assembled in three composite continents or continental cratons, which stayed close to the equator. The supercontinent Gondwana was the largest, including what was then South America, Africa, Antarctica, Australia, parts of South Asia, and a few other small pieces. The continental craton Laurentia was the second largest, including what would become North America, parts of Europe, and Northern Asia; and the third one was the Baltica continental craton which included parts of Northern Europe and Asia. As the reconstruction of past climates shows, large landmasses concentrated around the equator tend to produce warm and dry climates. Global temperatures throughout the Cambrian seem to have been significantly higher than the modern thirty-year average. Not only were the continental arrangements quite different from today but so was the continental topography. The topography that we see today had not yet been built. Much of the folding and uplifting that would be a consequence of plate tectonics over the next few hundred million years was still to come. There were no Alps or Rocky Mountains, nor where there Andes or Himalayas. Earth's lands were still rather flat with little in the way of topography that could stop rising seas from inundating large continental areas. Changes in sea levels back then had a vastly more dramatic effect than today as already slight increases in seas levels could create enormous inland seas. As we picture the evolution of life in the Cambrian, the canvas that we should assume as a backdrop is a combination of shallow seas and low lying continental areas representing a much larger ecosystem than it is today. In a way this was fortuitous because theses shallow seas were exactly where the explosion of life in the Cambrian could take place.

With the beginning of the Cambrian period, we see a proliferation of what paleontologists refer to as small shelly fossils. None of the soft bodies that inhabited those shells have been preserved so we can only guess as to what they must have looked like. However, going just by the shells, they seem to have been different from the shelly animals that came later. By the Cambrian, animals had learned how to make skeletons and shells, and it looks like that happened on an evolutionary time scale within the blink of an eye; and not just one type or kind of shell. Shells all of a sudden come in many different forms and all of them small. These shells were made out of calcium carbonate, the material that animals mostly used back then and marine animals still use today to construct their shells. However, in the Cambrian they also used calcium phosphate and much more so than today. This likely was because of a higher phosphate content in the late Precambrian and early Cambrian oceans, something that is well known from rich phosphate deposits left on the floors of what back then where shallow seas.

For paleontologists the Cambrian is a period of many firsts in the fossil record, shells just being one of them. Literally, the Cambrian gives us the very first footprints of life. Trace records left behind by tiny creatures moving across the sand or boring into it, leaving their footprints as the earliest fossil evidence we have for the presence of animal legs. There are the first eyes such as those unique faceted eyes of the trilobites, the most abundant animal species we know of from this period. Trilobites are the animal species that most of us would associate with the Cambrian because they have been so prevalent and we have many well-preserved examples of them. Maybe we perceive them as the most abundant species of the Cambrian because their skeletal features preserve so well and because we cannot find fossils of the many other species that did not fossilize as well or not at all. We will likely never know. As the name indicates, their bodies had three lobes, two side lobes and a larger central lobe to which head and tail parts attached. They had eyes and they had many legs as any self-respecting arthropod could be expected to muster. Not every Cambrian fossil body plan was as straightforward as that of trilobites.

A number of Cambrian fossils have caused quite some head scratching among paleontologists. It was the peculiarity of the early Cambrian fossils that gave rise to the earlier discussed debate as to what would likely happen if one could replay evolution. What triggered this whole debate was the fact that the early Cambrian fossils did not really seem to indicate any continuity with most lifeforms that would emerge later in the Cambrian, expect for a few species such as jellyfish. This contrast of lifeforms populating the early Cambrian or late Ediacaran against what populated the later Cambrian seemingly came into even starker relief with the discovery that the soft bodies of some Cambrian animals had been uniquely preserved through the peculiarity of fossilization in the layers of the Burgess Shale in British Columbia, Canada. It looked as if evolution had run many failed experiments with earlier Cambrian lifeforms as almost nothing seemed to connect to later Cambrian lifeforms. If that was the case then clearly, if evolution would be replayed, a single chance event could drastically change the outcome given that evolution seems to have pursued so many different lifeforms of which only a few would continue.

However, after studying the Burgess Shale fossils in detail, it eventually was discovered that most of the early Cambrian lifeforms did have connections to what came after them. One of the more peculiar early Cambrian fossils was given the intriguing name of Hallucigenia; the name seems to indicate paleontologists may have believed they were hallucinating when they first saw this animal's body plan. But once they realized that it was them who had turned the animal on its head, interpreting what were actually the upper parts of this animal as its lower parts and vice versa, things started to make more sense again. While there are still some evolutionary oddities of the early Cambrian that seem to fit in no known animal group, most of them now do. There was even some continuity as life transitioned into the Cambrian with a few Ediacaran species still around in the early Cambrian. There is a critical lesson here. It just takes time and a lot of work to study the evidence preserved in the fossil record before definitive conclusions can be made.

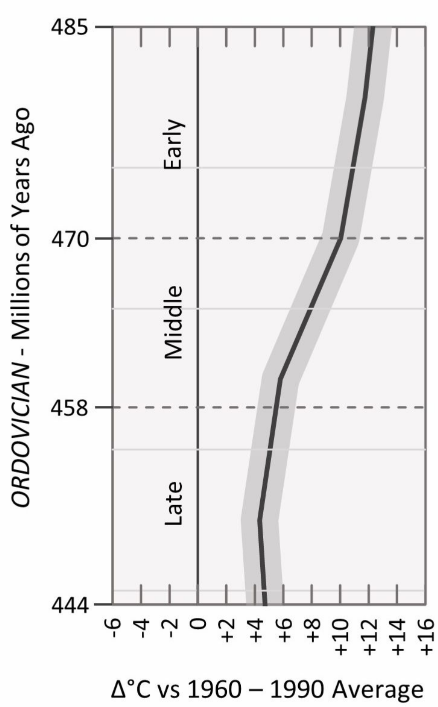
The seeming explosion of life the rich fossil record of the early Cambrian period bears witness to is still very much a mystery. All of a sudden, around 540 million years ago, a diversity of lifeforms seemingly appears in the fossil record practically coming out of nowhere in what in evolutionary and geological terms is a very brief period, only a view million years at best. However, in fairness, evolution can achieve a lot of change in a few million years; we only have to look at the brief history of humans who split from the chimpanzee line not more than some six million years ago; and what a difference these few million years have made. The Cambrian fossil records shows us that most major animal body plans, or phyla as the experts call them in Latin, that continue into our time, make their first appearance in this period. The Cambrian period bestowed on us the large majority of blueprints for animal life that we are familiar with today or that we know from the later fossil records from species that are now extinct.

Clearly, something had to precede the Cambrian explosion, it cannot have come out of nothing and there has to be another explanation. Just as primate evolution preceded human evolution the animals of the Cambrian must have a lineage going back in time before the Cambrian. One hypothesis that seems to make the most sense given the available data is that the ancestral lifeforms of what we see emerge with the Cambrian period may have just been very tiny. So tiny that we do not find a trace anymore when going back in time more than 540 million years. It could have been that a higher level of available oxygen triggered the evolution of larger lifeforms, large enough to pass the threshold needed for evidence of their presence to preserve for hundreds of millions of years. The fact that animals had learned how to produce shells and skeletons certainly also made it much more likely that we would find their fossil records today; the same cannot be said for life in the Precambrian. Another hypothesis to consider is that a change in sea levels may be responsible for us not finding the same abundance of life before the Cambrian. All land at the time was still devoid of life and therefore none of the rock formation on land would have carried evidence of life. We also know that life in the Cambrian, or any life before it, flourished in shallow waters. Putting these two facts together, the hypothesis goes like this. For whatever unknown reasons the oceans expanded at the beginning of the Cambrian and new land came under water not previously covered by oceans before. Hence, the result is that we find a lot of fossilized life in these virgin lagoons preserved in Cambrian sedimentary rocks, but virtually nothing before it. It is an intriguing argument and can arguably explain the sudden jump in the richness of the fossil record we see with the Cambrian. However, so far it is a hypothesis and not a fact. Sometimes we have to accept that there are no satisfactory explanations because we have too little data to substantiate one. Scientists continue the search for more clues in the records of life and our methods and capabilities to do so have and will continue to improve. Eventually, their hard work and patience is likely to lead us to increasingly better explanations, as it has done many times in the past.

The Cambrian was punctuated by extinction events, including one in the middle of the period and one towards its end. Decreasing ocean oxygen levels are thought to have caused the Middle-Cambrian extinction. This is also believed to have triggered the radiation of new species, which the fossil record shows after this event. Life in the Cambrian sea flourished in shallow coastal waters populated by creatures that we would not readily recognize but the ecological niches they occupied would be familiar ones.

Arthropod species were the most abundant among animals and trilobites were their most prominent Cambrian representatives; that is what the fossil records tells us, how complete this picture is we do not know. Other invertebrates included mollusks, worms, sponges, and animals with radial body symmetry such as jellyfish which were already present in the Ediacaran and are likely one of the oldest larger lifeforms. Cambrian fish were jawless but some of them had teeth. Such as the conodonts who seem to have appeared in the later part of the Cambrian and are aptly named for the cone shaped teeth which are among the most abundant microfossils of the Cambrian.[51] For a long time conodonts were only known from their teeth. It remained a mystery as to which animals such teeth coming in various shapes and sizes actually belonged. That is until first a small complete specimen was discovered followed later by a larger example the size of a small fish. They revealed a body form that looked much like an eel with a pair of eyes on its front end, fins, and a flexible cartilage cord down its back which makes it a vertebrate; one of the first ones. At the base of the Cambrian food chain remained single-celled algae. As we look at the evolution of complex animal and plant life, we should never forget that their existence, even though several stages removed, was completely dependent on the diversity, multitude, and plenty of algae and plankton matter in the oceans.

The Ordovician: 485 - 444 Million Years Ago



Ordovician: Global temperature vs. modern average.

During the Cambrian, the composite continents that made up Gondwana-Land had slowly rotated away from the equator towards the South Pole and by the end of the Ordovician period, parts of what is today Africa would be located at the South Pole (see fig. 2.4b). Global temperatures remained significantly above the modern average throughout the Ordovician period. However, during the early and middle Ordovician, they declined from the high levels seen in the Cambrian to settle around a few degrees above the modern average in the late Ordovician. During the late Cambrian and into the Ordovician period, plate tectonics resulted in the closure of a part of the southern ocean, referred to as the Iapetus Ocean.[52] As this happened, the continental landmasses of what were then Laurentia, Baltica, and Avalonia closed in on each other and in the subsequent collisions new mountains were rising. Avalonia, as its name implies, was a small composite continent that included parts of southwest Britain, Western Europe, and the east coast of North America. It had broken off from Gondwana earlier and on its journey west and then to the north it started colliding with Laurentia and Baltica which were approaching each other at the equator; Baltica moving in there from the northeast and Laurentia moving there from the west. The mountain building, or to use the technical term, the orogeny that resulted from the subsequent collisions is called the Caledonian orogeny, Caledonia being the Latin name for Scotland.[53] This mountain forming process took place over roughly one hundred million years, from the early Ordovician, through the Silurian, into the early Devonian period. As the naming suggests, traces of the mountain ranges formed during this Caledonian orogeny can still be seen today in Scotland and parts of Scandinavia after almost 400 million years of weathering and erosion. Similar to the Cambrian period, wide spread shallow seas and large inland seas, combined with an overall warm climate for most of the Ordovician, continued to provide favorable conditions for marine life.

In the preceding Cambrian period, many of the blueprints of life had made their first appearance even though most of their specific representatives would not survive. The Ordovician period played a similar role for the emergence of the ecological system blueprints that are still the ones we recognize today. In the Ordovician, these ecological systems would be populated by different lifeforms than today, some of them looking more familiar than others do; but they would play roles we would perceive as comparable to those that modern lifeforms play in such systems today. Of course, this only applies to marine based ecological systems, as there was still no life on land, coral reefs being a good example for such marine ecosystems. The marine food chain also would start to resemble what our modern oceans look like today except for fish, which were still quite small and still jaw-less.

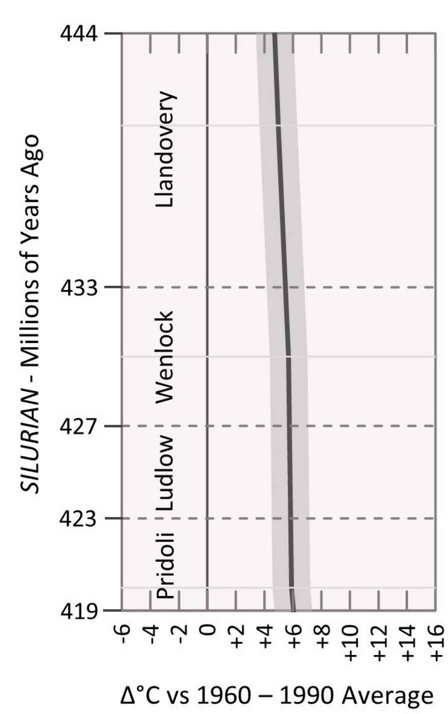
The dominant lifeforms in the Ordovician oceans were invertebrates such as mollusks and arthropods; at least that is what we can read from the fossil record. The fossil record can sometimes bias our views because of what it preserves and what it does not. Acknowledging that, it does look like that there was quite a bit more predation going on in the Ordovician than the Cambrian. The thicker shells we find in the Ordovician are a good indicator for that. The most fearsome predators at the time seem to have been the many different kinds of nautiluses, and mollusks. Many of the nautiluses being of the linear version rather than the coiled-up ones whose smaller sized descendants are still around today. Nautiluses were growing quite large back then. Some of the fossilized nautilus shells approach the size of large truck wheels and fossilized remains of straight nautiluses of up to six feet length have been found as well. Other no less ferocious and nightmarish predators at the time included large sea-scorpions, arthropods the size of a man which are extinct today; and there were of course smaller predators including starfish, sea urchins, or shrimp-like arthropods who had the tool sets to devour even smaller lifeforms.

Then, of course, there were creatures such as the conodonts, a species we already encountered in the Cambrian. During their existence, conodont species seemingly populated every marine habitat on Earth in large numbers and in many changing varieties. Because they were so prevalent, rocks of different ages from the late Cambrian up to the Triassic include many different conodont micro-fossils. This makes conodonts so-called index fossils, natural chronometers, which scientists use to accurately date rocks starting from the late Cambrian up to the Triassic-Jurassic extinction event some 200 million years ago, when conodonts finally disappear from the fossil record. Another kind of index fossils for the Ordovician are the so-called graptolites, tiny floating colonial plankton feeders that were around from the late Cambrian until the early Carboniferous epoch. They have been very helpful in understanding the stratigraphy of rocks from the Cambrian up to the Devonian and were particularly useful in establishing the boundaries of the Ordovician period itself.[54]

Sometime in the Ordovician, the first plants must have started their landward journey as the oldest fossilized airborne plant spores have been found in late Ordovician rocks. Land plants evolved from algae and the first ones to venture on land may not have gone far and stayed close to the ground, likely plants similar to the liverworts and mosses of today. These are pioneer plants which to this day will be the first ones to colonize new territory. Botanists call them non-vascular plants, and sometimes they refer to them as primitive plants, because they lack the vascular tissue that allows modern plants to grow tall. A plant's vascular tissue transports water and nutrients sucked from the soil in which it is rooted as well as sugars produced in its leaves to where the plant needs them for growth. The lack of a vascular tissue restricts the size of plants and keeps them to the ground. The first plants derived from algae, which in their aquatic environment were practically swimming in nutrients, allowing their cells to absorb directly whatever they required without the need for transport. Because of that, these first plants needed to have water constantly available, they needed to stay in close contact with water and this is why they would have stayed at the water's edge.

The Ordovician ended with a cold period referred to as the Andean-Saharan glaciation or also the Ordovician ice age lasting from 460 – 440 million years ago. Fig. 2.4b shows what the continental map of the late Ordovician looked like, with Africa being part of the Gondwana supercontinent and practically budging up at the South Pole. In the previous chapter, we saw how changes in Earth albedo moderate or amplify the impact of the solar forcing pattern on Earth climate. The concentration of much of Earth landmass close to one of the poles - in the Ordovician at the South Pole - creates the conditions for just such an albedo altering effect. Therefore, we should not be surprised to find evidence for an ice age in the late Ordovician in what is today's African Sahara. It is this ice age, which scientists believe triggered the Ordovician-Silurian mass extinction event. This mass extinction event marks the boundary between the Ordovician and Silurian periods; and it also marks one of the most significant reshuffling of life on Earth with many species becoming extinct. The lifeforms surviving this event or emerging from it, as it was more like a protracted extinction period, would provide the new stock, the new cast of characters, from which future life was going to evolve, including ourselves. As animal life was still bound to the oceans, marine organisms practically comprised all complex multicellular life at the time. Therefore, the Ordovician-Silurian extinction event was an exclusively marine mass extinction with a large number of marine families vanishing forever; estimates range between 60-70% of all marine species at the time being wiped out. If anything had played out differently during that event, leaving behind a slightly altered cast of life's characters, the world could be quite dissimilar from ours today and maybe other creatures than us would now be wondering about what happened back then. The evolution of life throughout Earth history is in many respects a lottery, on the very large scale as well as on the very small one. Over the enormous temporal expanses of geological time, it produced winners and losers among species in the past and will continue to do so in the future. On a very much shorter time scale, it results in some of us luckily born into prosperous and nurturing societies while the unlucky ones are born into poverty and war.

The Silurian: 444 - 419 Million Years Ago

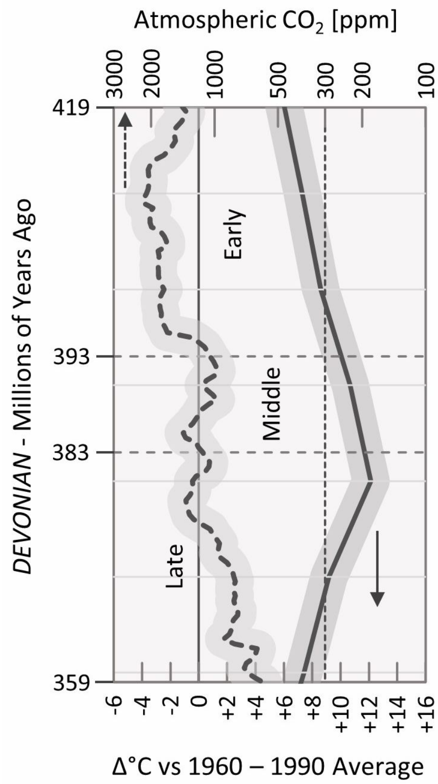


Silurian: Global temperature vs. modern average.

During the Silurian, the Gondwana supercontinent continued its slow drift towards the South Pole but large landmasses remained close to the equator where the climate staid warm and tropical. In the west, the composite landmasses of Laurentia, Baltica, and Avalonia, converging around the equator, were joined by Siberia from the northeast to form what would become in the Devonian the composite landmass referred to as Euramerica or Laurussia.[55] The early Silurian climate remained similar to the climate that prevailed throughout the late Ordovician with glaciers occurring at the higher southern latitudes. However, as the Silurian progressed, Earth entered a long greenhouse phase with an arid and warm climate dominating moderate latitudes. In equatorial regions, large areas of the continental landmasses continued to be covered by warm shallow seas. This was particularly the case in the early Silurian when sea levels were still high following the end of the Andean-Saharan glaciation that had dominated the late Ordovician climate. The Silurian period was a phase of transformation, particularly the late Silurian when plants and animals moved gradually landwards. As plants evolved to survive in drier conditions away from the water's edge, Earth began to green. This process started in the late Silurian with the first green sprinkles on otherwise barren lands and it would take well into the Devonian before we can speak of a *Green Earth*. The development of higher plants is a major milestone for life on Earth. Ferns came first among the vascular plants with roots, leaves, stems, and trunks. With a circulation system that let them grow large, plants rose from the grounds towards the sky. This magical transformation converting the barren rock surfaces of Earth's continents into landscapes covered in lush green took many millions of years and it was a collaborative effort. Parallel to plants, the first animals were also moving onto land. We know this from fossil trace evidence, interpreted as the first preserved footprints we have of animal life on land. Scientists identified them as belonging to arthropods; a small step for an arthropod, a huge step for life. However, there is still some controversy. The oldest trails of animal movements date back to the Cambrian, left by unknown creatures in an ancient sea floor. Interestingly, there are also trace marks indicating some animals may have ventured onto Cambrian shores but never seem to have colonized land. We do not know which animals left those trace marks and why they left the water at all when there were no food sources available on Cambrian lands; seemingly, mating could be one reason. However, it is only in the Silurian that we find the evidence that animals finally moved onto land for good, plant eaters following their food sources and predators following their prey. Both, plants and animals, had to go through major evolutionary changes to colonize land. Plants had to make sure they could hold on to water and not lose it through evaporation while at the same time they needed to maintain their capability to absorb carbon dioxide for photosynthesis. The oldest fossilized plant spores we have date back to the Ordovician. However, dispersion through windblown spores does not ensure that such spores can survive on land, something more is needed. As plants evolved airborne spores that could spread their kind all over the planet, the solutions devised within the kingdom of plants would be the development of waxy skins, vascular systems, roots, leaves and stomata.[56] Animals had to go through their own challenges to conquer land. For plants, the primary challenge was to hold on to water, for animals it was to figure out how to breathe on land. Today's aquatic arthropods have feathery fish-like gills to absorb oxygen, which reasons that this is where arthropods back then also started. To move onto land, arthropods evolved the capability to breathe through their skin using a system of tubes biologists call tracheae.[57] Our own ancestors took a different path as they evolved lungs, and it would take longer for them to develop this capability. For some time, arthropods were on the top of the food chain of animal life on land while continuing to be one of the dominant marine lifeforms. There is justification to look on animal life up to the Silurian as the Age of Arthropods. Arthropods continue to this day to be one of the most resilient, numerous, and most diverse lifeforms. However, never again would they be as dominant as they were back then when they were the vanguard of animal life on land; and never would they be as big in size. The giants among arthropods grew in length to the size of a man, likely testing the very limits of arthropod growth. One factor limiting the size of arthropods is their exoskeleton. Unlike the internal skeletons of mammals and reptiles, exoskeletons cannot support the body sizes that we later would see with the large giant reptiles and mammals. A likely even bigger restriction is the way arthropods take in oxygen directly through the surface of their exoskeleton to diffuse it throughout the body tissue. To a first approximation, the surface area of a body scales with the square of its linear dimensions while its volume scales with the third power of the same. There are ways to increase a body's surface area through intricate folding of tissue but while that works with lungs, it does not work with tracheae. Therefore, with increasing size arthropods will struggle to sustain themselves with a sufficient amount of oxygen. As their prey continues to evolve to become more elusive or if it becomes scarcer because other predators hunt it as well or more efficiently, this poses a serious problem. At some point, such large arthropods must either evolve into very successful ambush predators or will have to downsize to become more mobile again so they can successfully pursue prey and make a living.

When we think about the evolution of life, we usually tend to focus on the larger animals and plants and we often give little attention to the much smaller lifeforms, sometimes barely visible, if at all, and the dirt they live in. That is a mistake. The Silurian is not just the start of a greening Earth and of animal venturing on land. It is the very beginning of fertile Earth itself. Before plants and animals moved onto land there were no soils, the land was literally barren. Over the hundreds of millions of years since then it is the combination of decaying plant matter and the multitudes of small organisms that decompose it, which produce the fertile soils in which we grow our food staples feeding billions of us today.

The Devonian: 419 - 359 Million Years Ago



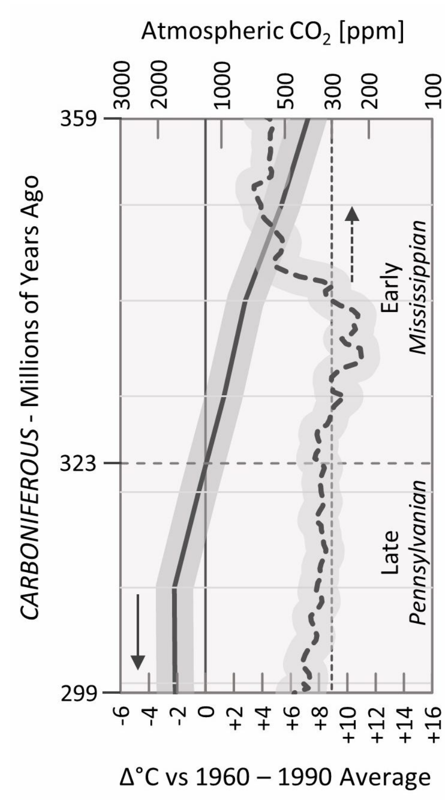
Devonian: Global temperature vs. modern average (solid line) and CO_2 levels (dotted line).

In the Devonian, the bulk of Earth's landmasses remained largely centered around Gondwana, much of which stayed close to the South Pole. The exceptions where Euramerica / Laurussia, Australia, and the South China craton, all straddling the equator; and some continental fragments of Southeast Asia as well as the North China and Siberian cratons lodged north of it. Life may have started to move landwards as early as in the Ordovician. We know with certainty that it did so at an accelerating pace during the Silurian. However, it was only with the Devonian that plant and animal life on land proliferated and diversified, and it is towards the end of this period that we see our distant animal ancestors making it onto land. In the early phase of the Devonian, the pioneering land plants may still have been restricted to staying close to bodies of water or in moist areas. Eventually plants became larger as the evolution of vascular systems progressed and it is in the Devonian when plants figured out how to grow large. The origin of lignin, the organic polymer in the cell walls of many plants providing the structural support that allows trees to grow large, is still being debated.[58] Lignin's original function in plants may have been to enable their vascular systems to transport water efficiently or to provide structural support. We do not know yet which one came first. Regardless, sometime in the Devonian, lignin started to fulfill both functions and trees ever since have been prominent features of Earth's landscapes. It is also in the Devonian that we see plants adding new reproduction strategies as the first seed plants emerge. An important addition to animal life on land were insects for which we find the first fossil evidence in the early Devonian. These insects are still flightless. It is intriguing to see one of the most important relationships in nature, the one between plants and insects, going back all the way to the time Earth started to green. Animal life in the seas continued to diversify and the Devonian is often referred to as the Age of the Fish. The first jawed cartilaginous fish such as shark-like fish and bony fish likely had already appeared in the late Silurian but they are certainly present in the Devonian. Importantly, the bony fish evolved into two groups; into fish that have rays in their fins, or ray-finned fish; and into fish, which have a central lobe in their fins, or lobe-finned fish. This central lobe of lobe-finned fish, as it evolved to contain bones and muscles, could provide the flexibility and strength for fish to move onto land and become amphibians. One such lobe-finned fish species, the coelacanths, still lives in our oceans today, as an endangered species though.[59] This veritable living fossil, thought to have gone extinct long ago, was rediscovered in the late 1930's. The first amphibians, appearing in the late Devonian around 370 million years ago, evolved from similar lobe-finned fish. To conquer land, they needed not only to lift themselves off the ground, but they also needed to develop lungs. Lungfish first developed in the Devonian and several species of lungfish can still be found in different habitats on Earth today. Amphibians evolved from such lobe-finned lungfish whose lobes had bony structures. The first vertebrates lifting themselves onto land to become the ancestor of all four-legged animals must have gone through a similar evolution. The fossil named Tiktaalik is an example for such a transitional animal.[60] Its first fossil specimen was discovered in 2006 and dates back to 375 million years ago. Tiktaalik combined fish-like features such as gills, scales, and fins with features that would become characteristics of four-legged animals such as lungs and a movable neck. How Tiktaalik's bony fin structures fit into the evolutionary path of wrists is still controversial. If Tiktaalik represents the missing link that connects the evolutionary development of different species through being their last shared ancestor is also still open to debate. But what we can say with some certainty is that Tiktaalik seems to be closer to being that missing link than other transitional fossils found before, but it may not be the missing link. Regardless, Tiktaalik has very much expanded our knowledge about what the transition could have looked like which eventually resulted in four-legged animals evolving from fish as undoubtedly it happened. And we can hope that further fossil finds will bring us ever better understanding of how this transition took place. One thing however, we already know. There could have been different path towards the evolution of four-legged animals, some of which may have led to us having more or less than five digits on each of our hands and feet. Evolution seems to have been experimenting with hand designs having different numbers of digits and for reason we do not know, a creature with five digits became the ancestor of all four-legged animals. It may have been just another role of the dice in the evolutionary lottery as we do not know of any reason that would make five digits on our hands preferable as compared to say for example, four or six digits.

During the late part of the Devonian, Earth climate started to cool again. As plants colonized more and more land surfaces, increased weathering of rocks and soils must have resulted in larger amounts of soil minerals washed out into marine environments. There, as the hypothesis goes, it produced widespread algal blooms. Together with the greening of the land, scientists conjecture, this algal bloom led to an increased sequestration of atmospheric carbon dioxide. In addition, the algal bloom also starved the deeper ocean of oxygen and resulted in a build-up of carbon sediments as dead algae sinking to the bottom of the sea would not decay due to the lack of oxygen. Eventually this led to a cooling climate and renewed glaciation at the end of the Devonian. This marks the beginning of the second ice age of the Paleozoic era after the earlier Ordovician-Silurian ice age. Often referred to as the late Paleozoic ice age, it is also known as the Karoo glaciation, or Karoo ice age, named after the Karoo basin in South Africa. Climate cooling and oxygen starvation of the oceans caused extinction events on land and likely much larger mass extinctions in the oceans. In the Devonian, much of animal life was still in the oceans and therefore, the late Devonian mass extinction is mostly a mass extinction of marine life. This late Devonian extinction lasted from 375 to 360 million years ago and is thought to have resulted in the extinction of up to between 80 to 90% of all species.

The Carboniferous: 359 - 299 Million Years Ago

The continental configuration that produced the Karoo ice age would also shape the Carboniferous period. Large ice sheets would continue on the South Pole as the supercontinents of Gondwana and Euramerica / Laurussia moved ever closer together to form the supercontinent Pangea. While the climate was cold in the South Polar Region of this huge landmass, much of the emerging Pangea enjoyed temperate and warm climates. The Carboniferous received its name for the large coal deposits from this period that would in the distant future fuel the industrial revolution. These coal deposits owe their formation to a number of factors. Foremost, it is in the Carboniferous, that Earth's plant life starts to green and brown landscapes everywhere as those were mostly the two colors of nature because flowering plants did not yet exist. These photosynthesizing plants covering much larger areas of the planet very much increased the amount of oxygen in the atmosphere. The result was an atmosphere with a much higher oxygen content, up to 35% versus today's 21%. Other factors include the emergence of large swamp and forest systems around the equator and a much higher lignin content of Carboniferous trees. This mostly is due to the fact, that trees back then had a significantly thicker bark than our trees have today. The latter mattered even more because of the seeming lack of organisms that could break down the lignin of dead Carboniferous trees as fungi do with our dead wood today. Another potential explanation is that even though fungi capable of breaking down lignin may have been around at the time, a combination of continuously wet tropical conditions and extensive swamps ensured that massive amounts of plant material would be buried before they could decay. During the Carboniferous the formation of the supercontinent Pangea was well underway and this seems to have produced a rather uniform growing climate with little seasonality. In turn, this is likely to have favored the abundant plant growth that we see in the Carboniferous. Without wood being broken down, the carbon in the buried organic matter remained biologically fixed, removing it from the carbon cycle through which plants generate oxygen via photosynthesis through absorption of carbon dioxide and return it once they decay. Plant life in the Carboniferous colonized much of the land including the colder areas. New plants developed in the favorable climate of the Carboniferous including a much larger diversity of seed plants, tree-sized mosses, and the huge ferns that are often prominent in artistic renderings of Carboniferous plant life. If one wanted to call the Carboniferous the Age of Trees, there is certainly justification for that. Never before, had they grown

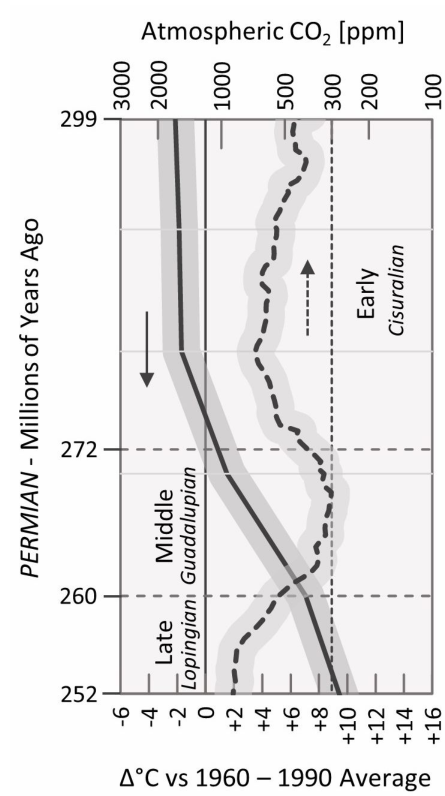


Carboniferous: Global temperature vs. modern average (solid line) and CO_2 levels (dotted line.

so numerous and so tall. The reason we usually do not refer to the Carboniferous as the Age of Trees is that trees of all kinds continued to be a dominant feature of Earth plant landscapes well after the Carboniferous and right into our time. Sometime, maybe as early as in the late Ordovician, but certainly in the Silurian and Devonian, life had for the first time left the sea to prosper also on land. It is in the Carboniferous that life takes into the air when insects become the first animals to evolve the capability to fly. How they did that exactly we do not know yet, and maybe we never will. Small insect bodies and specifically the delicate membrane wings of insects do not fossilize that well. Therefore, we likely will never find fossilized remains of the first insect wings. Animal life changed also in other ways as the high oxygen content of the atmosphere allowed animals to grow much larger. Animals whose growth otherwise would have been constrained by the efficiency of their respiratory systems were among the ones that grew to giant sizes in the later Carboniferous. We may all have seen the pictures of dragonflies the size of large birds. The Carboniferous ancestors of modern-day dragonflies could have wingspans of over seventy centimeters and were for all we know the largest flying insects ever. This gigantism may also have been encouraged by the lack of predators. Birds for example, prime predators of insects, did not exist back then. Flying insects such as the dragonflies were not the only giant arthropods in the Carboniferous. Surface dwelling insects such as millipedes grew to giant sizes as well, exceeding two meters in length, and may have well been the largest known land invertebrates ever.

Not only arthropods thrived in the Carboniferous, but amphibians also did well. The huge swamp areas of Carboniferous landscapes gave amphibians the opportunity to become the dominant land vertebrates. With that, sometime in the late Carboniferous, the first reptiles evolved from an amphibian ancestor. As reptiles branched off from amphibians, this is also the first time that land vertebrates stopped going back to the water to lay their eggs. Instead, reptiles started to lay eggs on land. In an amniotic egg, as the reptilian egg is called, the embryo develops inside the so-called amnion, a membrane that surrounds the fluid-filled embryo sac of reptiles; and similar also that of birds and mammals.[61] The four-legged vertebrates for which the evolution of the amniotic egg broke the connection between water and their reproductive cycle could now venture further away from the bodies of water where amphibians still needed to lay their eggs into and in which their young evolved. This opened up new land ecosystems to them that amphibians had not been able to take advantage of. Towards the end of the Devonian some fish species such as the heavily armored fish, the placoderms, became extinct.[62] They were replaced in the Carboniferous by more modern looking fish we would not be surprised to see swimming in our oceans today. Among the marine fossils of the Carboniferous, we find the distant ancestors of modern sharks. The warm and shallow seas of this period were home to many different kinds of brachiopods, mollusks long gone extinct now. Sea lilies were another plankton feeder prospering in the warm Carboniferous seas. Over millions of years, countless layers of their fossilized remains built many of our coastal limestone cliffs, still giving testament to their multitudes that endured throughout the Carboniferous. Toward the end of the Carboniferous, the climate became colder and eventually turned into a new glaciation phase lasting into the Permian period and therefore is commonly referred to as the Carboniferous-Permian glaciation.

The Permian: 299 - 252 Million Years Ago



Permian: Global temperature vs. modern average (solid line) and CO_2 levels (dotted line).

The Permian, the final period of the Paleozoic era, began with a cold period, the Carboniferous-Permian glaciation. The climate of the Permian continued to be shaped by the supercontinent Pangea shown in fig. 3.11, which would be fully assembled towards the end of this period. Throughout the Permian the southern part of Pangea located around the South Pole remained covered in ice. The vast interior of Pangea's huge landmass, centered at the equator, was far removed from the ocean that surrounded this supercontinent. Without the moderating influence of large bodies of water close by, Pangea's interior was dominated by a continental climate with high seasonality. Seasonal rainfalls and large temperature variations set in an overall arid climate must have made for desert-like conditions, a common feature in Pangea's interior. Only in the late Permian and early Triassic would wetter and overall warmer conditions prevail again and once more support vast rain forests near the equator. These drier conditions during much of the Permian favored seed plants that enclosed their seeds in a protective cover. Modern trees such as conifer, ginko, and cycad trees make their first appearance during the Permian. Plants that had been prominent in the wetter Carboniferous, such as ferns, retreated. Insects continued to thrive in the Permian including the insect giants from the Carboniferous period. Primitive relatives of today's cockroaches had already been around since the later part of the Carboniferous and by the beginning of the Permian, cockroach-like arthropods made up a large fraction of all insects, with some estimates going up to 90%. Currently, the oldest herbivore fossils date from the end of the Carboniferous. With lush vegetation becoming a feature of Earth's landscapes during that period, this is no surprise. However, it is only with the Permian that we see large herbivores and with that also large carnivores. The ancestral reptile species from which eventually dinosaurs and their flying relatives, the pterosaurs, would evolve, made their debut. Modern reptiles such as turtles and crocodiles can as well trace their ancestry to this same group of early reptiles. Importantly, it was also in the Permian that mammal-like reptiles first appeared. With the Permian, we finally have all the ancestral groups present that would in whatever circuitous way give rise to today's animal families. The evolution of all these new Permian animal species most certainly was conditioned by the need to adapt to the prevalent arid climate. However, with larger animal populations and diversity, competition for resources and the arms race between predators and prey must also have played an increasing role. Reptiles and insects seem to have coped best with the challenges of the Permian.

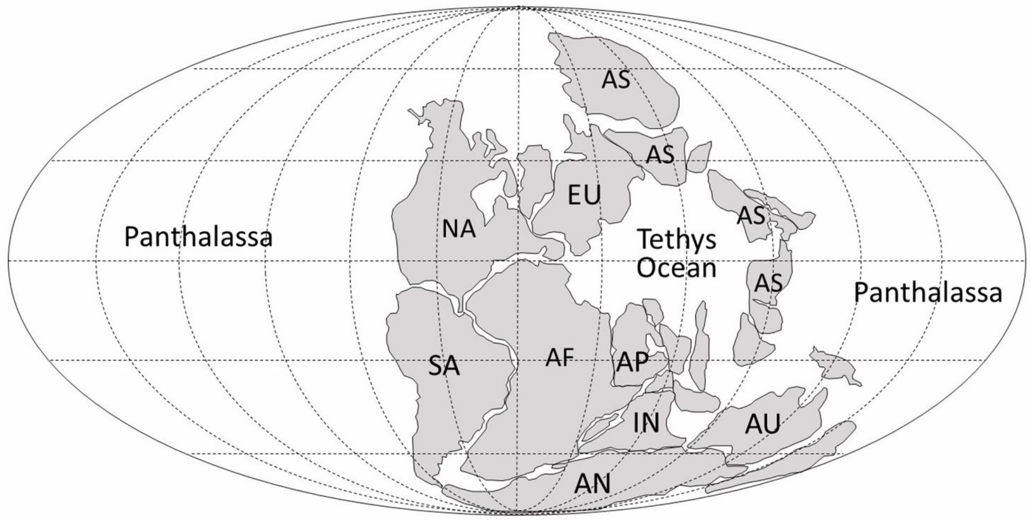
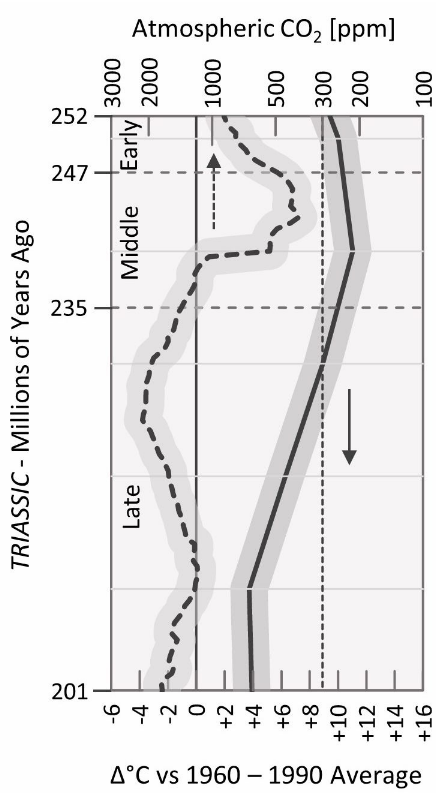


Figure 3.11: The Pangea supercontinent of the Permian. AF: Africa; AN: Antarctica; AP: Arabian Peninsula; AU: Australia; AS: various parts of Asia; EU: Europe; IN: India; NA: North America; and SA: South America. Map modified and adapted from Ref. [12].

This also included mammalian type reptiles; amphibians however, declined. The latter is not surprising given that by the middle of the Permian most of the Carboniferous swamp forests were gone and with that a large habitat that had been favorable for amphibians had vanished. The supercontinent Pangea, as shown in fig. 3.11, was surrounded by one single huge ocean called Panthalassa.[63] Little is known about this superocean. Ocean floors are recycled roughly every few hundred million years or so and only fossil evidence from shallow waters around continental shelves survives. The map in fig. 3.11 shows no islands at all in this large ocean surrounding Pangea because we do not know about any of them. Our Pacific Ocean today is home to a number of islands many of which support unique ecosystems. For all we know there could have been a Permian Hawaiian island chain but if so, it is quite unlikely we will ever know about the life it supported. What marine fossil records we have indicate large and thriving reef communities, home to a diversity of ecosystems including numerous sponge and coral species. Bony fish replaced the lobe-finned and spiny fishes that had given rise to the amphibians of the Devonian and Carboniferous; cartilaginous fish such as sharks and rays had been around for some time and continued to be successful designs. Nautilus species diversified during the Permian period; some were similar to the nautilus species we find today and most of them were now small. The northern part of Pangea, including a good part of the landmass that would eventually become part of Northern Europe, was covered by a shallow inland sea, the so-called Zechstein Sea. Zechstein is the German name for the sedimentary rock deposited over large areas in Northern Europe in the late Permian.[64] Inland seas such as for example the Baltic Sea today are mostly or sometimes even completely closed-off large bodies of water. If they are shallow and without connection to an ocean they will eventually evaporate ever so slowly and in the process precipitate their mineral richness. In the case of the Zechstein Sea, this richness was sealed and preserved until in modern times industrial mining of these mineral riches provided the foundation for a good part of Europe's industries. The Permian also saw the rise of the Ural Mountains as Siberia joined with Europe and southern Asia. As much effort as nature put into the evolution of species during the Permian, it almost all came to naught. At the end of this geological period, the greatest mass extinction we know about in the history of life on Earth removed not only many species but also extinguished many genera and eradicated complete taxonomic families. The Permian mass extinction took place about 252 million years ago and as it marks the boundary to the Triassic period, it is sometimes also referred to as the Permian-Triassic extinction event. The enormous loss of biodiversity that this mass extinction caused has earned it another name more eloquently expressing the sentiment of paleontologists confronted with the grim evidence of it - *The Great Dying*. Marine life took the greatest hit with up to 95% of species becoming extinct. However, this time, life on land was hit almost equally hard with maybe close to 70% of vertebrate species being lost. This mass extinction event has the singular distinction of also having nearly wiped out all insect life as well. Up to 90% of Permian insects seem to have been cockroach-like and if they were anything as resilient as their modern cousins are, this really says something about how devastating the Permian mass extinction was for life on Earth. Finally, the bell also rang for Trilobites, the perennial survivors, still around at the end of the Carboniferous, but in much smaller numbers. Trilobites had been part of Earth marine ecosystems for almost 250 million years, at times, specifically early on, thriving like few other marine species; and then, more resilient than most others, with some species surviving whatever nature was throwing at them. What an endurance record for a species, being around for such a long time and having survived all previous mass extinctions. For a comparison, we humans do not come even close to that. The earliest humans split of the chimpanzee line some six million years ago and the first modern humans originated in Africa only some 200,000 to 300,000 years ago; we have a very long way to go to match the resilience of Trilobites. As to what caused *The Great Dying*, there is still much speculation. From what it looks like, it was not just one event but more likely two separate events that bracketed the most trying time for life on Earth that we know of. The first extinction event seems to have occurred towards the end of the Guadalupian epoch which ended 260 million years ago and the second one about ten million years later at the end of the Permian. Maybe this double-whammy made the Permian mass extinction so much worse than the others that we know of. As to what the specifics are behind the early and the later parts of the two extinction waves, we know little. Paleontologists considered all the usual culprits such as potential catastrophic events including an asteroid impact, volcanic activity, or a sudden release of methane bound in methane hydrate on the sea floor. No evidence so far points towards devastating asteroid impacts comparable to what roughly 200 million years later would kill the dinosaurs; or to any of the other usual suspects. In addition, scientists looked into some of the more gradual processes that could also have been responsible. Among those are changes in sea level accompanied by oxygen depletion of the oceans as well as much drier conditions on land. It may well be that towards the end of the Permian multiple factors just changed in ways that Earth climate took a turn for the worse. Whatever the specific reasons, they certainly produced extreme greenhouse conditions. Global temperatures had started warming towards the end of the Cisrulian epoch about midway into the Permian period and by the end of the Permian average temperatures were almost ten degrees warmer than today's modern average. Carbon dioxide levels at the end of the period were very high with concentrations of almost 1,000 ppm or roughly two-and-one-half times the atmospheric carbon dioxide levels we have reached today. We may need to wait for more evidence to emerge, before we can hope to gain a better understanding of this Permian mass extinction event that resulted in the demise of so many species. Not long ago, scientists thought that the Permian mass extinction was the result of a cooling climate but newer discoveries have invalidated this hypothesis. A further alternative hypothesis is a much more protracted Permian extinction including a third, somewhat earlier extinction event. This could have started off the *The Great Dying* earlier than we assume until now, making it even more difficult to survive. We have to wait for this picture to become clearer but we do have a quite good understanding of the result. The beginning of the Mesozoic era provided a clean slate for new species to rise to dominance. As it turned out, the reptiles would make the most of it.

The Triassic: 252 - 201 Million Years Ago

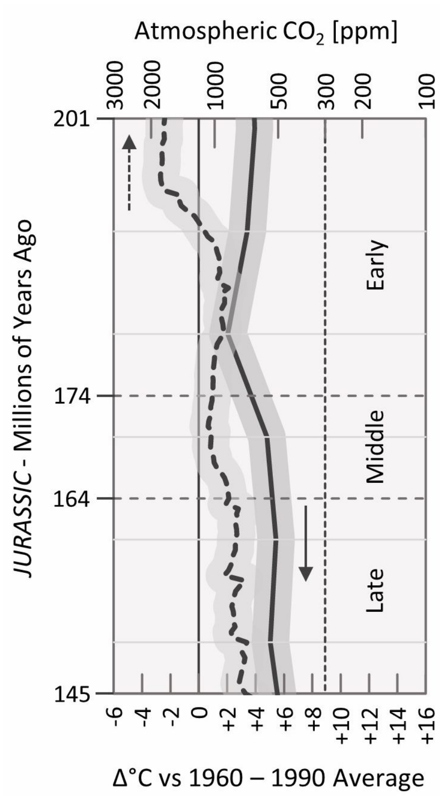


Triassic: Global temperature vs. modern average (solid line) and CO_2 levels (dotted line).

With the Triassic, we enter the Mesozoic era, which spans from 252 to 65 million years ago. First, the Triassic climate remained similar to the Permian climate, hot and dry with large seasonal rainfalls, dominated by the huge landmass of the supercontinent Pangea. In the middle of the Triassic, Pangea started to drift apart into a southern and a northern supercontinent and the climate became more humid. Gondwana, the southern continent, included South America, Africa, India, Antarctica, and Australia while the northern supercontinent Laurasia included North America and Eurasia. Polar ice had already melted in the late Permian and the poles would stay ice-free throughout the Triassic. In the early part of the Triassic, plant and animal life was recovering from the devastation of *The Great Dying*. It seems that most large trees were lost during the Permian-Triassic extinction. The fossil record shows that the biological traces preserved in rock samples before the extinction event contain many tree pollen. Samples after the extinction event lack them but indicate a highly elevated presence of fungi. Evidently, fungi were thriving on the decomposition of all the dead trees. It took time for plant life to grow back, which it eventually did, including the recovery of the conifer forests. Insect life recovered with plant life and during the Triassic, many new beetle species emerged. Completely new insect groups appeared too, such as grasshoppers. Extinction events clear the slate for new life to take center stage and that is often what paleontologists see in the fossil record of the first several million years after such an event. However, it looks like plant life did not take advantage of this in the Triassic, as it seems to have been only recovering what was lost with the Triassic-Permian extinction while seemingly little new evolutionary change took place. This was different for animal life where an increase in the diversity of land animals can be seen in the Triassic. Dinosaurs were a prominent part of that but more important for us, true mammals also made their debut.[65] Dinosaurs were not the first new reptilian group to emerge in the Triassic, turtles beat them by a few million years. The first dinosaurs appeared around 240 million years ago; as did the first true mammals which evolved from the mammalian-like branch of the reptiles. These earliest true mammals were small shrew-like creatures, their length not exceeding the size of an average human hand. The assumption is that these first mammals mostly lived off insects and in all likelihood were nocturnal animals, as suggested by their relatively large eyes. This adaptation, combined with them making their habitat either high up in trees or in burrows, would certainly have helped these small mammals to stay out of the way of the larger reptilian or amphibian predators. What makes them the first true mammals is of course that they did suckle their young. However, they did not give live birth but still laid eggs instead, just like reptiles. In contrast to these small mammals, the first dinosaurs, both the carnivorous and the herbivore kind, were getting large very quickly. With pterosaurs, reptiles also took to the air becoming the first flying vertebrates.[66] Dinosaurs and their flying relatives were not necessarily yet the top predators. Modern-day amphibians are mostly small animals, nothing many of us would fear. However, their distant Triassic relatives could become quite large, some of them more similar to large crocodiles than what we would recognize as an amphibian; and these amphibians not only had the looks of a crocodile but also the jaws and bite to go with it.

The Great Dying had devastated marine life even more so than life on land. The reduced diversity paleontologists see in the fossil fish records from the early Triassic is an indicator that not many fish species had survived. Nautiluses did survive and their coiled variety, the relatives of which still populate our seas today, as well as mollusks and sea urchins seem to have diversified quickly. In the Triassic we also observe for the first time land animals returning to the sea.[67] This includes the ichthyosaurs and the plesiosaurs, both large predators who similar to modern whales had to come to the surface from time to time to breathe air; and just as whales, they gave birth to live young.[68] The ichtyosaurs became the top predators of the Triassic seas and they would continue to terrify sea life for more than one hundred million years. Another mass extinction, the Triassic–Jurassic extinction event, marks the boundary between the Triassic and Jurassic periods. Again, we have no good understanding as to what caused this extinction event. Maybe it was somehow linked to the increased tectonic activity due to the ongoing break-up of Pangea but we do not know this for certain. While less severe than the Permian-Triassic extinction event, it still seems to have reduced marine biodiversity by about one third but the large reptilian marine predators, the ichthyosaurs and their kind, survived. On land, it looks like close to half of all species became extinct with most of the large predators perishing, including many large amphibians. However, several groups of large reptiles including the dinosaurs and their flying relatives survived. This set the stage for the dinosaurs and their kind to dominate the animal world of the Jurassic.

The Jurassic: 201 - 145 Million Years Ago



Jurassic: Global temperature vs. modern average (solid line) and CO_2 levels (dotted line).

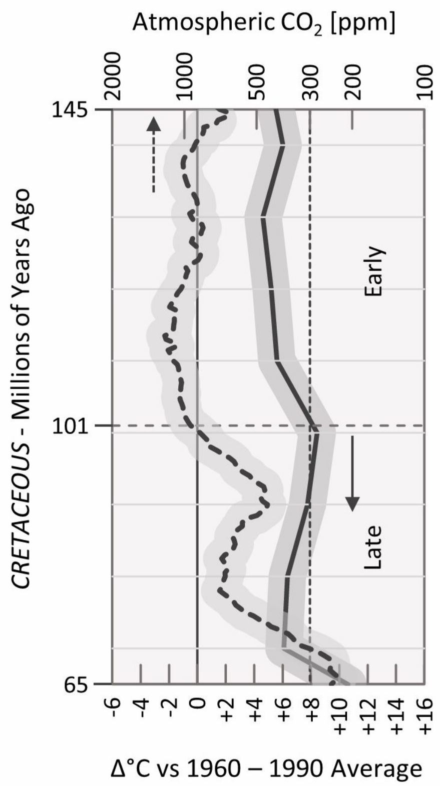
During the early Jurassic period the western half of Gondwana, Africa and South America, began separating from the eastern half, which included Madagascar, India, Australia, and Antarctica. The arid continental conditions that had been prevalent in the Permian and Triassic gave way to more humid climates as the break-up of Pangea and Gondwana progressed. Warm and tropical conditions helped foster more lush landscapes, still dominated by plants that already had been characteristic of the previous periods, such as conifers, ginkos, and cycads. As in the Triassic, a diversity of conifers continued to dominate the tree landscapes of the Jurassic. Cycad trees were also abundant and diverse and because they have a somewhat palm-like look, though they are in no way related to modern palm trees, one often finds renderings of Jurassic landscapes and movies studded with palm trees. The Jurassic period is the time when enormous dinosaurs began to roam the Earth. For many of us, this association with dinosaurs makes the Jurassic period the most familiar episode in the history of life on Earth. Who has not seen the Jurassic world portrayed more or less accurately in popular movies? The Jurassic saw the largest plant-eating dinosaurs ever but the terrifying *Tyrannosaurus rex*, the dinosaur star in many of the movies pretending to portray the Jurassic world, did not evolve until after the Jurassic. Even though, with some artistic license, the movies correctly portrayed the Jurassic as a world of dinosaurs, just populated with quite a few of the wrong animal characters. It seems as if in the Jurassic, almost everything was trying to grow as big as possible. It is in this geological period that we encounter the largest land animals ever to roam Earth, the gigantic sauropods.[69] The titans among those giants weighed in at more than one hundred metric tons with body length up to forty meters, including very long necks, in some cases exceeding fifteen meters, and corresponding tails for counterbalance and defense.

They must have been an impressive sight to see, herds of gigantic sauropods moving across lush Jurassic landscapes. However, not all dinosaurs were big. As a species, dinosaurs like all other species in evolution started out small. Many of them remained small and most of them would continue to look small when compared to the giants among the dinosaurs, which only arrived late in the Jurassic. Dinosaurs just have been extremely successful as a species radiating out to fill practically every niche in the ecosystems of the Jurassic. What kept mammals small while dinosaurs were around were not the giant dinosaurs but the many smaller ones for whom they were prey. In a sense, the dinosaurs that kept early mammals small were the equivalents of today's small animal carnivores whose predation keeps most rodents small. Later in Earth history, mammals would be equally successful and if we imagine all the ecological niches that in our time mammals occupy as populated by various dinosaur species we would not be too far off the mark. Of course, there are some differences as the giant mammals are gone now so the size range in the Jurassic, the difference between the very small and the very large, would have been greater in the Jurassic; and of course, there was no equivalent back then to humanoids such as dinosauroids. As long as dinosaurs dominated life on Earth, mammals seemingly would have to stay small and as much out of sight as possible.

An important dinosaur offspring that appeared towards the end of the Jurassic is the Archaeopteryx.[70] This fascinating fossil, first discovered in 1861 in the 150-million-year-old limestone deposits of a quarry in southern Germany, is one of those very precious links in evolution that paleontologists often look for in vain. Not in this case, as the Archaeopteryx is most likely a transitional fossil between dinosaurs and modern birds. However, we do not know if it is "the transitional fossil" linking dinosaurs and birds. For a long time, the Archaeopteryx had mistakenly been viewed as the earliest known bird. However, more recently it was discovered that its mode of flight was more like a gliding or flapping flight and not like the flight of modern birds. The relationship between dinosaurs, Archaeopteryx, and modern birds is still not clear but even if it should turn out that Archaeopteryx was not a direct ancestor of birds, it certainly announced their imminent arrival on Earth.

In the Jurassic oceans, the ichthyosaurs and plesiosaurs continued their reign but there were also other predators to avoid, such as giant marine crocodiles. Modern shark groups begin to appear as well and bony fish continued to thrive. Cephalopods are relatives of today's squids, nautiluses, and octopuses and had been around since the early Ordovician.[71] Squid-like cephalopods seem to have been abundant in Jurassic seas, which incidentally were also home to large numbers of coiled shell ammonites. Bony fish, cephalopods and ammonites were likely prominent among the food items sustaining the large marine predators although the bigger ammonites would have been hard to crack. The coiled shells of ammonites ranged in size from a few centimeters to about one meter and towards the larger size they were likely to be formidable predators themselves and less prone to become prey. The tropical climate also favored coral reef growth in the warm waters of the shallow seas created by the break-up and rotation of Pangea. In these reef ecosystems, the character casts that we find typically making modern reefs their home must also have flourished, including sponges, snails, and mollusks. The Jurassic also was a special time for plankton species. Four or five of the roughly one dozen plankton species we know from the fossil record evolved during the Jurassic. It has been suggested that an increase of free calcium washed into the oceans may have allowed evolution to experiment with plankton shells to an extent that was not possible before; maybe so and maybe warmer seas also contributed to the flourishing of plankton species in the Jurassic. A consequence of the proliferation of phytoplankton may have been algal blooms, which could have turned coastal ocean areas red. As we move from the Jurassic into the Cretaceous period, for once we have a transition between two geological periods not marked by a major mass extinction event. However, a minor extinction event still seems to have taken its toll but its impacts on fauna and flora are for now only poorly understood.

The Cretaceous: 145 - 65 Million Years Ago



Cretaceous: Global temperature vs. modern average (solid line) and CO_2 levels (dotted line).

During the Cretaceous period the break-up of the supercontinent Pangea progressed further. Seafloor spreading continued to widen the Atlantic Ocean and with the African and South American continents drifting further apart it started to extend itself much further south (see fig. 2.4d). A more pronounced geographic differentiation of life takes place during the Cretaceous as the dispersion of plants and the movement of animals across continents becomes limited by the surrounding oceans that increasingly separate the continental landmasses as they drift apart. It is in the Cretaceous, that Earth continental map begins to take its modern form. This includes India splitting from Africa in the Late Cretaceous. India, until then still attached to Madagascar, starts to break free from Africa and by about eighty million years ago India and Madagascar had separated and were now both isolated by the surrounding ocean as India began its journey northward to meet up with Asia. At about the same time, around eighty-five million years ago, Australia and Antarctica also started to break apart with the ocean increasingly separating these two continents that up until then had shared a very similar past. It would take another fifty million years before the separation of Australia and Antarctica would be complete in the early Oligocene epoch about thirty million years ago. Finally, South America and Africa also split with South America setting out on its journey to rendezvous with North America. At the same time, Africa moved north and closed the gap that was once the Tethys Ocean. While doing so Africa crashed into Europe and the Alps started to rise. The climate at the end of the Jurassic had started to cool and this cold spell lasted into the early Cretaceous. However, this happened in the Mesozoic climate environment, which was generally warmer than ours is, so a cold spell in the Cretaceous could still be relatively warm and humid. Both poles had small permanent ice caps and the cooler temperate forests surrounding the poles may have seen snowfall. This cool period at the beginning of the Cretaceous did not last long and by the end of the early Cretaceous, temperatures had risen to make for a tropic and humid climate to last for most of the rest of the period. Eustatic sea levels during the Cretaceous were much higher than they are today. By the middle of the Cretaceous, large inland seas covered vast areas of what is today North America, Africa, and Australia. This certainly had a significant impact on climate but also on marine biology. Carbon dioxide levels at the beginning of the Cretaceous were high and they stayed high until the late Cretaceous when they eventually started to fall below modern-day levels at the end of the Cretaceous. It looks like the trends for average temperature and carbon dioxide levels decoupled during the Late Cretaceous. For most of Earth history, we find that average temperatures and carbon dioxide levels are coupled, even if there is often a time delay as changes in one level catch up to changes in the other. This behavior makes sense as Earth in essence is a buffered system, with Earth's oceans acting as the most important climate buffers. Because of that, there can be a significant time lag before the full effect of a carbon dioxide increase materializes or ice age induced temperature variations are reflected in lower carbon dioxide levels. For example, some of these delays can be due to lower ocean temperatures, allowing more carbon dioxide to dissolve or they can be caused by less carbon dioxide being produced by diminishing amounts of decomposing organic materials as a result of a reduction in intensity and acreage of vegetation zones. Earth is a complex and interactive system and while today we can measure all the data required to correctly deduce which changes are triggered by whatever causes we do not have this information for the Cretaceous. Therefore, for now, we have to accept that we do not quite understand the relationship between carbon dioxide and average temperature levels in the Cretaceous, but hopefully this will be only temporary until new data becomes available.

Up until the Cretaceous, Earth plant landscapes were painted in green and brown, no red, no yellow or blue colors anywhere in Earth vegetation; except maybe for the occasional red algae bloom. This finally changed with the arrival of the first flowering plants. The plant life of the Jurassic continued into the Cretaceous with its ferns, cycads, and conifers but at the same time it started to look more similar to present day, as some forests now also included oaks, hickories, and magnolias. We know that in our modern world, plants and insects deeply depend on each other. The arrival of flowering plants in the Cretaceous marks the beginning of those profound relationships in nature where plants use insects for pollination while providing them food. Relationships on which our own survival very much depends, as without insects pollinating our crops we would have a much harder time to feed ourselves. Not surprisingly, the arrival of flowering plants seemingly goes hand in hand with the appearance of many modern insect groups in the Cretaceous, among them the first butterflies and ants.

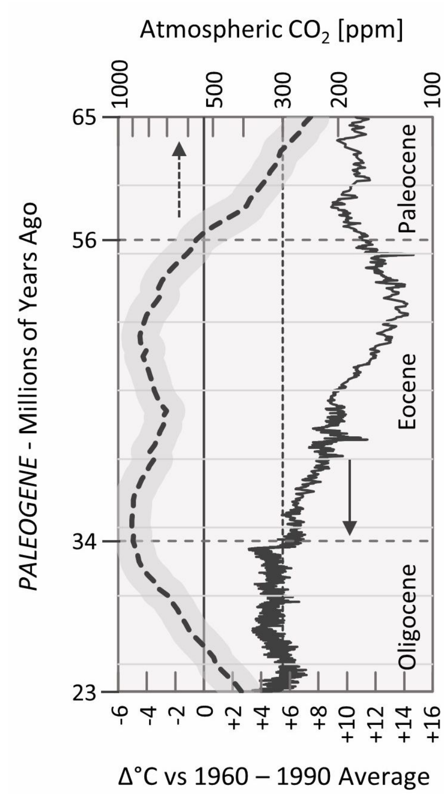
On land, dinosaurs and their kin continued to dominate animal life throughout this period. While some of the largest dinosaurs became extinct, dinosaurs continued to diversify with many new species evolving that now would be geographically constrained. To some extent, the separation of continents encouraged this multiplication of regional species. In many ways, dinosaurs are more representative for the Cretaceous than for the Jurassic. Among the many new dinosaurs were also large numbers of bird-like dinosaurs. These bipedal reptiles ranged from the very small to the very large and included such terrifying species as the *Tyrannosaurus rex*. The way we picture this animal in our mind may have to change as the simple reptilian skin that this top predator is often shown with may be far from the truth. Skin and feathers are not easily preserved in fossils, but it looks like that *Tyrannosaurus rex* may have been a lot more colorful and feathery than we thought; not that this would have made this apex predator any less terrifying. The flying reptiles, the pterosaurs, are still around for much of the Cretaceous but towards the end of it, they had almost vanished. It is not clear if birds had anything to do with that but in the Cretaceous, pterosaurs did not have the skies to themselves anymore. Increased competition by birds may have driven many of the smaller ones to extinction as seemingly only a few large pterosaurs were still around at the end of the Cretaceous. Mammals continued to evolve in the shadows of the dinosaurs, including placental mammals and marsupial mammals as well as monotreme mammals.[72] The evolution of the placenta in a sense moved the membrane that surrounds embryos in an amniote egg into the body and with that, mammals could give live birth. Marsupials also have a placenta, but the embryo spends only a short time in it before it is born and finds its way into the marsupial's pouch were the embryo matures. Monotreme mammals continue to lay eggs just like their earliest mammal ancestor did. The Cretaceous also sees the evolution of modern crocodiles. Some of these fresh-water crocodiles grew to huge sizes and they were certainly among the top predators of this period. Another successful reptilian branch were turtles, which now include some of the groups that are still with us today. The large predatory amphibians had mostly vanished during the Jurassic with seemingly only one large amphibian species surviving into the Cretaceous. However, the remaining large amphibians would not survive the Cretaceous and only amphibian groups whose descendants still exist today, such as frogs or salamanders, would continue.

Marine life in the Cretaceous oceans also was changing. The top predators of the Jurassic oceans, the species belonging to the group of Ichthyosaurs, were in decline and finally became extinct in the later Cretaceous. They may not have been able to adapt to the new kinds of fast swimming modern fish which had evolved in the Triassic and which by the Cretaceous had become widespread. Plesiosaurs were able to adapt better and were joined by new reptilian predators, among them the Mosasaurs.[73] Mosasaurs were large and fast swimming predators that like plesiosaurs gave birth to live young and needed to come up to the surface to breathe. Among the other marine reptilians, sea turtles seem to have done well with some growing to very large sizes but sea crocodiles vanished in the early Cretaceous. During this period, modern sharks also made their appearance. As for marine invertebrates, ammonites were still plentiful, the straight and coiled versions being joined by more bizarre shell shapes, the reasons for which are not known. Coral reefs continued to thrive in the warm seas of the Cretaceous, populated by essentially modern reef dwellers. The seafloor spreading driving the widening of the Atlantic seems to have happened at a much higher rate than we observe today. Scientists see the sharp increase in diversity and biomass of small marine organisms with mineralized skeletons that happened towards the end of the early Cretaceous as linked to this increased rate of seafloor spreading.

The Cretaceous–Tertiary extinction event marks the boundary between the Cretaceous and the Tertiary. This extinction event, often referred to as the K-T extinction event, wiped out about 75% of all species.[74] With few exceptions, all large land animals became extinct, including the dinosaurs. Something similar happened in the seas where the large reptiles vanished as well; and so did the ammonites and many other marine organisms. Even though many reptiles and their descendants such as birds would continue as nature recovered, this extinction event marks the end of the Age of Reptiles and the beginning of the Age of Mammals. As bad as it was for all the species that perished because of it, we have to be grateful. Without it, mammalian life and with it eventually humans would likely not have evolved the way it happened after that. There is agreement that the most likely cause for this mass extinction was the massive asteroid impact creating the Chicxulub crater in the Gulf of Mexico.

The Paleogene / Early Tertiary: 65 – 23 Million Years Ago

Tertiary has been the traditional name of the geological period stretching from 65 to 2.59 million years in the Cenozoic era. However, in the modern geological time scheme, scientists agreed to split the Tertiary into two periods, the Paleogene from 65 to 23 million years ago and the Neogene from 23 to 2.59 million years ago. For the purpose here, we will continue using the old Tertiary period name while at the same time folding in the new scheme. The climate of the Tertiary was warm and early in the period, during the Eozene epoch about 50 million years ago, global temperatures reached their highest values in the Cenozoic era with around 30 °C. This temperature increase was accompanied by a more than tripling of carbon dioxide levels between 65 and 50 million years ago. During this warm period, the Polar Regions remained ice-free but that would not last. Antarctica had been separated from the southernmost part of the supercontinent Gondwana in the Cretaceous. During the Paleogene, Antarctica and the rest of southern Gondwana continued to drift apart. As Antarctica lost its connection to South America, Africa, and Australia, it became increasingly isolated. With Antarctica surrounded by ocean at its southern polar location, Earth's ocean conveyor belt system started to change towards what is still in place today, including the Antarctic circumpolar ocean current. Because there is no landmass connecting Antarctica with any other continent, this ocean current is circumpolar, flowing clockwise around Antarctica from west to east (see fig. 2.10). It is this circumpolar current, keeping warm ocean waters away from Antarctica, which has enabled the build-up of the huge Antarctic ice sheets. Once lodged over the South Pole, isolated from the other continents and with the circumpolar current established, Antarctica's future was locked in, or putting it more picturesque, literally frozen in. The Antarctic climate started to cool and with the Oligocene, the Antarctic glaciation was well underway, eventually turning Antarctica into the vast continent of ice that we know today. Carbon dioxide levels continued to stay high until the end of the Oligocene when they started to drop to levels close to what they are today. At the end of the Paleogene, global temperatures had dropped significantly but were still several degrees higher than modern averages. This is one good example that glaciation phases and overall higher temperatures averages than today are not mutually exclusive. Eustatic sea levels (see fig. 3.9) were higher throughout the Paelogene than today, more than 200 meters higher for most of the Paleocene and Eocene epochs. However, they started to drop with the onset of the Antarctic glaciation at the beginning of the Oligocene to levels approximately 100 meters



Paleogene: Global temperature vs. modern average (solid line) and CO_2 levels (dotted line).

higher than today. Not only Antarctica separated from Gondwana in the Paleogene but all the continents drifted apart with South America now being isolated from Africa and evolving its own unique fauna and flora. As they did so, vast stretches of oceans opened up and the continental map slowly started to resemble our modern world. This process not only shaped Earth's climate but also reshaped life in the seas. Early in the Paleocene epoch when sea levels were higher, much of what is today North America, Africa, and Australia was still covered by the large inland seas that had existed since the middle of the Cretaceous period. This started to change and by the end of the Paleogene these inland seas had mostly vanished.

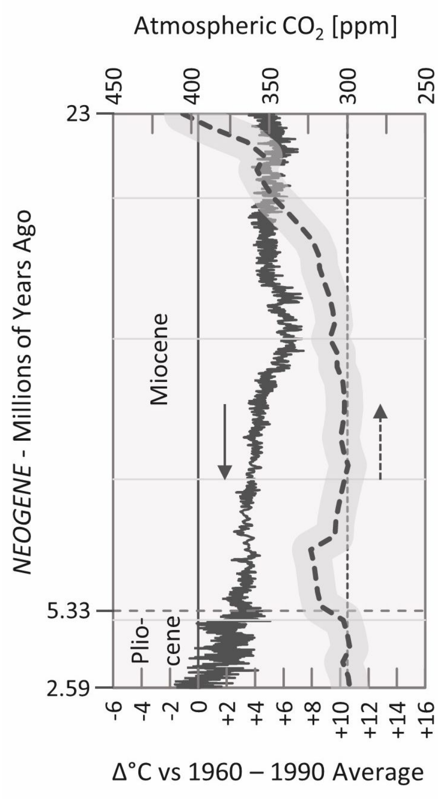
During the Paleogene, much of the modern plant life we can see today began to emerge. Flowering plants, which had first appeared in the Cretaceous, flourished during the Paleogene. The climate cooling towards the end of the Paleogene led to the spread of deciduous forests throughout the continents of the northern hemisphere. Different from rain forests, deciduous forests are home to tree species that shed their leaves annually. At the same time, rain forests and jungles became increasingly restricted to equatorial regions. Importantly for all the grazers that would eventually populate our planet, the first grasses started to appear towards the end of the Paleogene period. Of course, this is not only a milestone for herbivore evolution on our planet but eventually, this would also provide the foundation for our grain-based societies as all of those cereal grains that feed humanity today would evolve from grasses.

With the dinosaurs out of the picture, except for the line evolving into our modern birds, the Paleogene provided new evolutionary opportunities for the species surviving the K-T mass extinction to fill vacant ecological niches. Just like in stage plays, there is a limited number of roles to fill in the drama of life playing on Earth. Granted, the number of characters in the drama of life is much larger than in any stage play we would be watching in one of our theaters, but it is still limited. Over the past several hundred million years, different character sets have taken on these roles in the various ecosystems of life on Earth; like in a change of guards, as one set of characters leaves, a new one enters. The beginning of the Paleogene represents one of these character swaps and this time it is our mammalian ancestors moving into prominence and taking on lead roles.

The radiation of mammals is what characterizes the early Paleogene, the Paleocene epoch. Of course, it takes time for small mammals to grow to the sizes of large dinosaurs and to not only evolve into large size plant eaters but also to become top predators. For a long time it had been assumed that mammals were small in size as long as dinosaurs ruled the world but more recently quite large mammalian-like reptile fossils have been discovered that are dated back to the Triassic [47,48]. Maybe, after all, not only the dinosaurs had evolved to large sizes, mammalian-like reptiles must have been under similar evolutionary pressures. In how far this potentially implies also an earlier presence of some larger mammals we do not know yet. Regardless, it seems that all large herbivores and of course the carnivores depending on them for their meat supply vanished with the K-T mass extinction. Small mammalians, which were either more numerous or more successful, eventually became the dominant animals on Earth in the early Paleogene; except of course for the reptilian ancestors of modern birds. It took them some time to grow to bigger sizes though. In the Eocene epoch we see the first small ancestors of modern-day hoofed mammals; and small really means small here with for example ancestors of horses, camels, or elephants all not much bigger than the size of small dogs. Speaking of dogs, towards the end of the Paleogene, dogs, cats and pigs were abundant and so were different kinds of rodents. The first primates such as the ancestors of modern lemurs also made their appearance, as did the first modern birds. Some flightless birds grew to very large sizes, as there were seemingly no competing mammals yet to challenge them in the role of top carnivores. Among the reptiles, crocodiles not only survived the K-T extinction but thrived and the distant ancestors of today's turtles and snakes appeared towards the end of the Eocene. Like their land-based cousins, sea-based dinosaurs also vanished with the K-T extinction event. Shark species took advantage of that as they evolved to become the top predators of the Paleogene seas and have held that position ever since. The Eocene also saw the return of some mammals to the oceans including the ancestors of our whales even though they may have still been semi-amphibious.

The Neogene / Late Tertiary: 23 – 2.59 Million Years Ago

It is during the Neogene, that plate tectonics put the finishing touches on Earth continental map as we know it today. Since the Indian part of the Indo-Australian plate started colliding with the Eurasian plate some 40 million years ago the Himalayas have been rising, forming a mountain chain of about three thousand kilometers in length. The groundwork for the Tibetan Plateau was laid earlier, when the ocean crust that had separated India from Asia was ploughed under the Eurasian plate, as India slammed into Asia over millions of years. This resulted in a crustal thickening that provided the initial lift for the Tibetan plateau. As the Himalayas rose, the Tibetan Plateau was lifted further to an average height of around four thousand meters today. In Western Asia, the Caucasus Mountains arose while in Europe the Alps were rising as Italy was attached to Europe; and the Pyrenees started to rise as well as plate tectonics butted Spain into France. The Arabian plate rotated away counterclockwise from Africa thereby creating the Red Sea and starting the formation of the East African Rift. As we move from the Miocene to the Pliocene epoch the jostling of the tectonic plates in and around the Mediterranean Basin led to the Mediterranean Sea being closed off from the Atlantic several times and eventually leading to the Mediterranean Sea completely drying out. Once the Mediterranean Sea was cut off for good from the Atlantic, it likely took not much more than one thousand years to completely evaporate and leave behind an enormous salt flat, in an event that is referred to as the Messinian salinity crisis.[75] Once the Straits of Gibraltar opened again and water



Neogene: Global temperature vs. modern average (solid line) and CO_2 levels (dotted line).

poured in from the Atlantic, the Mediterranean was quickly refilled, maybe in as short a time as a couple of years; however, there are also scientists arguing that it took much longer to refill the Mediterranean Basin.

On the other side of the globe, the Indo-Australian plate moved Australia towards its current position. When this part of the Indo-Australian plate also began to collide with the Eurasian plate it gave rise to New Guinea, several Indonesian islands, Taiwan, the Philippines and other islands. Eventually, this process resulted in the emergence of the geological and topographical map of the western part of what we call today the Pacific Ring of Fire. On the North American continent, the Neogene witnesses the final uplift of the Rocky Mountains and the rise of the modern Sierra Nevada; the latter likely causing the opening of the Gulf of California. It was in the late Neogene, that North and South America were finally connected and the East and West coasts of the Americas permanently separated, isolating marine life in the Gulf of Mexico and the adjacent North and South Atlantic coasts from marine life in the Pacific. With the new land bridge between South and North America in place, ocean circulation changed and moved closer to the modern patterns. The Nazca plate had been pushing its way under the South American Plate since the early Cretaceous some 140 million years ago. However, it is thought that the uplift of the current Andes only start just before the Neogene period, some 25 million years ago, with the Andes rising in what looks like separated growth spurts to their current height throughout the Neogene.

The climate continued to cool in the Neogene, a trend likely initiated by the rise of the Himalayas. About twenty million years ago, in the early Miocene, Antarctica was covered by a solid sheet of ice. A little later, glaciation also started in the North Polar Region. Low sea levels and the emergence of a permanent North Polar ice cap resulted in North America and Greenland connecting. About 20 million years ago, the island of Iceland emerged from beneath the waves, pushed up by volcanic eruptions in the Mid-Atlantic ridge that separates the Eurasian from the North American plate. With the colder climate, carbon dioxide levels also dropped during most of the Neogene. Eustatic sea levels continued to decline as the Antarctic and Arctic ice sheets grew. As glaciation intensified during the Pliocene, sea levels started to drop below modern levels. It is believed that plate tectonics in the northern hemisphere and the low sea levels effectively prevented warm equatorial waters from reaching the North Polar Region.

The South Polar Region had been thermally cut off by the Antarctic circumpolar ocean current established in the Paleogene; now the North Polar Region followed suit. This may have been the first time since the Late Ordovician Glaciation, some 450 million years ago, that both Polar Regions were isolated from warm equatorial waters. The Neogene climate did not only get colder but also drier. In the temperate regions of the northern hemisphere, large areas of Paleogene deciduous forests retreated, pushed back by grasslands, steppes, or tundra. Some of that happened also in Australia, shrinking its once large rain forests. The tropical rain forest areas of South America and Africa also shrank. However, geological changes such as the rising Andes and the resetting of the ocean currents following the closure of the Isthmus of Panama helped preserve these reduced rain forest areas which became the cradle for much our modern-day rain forest biodiversity. On land the changing climate gave rise to expanding deserts in Northern Africa and Central Asia while the cooling oceans became home to the first kelp forests.

The changeover from deciduous forests to grasslands, steppes, and tundra must have been a major driver in animal evolution during the Neogene. Extensive grasslands and steppes favor different kinds of herbivores than dense forests, be it rain forests, conifer forests, or deciduous forests. Hoofed grazers of all kinds thrived and became larger, eventually becoming the herd animals of the grasslands that we can still witness today in some protected parts of Africa. Horses, first appearing in the Paleogene in North America grew to pony-size. Together with camels who also first evolved on the North American continent, they were among the animals that crossed the Bering Strait land bridge into Eurasia. Mastodons and elephants originated in Africa and spread to Eurasia and North America, while the ancestors of rhinoceroses originated in Eurasia and moved from there into Africa and North America. The ancestors of several hoofed animals had already made their appearance in the late Paleogene but their descendants and some new additions started only to thrive in the Miocene, including giraffes, deer, bison, sheep, and pigs. Among other familiar animals making their appearance in the Neogene we find hyenas, raccoons, and weasels.

The closing of the Isthmus of Panama ended the biological isolation of South America and fauna and flora started to migrate from the South to the North and vice versa. This included for example the giant sleuth and armadillo moving north and the raccoon moving south. The end of South America's biological isolation resulted in the majority of the large marsupials of South America, which in many ways resembled the marsupials of Australia, becoming extinct. Australia remained biologically isolated and retained much of its remaining rain forests with a thriving population of a diverse set of marsupial mammalians, filling all biological niche roles which on other continents placental mammals occupied, at least for the most part. Reptiles had lost their dominance to mammals, but some species such as crocodiles continued to thrive evolving to very large sizes; they must have been terrifying predators. The forebears of our modern snakes and turtles also grew much larger.

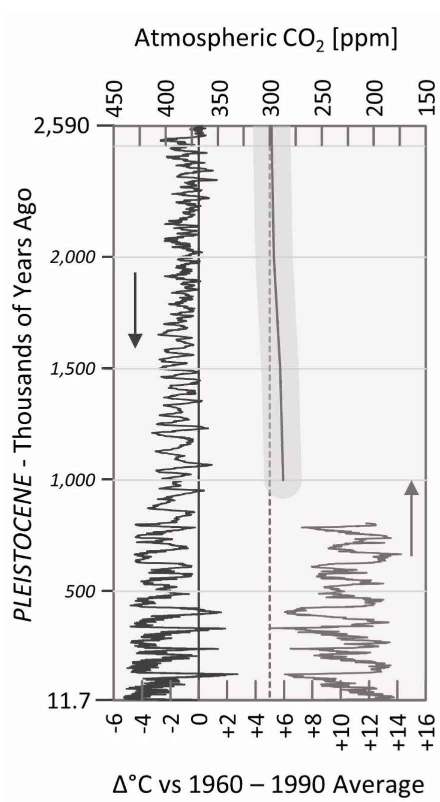
Among carnivores, there were still the large flight-less birds, the reptile's cousins, but towards the end of the Neogene, they were already close to extinction. The bird family also included several large species taking to the air with some of these birds weighing twenty-five kilograms and the largest species almost three times that weight. However, most birds remained small and continued their evolution towards modern forms. A number of carnivores preyed upon the large herds of hoofed herbivores, among them prominently the large cats, including of course the famous saber-tooth cats.

It was during the Neogene that prehistoric whales eventually became fully marine mammals. Sharks remained the top predators of the seas, some of them such as the megalodon sharks who first appeared at the end of the Paleogene grew to very large sizes, up to three times the size of the largest great white sharks of our time. The evolving marine kelp forests supported new species such as otters but also other species that are now extinct such as the dugongs, an animal species related to elephants and manatees.

The Neogene takes of course a special place in human evolution. Old world monkeys and apes are thought to have diverged from a common ancestor around the beginning of the Neogene. Proconsul, an extinct genus of primates that existed from 23 to 17 million years ago is believed to have been one of the first or maybe the first ape. Proconsul was a member of a primate species referred to as Dryopithecines, or tree apes, and a

Dryopithecine primate is thought to have diverged into two separate lines about 8-9 million years ago.[76] One of the two lines would eventually lead to gorillas with the other line evolving into what would become the common ancestor of humans, chimpanzees, and bonobos. The common ancestor of humans and the chimpanzee line, the latter including chimpanzees and bonobos, is thought to have diverged into the human and chimpanzee line sometime between five and seven million years ago. The chimpanzee line split about two million years ago into chimpanzees and bonobos and from what it looks today, bonobos retained more of the common ancestor of the human and chimpanzee line than did chimpanzees themselves [49]. It is the environments of the late Miocene and the Pliocene, that provide the stage for early human evolution.

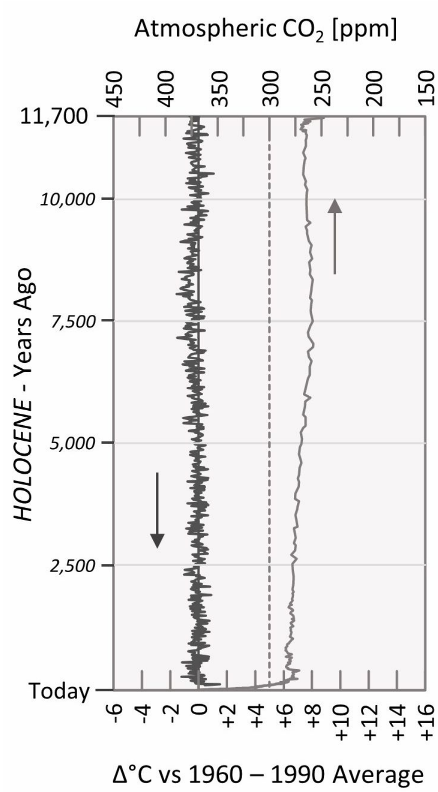
The Quaternary: 2.59 Million Years Ago - Today



Pleistocene: Global temperature vs. modern average (black line) and CO_2 levels (gray line).

The Quaternary is split into two epochs, the Pleistocene which lasted from 2.59 million years until 11,700 years ago followed by the Holocene epoch, in which we still live. The climate cooling Earth witnessed throughout most of the Neogene had set the stage for the severe glaciation phases of the Pleistocene. For good reason, the Pleistocene is often referred to as the time of the ice ages. The envelope of the Pleistocene temperature shows a continued drop throughout the epoch with the embedded temperature oscillations increasing as the Pleistocene epoch progressed. The carbon dioxide level shows a similar behavior for the later Pleistocene, where a sufficient number of data points becomes available at around 800,000 years ago. Before that time, the carbon dioxide curve has too little data points to reveal more details and in a sense represents a coarse average but that also shows a decrease in carbon dioxide levels as one would expect. The excellent correlation between the temperature and carbon dioxide data for the late Pleistocene was already highlighted in the previous chapter when we looked into the discovery of the ice ages and the efforts to explain them through the astronomical theory pioneered by Milutin Milanković. Earlier in this chapter, we have seen that the rise and fall of eustatic sea levels for the same time span, from around 800,000 years ago to the end of the Pleistocene, mirrors the temperature and carbon dioxide data. As glaciers advanced during cold or glacial periods and receded during warm or interglacial periods, sea levels fell and rose accordingly. It was the repeated lowering of sea levels during the Pleistocene which provided the land bridges enabling our human ancestors to spread to all continents, except of course to Antarctica or the North Polar Region. The alternating intervals of glacial and interglacial periods, specifically in the second half of the Pleistocene, led to large-scale migrations of plants and animals. During warm phases, vegetation would expand towards higher latitudes along with the animals that relied on such habitats. This is why fossil remains of animals whose living relatives we find today only on the African continent or in the tropical zones of Asia, turn up in northern latitudes. It may be difficult to picture for example northern Europe with a climate supporting elephants or rhinoceroses but that is what happened during the warm interglacial periods. Much of the animal world of the Pleistocene would have looked familiar to us but there were also some giant mammals that would look quite out of place in our world today. Animal giants, frequently referred to as the megafauna, had evolved on all continents among them giant kangaroos and marsupial lions in Australia, the giant armadillo and the giant ground sloths of South America, the North American mastodons, camels, dire wolves, saber-tooth cats and giant bears, and the Eurasian mammoth and woolly rhinoceros. Small numbers of some of them would survive into the early Holocene but most of them became extinct. The frequent climate changes of this period must have stressed animal populations and may have contributed to their extinction. Top predators are few in numbers and require large numbers of prey animals to sustain their small populations. If prey animals become fewer as vegetation becomes for example less abundant the first to go are the large predators. However, not only the large predators vanished from the scene, but the large herbivores also become extinct. We may not want to hear that, but the most probable cause that drove the megafauna to extinction is in all likelihood human hunting. On every continent where humans newly arrived, the very same thing happened. Everywhere we look, the forensic evidence of the fossil record shows that large animals started to become scarce and eventually became extinct not too long after humans arrived.

The Americas and Australia provide particularly good examples for this pattern. One mitigating circumstance humans could point to may have been the global glaciation period at the end of the Pleistocene that could have tipped the scales towards extinction for an otherwise already struggling megafauna. All of the Pleistocene period was embedded in an ice age that had started in the Neogene and that has not ended to this day as we currently are still in a prolonged interglacial period, which began with the early Holocene. The glaciation phases during the late Pleistocene were particularly severe and it is believed that there were at least eleven of these. Most severe was the last glaciation of the Pleistocene period, which was truly global with glaciers advancing far to the south. During this last massive glaciation, the so-called Laurentide Ice Sheet covered vast areas of North America. In some places, the thickness of this ice sheet exceeded three kilometers and it extended as far as New York City on the East Coast, swung southwest around Chicago where in the Midwest it extended below 40° latitude before it swung back north towards just south of the Canadian border and from there ran pretty straight westwards towards modern-day Seattle. As this ice sheet expanded and retreated, it gouged out the Great Lakes of North America. In Europe we also find a number of ice age lakes dating back to the Pleistocene ice ages but much smaller than the Great Lakes. As the ice sheets expanded and retreated, they changed the topography as advancing glaciers grounded down the rocks beneath them and receding glaciers left behind all the materials they had transported during their advance. Much of the landscape in today's temperate northern climates has been shaped by these ice ages.



Holocene: Global temperature vs. modern average (black line) and CO_2 levels (gray line).

For us, the most important development during the Pleistocene was of course the evolution of the human species including modern humans into which we will look in the following section. As we enter the Holocene epoch 11,700 years ago, modern humans are the only human species left. The Holocene epoch is a very brief epoch when compared with other geological time intervals. When we look at the temperature and carbon dioxide record over the few thousand years the Holocene has lasted so far, we see little change, similar for eustatic sea levels. If there is anything that our journey through the evolution of Earth and life on Earth should have told us so far, it is that over the long time scales that matter, the geological time scales, significant changes to Earth environment are the rule and not the exception. Therefore, we should not count on things to stay as calm for the next few million years as they have been over the past roughly 12,000 years. We are a bit handicapped here in our perception of time because our recorded history is so short compared to the history of life on Earth; two million years mean nothing to us but a few hundred or maybe even a few thousand years have meaning for us. Can anybody really picture where evolution could take the human species on a geological time scale? How about in say fifty million years, which is less than the time that has passed since the dinosaurs went extinct? Most likely not. The short time humans are around on planet Earth seemingly does not matter, or does it? It does, and to see that we just have to look around us to realize how much humans have altered the conditions for life on this planet during their relatively short presence on it. It is safe to say that with human evolution we see something very new and unique in the history of life on Earth. The most important event for life in the Holocene is the seemingly unstoppable rise of our species, literally in the geological equivalent of a blink of an eye, taking over the whole planet and becoming a major factor in the evolution of life on Earth itself. There has never been such a species on Earth before with the capability to change long-term environmental conditions for itself and every other life on Earth in such a massive way as we have been doing since the beginning of modern industry. True or not? Probably not, unless we discount bacteria, specifically the ones that did all the planetary scale engineering that provided the conditions for higher lifeforms such as us to evolve in the first place. Bacteria have been around for billions of years; modern human have not even pushed a million yet. Regardless, we are now similarly changing the conditions for life on Earth, for every life, including ourselves. We will look more into this predicament in the next chapter but now it is time to take a closer look at human evolution itself.

Human Evolution

It is not possible to give a complete picture of what we know today about human evolution in the space of a few pages. A number of excellent and highly readable books provide much more detailed accounts of human evolution than can be given here; the bibliography in the back of this book lists some of them. How much our understanding of where we come from has changed just over the very recent past should caution us that we do not know the story of human evolution fully yet, not by any means. Each new discovery or insight fills in another piece of the human evolution puzzle. We can expect this to continue for a long time to come, as we really do not know how big the mosaic is that we are trying to complete. Quite likely we will never be able to uncover the whole human family tree with all its numerous branches and twigs illustrating the evolution of our species and that of our close and distant human cousins over the past few million years. But there is much we can learn from a better understanding of human evolution. Only a few decades ago, our story of human evolution was very simple, easily fitting into a few paragraphs. Then we discovered that the story of how we became the species we are today is much more complicated than we had believed. There is no straight line connecting us to our earliest ancestors with only a couple of missing links that need to be bridged. Human evolution is a family story in the true sense. Today we know about many more family members than we could have even imagined a generation ago. To get a sense of how big this gap is, my generation only has to look back to what we learned about human evolution in high school. It is nothing short of amazing how dramatically our knowledge of human evolution has changed over a little more than one generation, the past forty years or so.

Before taking a closer look at this much more detailed picture of human evolution that science has given us, a few words are in order about how paleoanthropology gathers its records and builds our human evolution family trees. Finding a human fossil is one thing but understanding how, where, and when it fits into the larger story of human evolution is quite a different challenge. The first finds of archaic or ancient human fossils, to use the generic terms for humans predating our own species, have been made around the middle of the nineteenth century. Back then, science did not have the methods and tools which today allow researchers to understand where a fossil might fit into the story of human species evolution. So, all they could do was to hold newly found fossil bones next to those of apes and modern humans and analyze how they differed. Loosely speaking, the experts of the time were comparing shapes and dimensions of archaic fossil finds to what was known from modern humans and apes; and as more archaic human fossils were found they were of course also compared to each other. Such comparisons entailed measuring bones according to some metrics that the science of the time had agreed to as the most important ones to determine differences in sizes and shapes. Until a few decades ago, combining this kind of research with the best dating methods available at the time provided the only way to map out the landscape of archaic human evolution. Early dating methods derived the age of a fossil from the specific geological stratigraphy of a fossil site, associating the age of a fossil with the Earth layer it was found in. This method had been used for some time to understand the relative ages of bones of extinct animal species, so naturally it was also applied to date the first human fossil discoveries.

Relative dating was as close to an accurate chronology of archaic human fossils as the first generations of paleoanthropologists could get. As discussed in the previous chapter, the discovery of radioactive decay at the end of the nineteenth century led to the development of radioactive decay dating methods. With that a number of different radioactive clocks would become available for accurate dating of human fossils. We already heard about the uranium-based method that allowed scientists to date the oldest known rocks on Earth. This specific clock does not really work for human bones, but other ones do cover the time scale relevant for human evolution. Modern science and engineering have enabled a range of additional dating methods for paleoanthropology that today complement radiometric dating methods. These dating methods are not restricted to fossilized bones or teeth; they can also be applied to the archeological context of a fossil site itself. Dating of tools, animal bones, sediments, volcanic ash, or lava are only some examples. However, it does not stop there, as for example volcanic ash or the deposition of sediments can also provide relative dating in a much larger context. In this way, scientists can connect volcanic ash found at one site to a volcanic eruption that happened thousands of kilometers away, thereby linking two locations in distant time.

In addition, biology discovered its very own molecular clocks that rely on certain genetic changes also known as mutations, occurring at a steady rate per generation. Knowing the rate of change and being able to measure the accumulated change between DNA samples from two different archaic humans allows estimating the time elapsed since they separated from a common ancestor. Importantly, once DNA sequencing became available it was only a matter of time until scientists would be able to use this technique to reconstruct the DNA of archaic humans from DNA fragments preserved in bones. DNA does unfortunately not preserve well but as we will see, scientists have successfully reconstructed the DNA of some of our extinct human species cousins. Paleoanthropology has come a long way, from speculating about human relationships based on comparing bones to building a family tree of human evolution using molecular clocks and DNA analysis, and all that in a little more than one generation.

The first paleoanthropologists in the mid-nineteenth century had a very limited tool set at hand to apply to the few archaic human fossil finds known at the time. This allowed at best for only a myopic view on human evolution. Frequently, this situation was made even worse by strong preconceptions still widely held at the time. Back then, that certainly included the continued influence of religious beliefs on what in terms of human ancestry seemed to be acceptable and what not. However, even worse was the perceived sense of superiority that many derived from the political situation at the time where Europeans and their descendants had subjugated much of the world under their colonial regimes. Beliefs of racial superiority and membership in what seemed to be the master race had been justified with religious texts for some time, for an illustrative example one just needs to look at slavery. Eventually, some ideologies of racial superiority would even supplant religion. We all know how and where this ended and after the Second World War, ideas of racial superiority were not sociable anymore, at least not in public. Measuring human skulls to demonstrate the racial superiority or inferiority of one human being over another ceased to be respectable activities for academics seeking careers in biology or medicine. However, we should not kid ourselves; racial prejudices are still prevalent and in many places simmer just under the surface; given the opportunity, ready to pop out at any time. This is exactly why it is important that all of our children have access to education and learning that allows them to explore the facts of human evolution, untainted by biased beliefs they may find around themselves. Truly understanding our shared human evolutionary heritage is one of the best vaccinations against racial prejudice. In the face of human evolution, we are all the same which is not so different from what most religions proclaim anyway, that all humans are equal before their respective gods. Knowing is much preferable to believing. However, if access to knowledge is not available where one happens to grow up, for whatever reason, settling for the belief that we are all equal will do just fine; but we must live it.

Our very first distant human ancestors split off the chimpanzee line sometime between five to seven million years ago.[77] This range of two million years for the last common ancestor of human and chimpanzee lines leaves as lot of uncertainty as to when this critical step in human evolution really happened. Unfortunately, that is as accurate as scientists are able to pinpoint this event today; although there are some indications that the real date may have been closer to seven rather than five million years ago. Five to seven million years seems like a long time ago to us but on an evolutionary time scale, this is quite recent. In addition, we are not looking here at some run of the mill species that evolved certain capabilities and skill sets to compete for a while in the struggle for existence or the survival of the fittest, to stay in Darwin's picture.

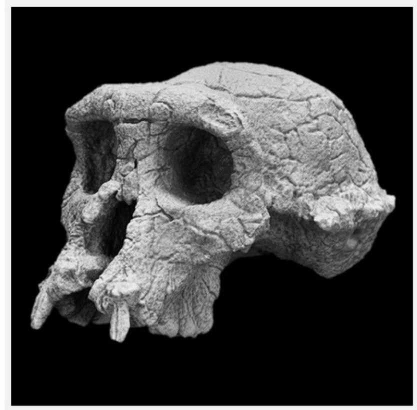
We need to look at human evolution through a different lens. We know of no other species which has evolved the creativity and learning capability on an individual and on a species level such as humans did. When we look at the evolution of species since the early Cambrian, we can see similarities and analogues as to how life biologically evolved and adapted. However, there is not a single example of a cultural or societal evolution of a species that remotely compares to what we see with humans; with human evolution, we are entering new territory. We have to be careful to differentiate human evolution into a segment where we can understand it in terms of the animal evolution that we share with so many other lifeforms that have come and gone on planet Earth; and into another segment where we have nothing against which we can compare the evolution of the human species. However, such binary views ignore that not everything is black and white, and that much of our reality is hidden behind various shades of gray.

From the outset, we have to acknowledge that at one end of human evolution we should be able to explain observations in the familiar language of animal evolution whereas on the other end we have to look for different explanations to account for human cultural and societal evolution. In-between those two ends will be a transition zone where it can get quite murky and where we will have to acknowledge that in many instances, we can only speculate but do not have good factual knowledge yet. In this chapter, we will start out with the animal evolution part of humans and reserve a discussion of human culture evolution for the next chapter. However, as we do so we will need to keep in mind that the animal evolution part we focus on in this chapter is increasingly adorned by evidence of cultural and societal evolution as we move from the distant human past to our more recent ancestors.

As far as we know today, human evolution produced seven different genus lineages. Our species belonging to the genus *Homo* is the only surviving one. How do we know that there were exactly seven? The answer is: we do not. Paleoanthropology is not only an exciting field of science but also a very competitive discipline. In some ways, the early history of paleoanthropology was more like a human fossil hunt, more of a competitive sport than what most would perceive as science, infested with a spirit maybe not to dissimilar from that driving the gold fevers of the nineteenth century when claims were staked and defended at all costs. Finding a new human species member entailed naming rights and unearthing a new species was akin to finding a gusher of a gold mine that could guarantee a scientific career next to bestowing fame, honor, and bragging rights. While things have changed and those times seem to be mostly over now, they have left us a human species classification that is still open to debate with some seeing more species and some seeing less. For all we know, we most likely have not yet discovered the fossil remains of all human species that once walked on Earth. Controversies about the number of genus lineages in human evolution are unlikely to subside and this debate will be rekindled every time a new fossil find is made that may not align with the current classification scheme. So, the number of seven human genus lineages should be taken with a grain of salt, it will be subject to continued discussion and in the future may be corrected either way, higher or lower; as it has been in the past.

Until quite recently, establishing relationships among and assigning fossil finds to specific human species relied exclusively on the correct reading and interpretation of fossilized bones. Then, over the last twenty years, the rise of modern genetics and increasingly sophisticated and fast methods to analyze DNA have enabled the study of ancient DNA from bone fragments or teeth. The boundary for the age of such ancient DNA or how fragmented ancient DNA can be to still provide valuable insights into human evolution has been pushed back far enough to significantly change our understanding of human evolution. As the species definition goes, a defining characteristic is that two separate species cannot successfully interbreed, that is, produce offspring that itself can procreate. Ancient DNA analysis has already shown us that there has been interbreeding between what were supposed to be separate species. Therefore, we can expect today's concept of seven human genus lineages and the number of species they include not only to be refined but also potentially change more significantly. We need to accept that the human species concept is likely much more fluid and certainly a lot more promiscuous than we had ever imagined. It is best to deal with the human species concept in approximations and not in certainties.

For many of us, the family names of our relatives that we will meet on our brief journey through human evolution, belonging to our genus or to one of the other six genus members of the human family, may not roll off our tongues easily. The next few pages should be looked at like a photo album from a family reunion where we get to see pictures of distant family members, some of them living generations ago that we never heard of. They may be members of families from different cultures and languages that happened to marry into our own family at some point in time, maybe a long time ago. As we will see, this is actually not so far removed from the truth as interbreeding seems to have played a much more important role in human evolution than previously thought. On our journey through human evolution, we will identify our extended family members by the names, science has given them, hoping they will not be too much of a tongue twister. If that however is the case, please give them nicknames if that makes them easier to remember; after all, this is just what scientists sometimes do.



Sahelanthropus tchadensis
Toumaï fossil

In chronological order, so they do frequently overlap, these seven human genus lineages are referred to as the genus *Sahelanthropus*, the genus *Orrorin*, the genus *Ardipithecus*, the genus *Kenyanthropus*, the genus *Australopithecus*, the genus *Paranthropus*, and the genus *Homo* to which our species belongs. The genus *Sahelanthropus* is the oldest we know and it contains only one species, *Sahelanthropus tchadensis*.[78] Paleoanthropologists reason that the first ancestor in the human family branch split off from the chimpanzee line between seven to five million years ago. Therefore, the species *Sahelanthropus tchadensis*, living sometime between seven to six million years ago must not only be one of the oldest species in the human family tree. It could also have been a contemporary species to the last common ancestor of humans and chimpanzees. The best known partial skull of this species has been nicknamed Toumaï [50].[79] Toumaï may or may not be among our ancestors, we do not know that. As the species name indicates, this fossil was discovered in an area that is part of modern-day Chad. This alone is a quite interesting fact by itself, as it makes it one of the fewer fossil finds not located along the corridor of East Africa's Great Rift Valley which extends from Ethiopia down to southern Africa and where most ancient human fossils have been found. From what the fossil remains show, *Sahelanthropus tchadensis* possessed a combination of ape-like and human-like features. Among the ape-like features are a small brain with a volume of around 350 cm^3, a sloping face, prominent brow-ridges, and an elongated skull. Our great ape cousins brain volumes fall within 275 – 500 cm^3, putting *Sahelanthropus tchadensis* closer to the lower end of this range. What points towards human-like features are smaller canine teeth and importantly the location of the spinal cord opening underneath the skull. The latter is an indication that the head of *Sahelanthropus tchadensis* was held in an upright position on its body. While this is usually associated with walking on two legs, we cannot really know how much use this species made of its bipedal capability. We do not know what became of the descendants of this species or if there were any in the first place. If there were descendants, they could have found their way into our ancestral line or into the ancestral lines of other species that split of the chimpanzee line but have since gone extinct.

Next in our chronology is the genus *Orrorin*, also with only a single species, *tugenensis*.[80] Its fossils were discovered in what is today central Kenya and are dated to around 6.2 to 5.6 million years ago. The fossils indicate that this species must also have been similar in size to chimpanzees. Since no skull has been found yet, we do not know what the brain volume of this species was but a small brain size similar to *Sahelanthropus* can reasonably be assumed. Like for *Sahelanthropus*, this species was given its own genus because of a unique combination of ape-like and human-like features that could be read from the fossils. This includes a bone buildup that is more typical for a bipedal movement and teeth more similar to modern humans than to chimpanzees. Just like for *Sahelanthropus tchadensis*, we cannot know how much use *Orrorin tugenensis* made of its bipedal capability.

Ardipithecus ramidus
ARA-VP 5/600 & 1/500
fossils

How much time did this species spend on the ground versus in the trees? We do not know. There is speculation that *Sahelanthropus* may be ancestral to *Orrorin* but we really do not know if that is the case either. With *Orrorin tugenensis* we face similar challenges and it is not clear yet how this species may relate to any of the later human species or if it went extinct before passing on part of its genes through interbreeding with other species. Given the fossil discoveries that continue to be made with new species being discovered as recently as 2017, it would not be surprising that future fossil finds may lead to a different scientific view of *Sahelanthropus tchadensis* and *Orrorin tugenensis* and how the two species they describe relate to other genus members of the human family. We can hope that at some point, we finally will be able to penetrate the fog that today still obscures our view on how sometime between five and seven million years ago, our human ancestors split from the chimpanzee line. That in turn could lead to a very different view of *Sahelanthropus tchadensis* and *Orrorin tugenensis* from how we see them today.

Next in line from a chronological perspective is the genus *Ardipithecus*. There are currently two species members of this genus: *Ardipithecus kadabba* and *Ardipithecus ramidus*, the fossils of which were both found in East Africa.[81] The fossils assigned to the species *Ardipithecus kadabba* are older and date back to around ∼5.8 to 5.2 million years ago while paleoanthropologists date the fossils of *Ardipithecus ramidus* to around 4.4 million years ago. It is not clear yet how the two species of *Ardipithecus* relate but it may be that *Ardipithecus ramidus* is descended from *Ardipithecus kadabba*. The genus *Ardipithecus* seems to have been closer to the chimpanzee line than to the next fossil in the chronology of our family tree belonging to the genus *Australopithecus*. Only for *Ardipithecus ramidus* do we have a skull, reconstructed from the cranial bones belonging to two different individuals [51]. Based on this reconstructed skull, its brain volume seems not to have changed significantly from earlier human species, remaining pretty much within the size range for modern chimpanzees. The current consensus is that the genus *Ardipithecus* and the genus *Homo* have a shared common ancestor, making them distant cousins at best rather than our genus *Homo* directly descending from the genus *Ardipithecus*. Essentially, *Ardipithecus* seems to have been a side branch of human evolution, closely related to the chimpanzee line but with adaptations such as smaller canine teeth and feet better suited for walking than those of a chimpanzee. We can look at the genus *Ardipithecus* as one of probably many evolutionary trials in the continuing split into the family branches of chimpanzees and humans. We do not know if we will ever be able to pinpoint a single common ancestor between the human and chimpanzee lines. Scientists have identified the Dryopithecine primate as the common ancestor who diverged some 8 to 9 million years ago into the gorilla line and the evolutionary line that became the common ancestor of humans, chimpanzees, and bonobos. This leaves a time window of some 2 to 3 million years for the common ancestor of the human and chimpanzee lines to emerge and to split into those two lineages.

Kenyanthropus platyops
KNM-WT 40000 fossil

However, we do not know who that common ancestor was or where he lived. The evolutionary process that resulted in this split between chimpanzee and human lines played out over a very long time, not just hundreds of thousands of years but more like on the order of a million years. At any given time multiple different species may have lived that had the potential to evolve into different family branches. Members of those branches may have retained their ability to interbreed over long times as we know from much later human species cousins, thereby complicating the picture even further. Some of them may have been closer to chimpanzees, some of them may have had more adaptations that are human-like; and there may have been species that were on a different trajectory altogether from both, chimpanzee and as well as human lines, but never evolved into a separate family we know of because they became extinct and left no trace. As we now know from our own genus *Homo*, separate species can and do interbreed long after splitting from a common ancestor. The same would have been true for this long time span during which the evolutionary tree developed into what we now see as the family branches of chimpanzees and humans, from the appearance of the Dryopithecine primate, to the emergence of the common ancestor of chimpanzee and human lines, to its eventual split into those two lineages.

Before we come to what may be our first direct ancestor we have to deal with one more genus, *Kenyanthropus*.[82] For this genus, we again have only one species representative, *Kenyanthropus platyops*. Our knowledge about this species is still very limited. The species fossils date back to ∼3.5 to 3.2 million years ago. So far, paleoanthropologists could not link *Kenyanthropus platyops* to any of the earlier or contemporary species and therefore, just like for *Sahelanthropus* and *Orrorin*, it is assigned its own separate genus. The debate about how close this species is to the lineage that may have led to modern humans, versus the *Australopithecus afarensis* species to which we will be introduced in a few paragraphs, seems not to be settled yet. The species designation *platyops* translates into flat-face and this is the most remarkable human-like feature of this species. Before the discovery of *Kenyanthropus platyops* the earliest known flat-faced human skull fossil was dated to around two million years ago. Hence, this species pushes the first appearance of a flat-faced skull back by more than one million years. While we have no indication for tool use for either *Sahelanthropus*, *Orrorin*, or *Ardipithecus*, scientists are considering the possibility that *Kenyanthropus* could have used tools such as bones or even stone tools. This hypothesis derives from the discovery of roughly 3.3-million-year-old stone tools not far from where *Kenyanthropus platyops* fossils had been found earlier [52]. The brain size of this species is estimated to have been around 450 cm^3 and with that is higher than for the earlier human species members of the genus *Sahelanthropus*, *Orrorin*, or *Ardipithecus*; and well within the brain volume range observed for members of the genus *Australopithecus*. *Kenyanthropus platyops* was a contemporary to other species but what stands out is that fossils of the genus *Australopithecus* were found in the same area as the *Kenyanthropus* fossils.

Australopithecus africanus
"Taung Child" fossil

Different species likely overlapping in time and space is something that scientists had not encountered before. Paleoanthropologists realized that this species proximity in time and space might have opened the door to new possibilities in human evolution that they had not considered previously. As we will see below, human species did indeed interbreed at later points in our species evolution. Therefore, we cannot exclude the possibility of this also happening much earlier. This brings us to the next genus in the line of the human family tree, *Australopithecus*.[83] This genus could be a likely descendant of the genus *Ardipithecus* but we do not know for sure. However, for now it looks like that *Ardipithecus* was more of a human sideline and *Ardipithecus* and *Australopithecus*, while maybe sharing a common ancestor, were more like cousins at best rather than *Australopithecus* descending from *Ardipithecus*. Regarding the relationship of our own ancestral human lineage with *Australopithecus*, the current hypothesis seems to be that in some shape or form, the genus of *Australopithecus*, *Kenyanthropus*, or a hybrid species derived from both may have been one of our direct ancestor, or more formally, the ancestor of the genus *Homo*. The first species of the genus *Australopithecus* is *Australopithecus anamensis*, the fossils of which have been found in Eastern Africa near Lake Turkana. The species designation *anamensis* reflects the geography of the fossil site as *"anam"* in the local language means lake. *Australopithecus anamensis* dates to about 4.2 to 3.9 million years ago. If *Australopithecus anamensis* descended indeed from the genus of *Ardipithecus* it would be a link between this genus and the *Australopithecus afarensis* which is thought to have descended from *Australopithecus anamensis*. We do not fully understand yet how the different members of genus *Australopithecus* may relate to each other and the assumption for now seems to be that all continued their own lines, *Australopithecus anamensis* as well as *Australopithecus afarensis*. *Australopithecus afarensis* dates back to sometime between 3.9 to 2.9 million years. Its fossils have been found only in East Africa and it is because of that also referred to as the Northern *Australopithecus*. The species designation *afarensis* also reflects the fossil site geography as it refers to the Afar region of Ethiopia. The species most famous representative was discovered in the Awash Valley of Ethiopia and has been nicknamed Lucy, a female skeleton that dates back to about 3.2 million years ago.

The fossils of the Southern *Australopithecus*, or *Australopithecus africanus*, more precisely translated as the southern ape of Africa, date to between ∼3.7 to 2.1 million years ago and as its name indicates, all its known fossil sites are located in southern Africa. The oldest fossil remains of *Australopithecus africanus* skeletons we have to date also represent the most complete one we have of this species. However, there is an ongoing debate whether this skeleton may not belong to a completely new member of the *Australopithecus* family, dubbed *Australopithecus prometheus*, the species designation of which references of course the Titan Prometheus of Greek mythology. Whichever is the case here, if *Australopithecus prometheus*, sometimes also nicknamed Little-Foot, is a separate species or not, this skeleton, as we will see below, opened a new perspective on how humans may have become bipedal. With ∼400 cm^3 and ∼450 cm^3 respectively, the Northern and the Southern *Australopithecus* had cranial capacities comparable to modern chimpanzees. Another member of the *Australopithecus* genus is *Australopithecus bahrelghazali* whose fossils, dated to ∼3.6 million years ago, were found in the Bahr el Ghazal region of modern-day Chad; hence its species designation of *bahrelghazali.* Fossils of a younger species of the genus *Australopithecus* have also been found in the Afar region of East Africa and this species, *Australopithecus gahri*, dates back to around 2.5 million years ago. The species designation of *ghari* has the meaning of surprise in the local Afar language reflecting the unexpected recentness of 2.5 million years for a member of the genus *Australopithecus*. Because of its chronological closeness to the first species of the genus *Homo*, but little difference from earlier species members of the genus *Australopithecus*, it is considered unlikely that *Australopithecus gahri* could have been an evolutionary path leading to the early species of the genus *Homo*.

A more recent discovery of an *Australopithecus* species in South Africa revealed that this genus was still around in the early Pleistocene. These youngest fossils in the *Australopithecus* genus belong to the species *Australopithecus sediba*, the species designation *sediba* having the meaning of "fountain or well spring" in the Sesotho language spoken in the part of South Africa where the fossils were discovered. The fossil skeletons of this species date back to ∼1.98 million years ago, placing them well into the early Pleistocene. While retaining most features of the *Australopithecus* genus, *Australopithecus sediba* also had some more human-like features. During the time span from shortly after the start of the Pleistocene to about 1.8 million years ago, or roughly for six to seven hundred thousand years, we have three human lineages overlapping in time: the genus *Australopithecus*, the genus *Paranthropus*, and the genus *Homo*. More interestingly, these species also overlap partly in geography. This is for example the case for the species *Australopithecus africanus*, *Australopithecus sediba*, and *Homo habilis* in South and East Africa. *Australopithecus sediba* shares features with *Australopithecus africanus* as well as with *Homo habilis* but the surprise for paleoanthropologists with *Australopithecus sediba* was that this very young member of the genus *Australopithecus*, in time and geography so close to *Homo habilis*, was not even more human-like. This observation certainly challenges any view of human evolution assuming an almost linear progression; human evolution seems to be much more complicated than that and may have had more dead ends than we assumed. *Australopithecus sediba* is believed to have been walking upright on a regular basis and shows some of the associated changes in the pelvis found previously only in later species of *Homo*. However, its mode of walking seems to have been quite different from what we find with modern humans, which points to possibly a different evolutionary path towards walking then the one our direct ancestors took. No skulls have been found yet for *Australopithecus anamensis* and *Australopithecus bahrelghazali* so we do not know what their cranial capacities were. However, their brain volumes were most likely not too different from those observed for the other species members of the *Australopithecus* family, which is between 380 - 460 cm^3; *Australopithecus africanus* marking the high end of this range. The brain of *Australopithecus africanus* was still not much bigger than an average chimpanzee brain, but something must have already changed. Members of the genus Australopithecus were likely the first stone tool makers and maybe even contributed to the first tool industry in human pre-history - the Oldowan tool culture.

Figure 3.12: A 3.6-million-year-old human footprint (Laetoli, Tanzania).

The oldest evidence for humans walking upright is associated with *Australopithecus afarensis*, Lucy's kind. Two of her kin are thought to have walked some 3.6 million years ago over soft volcanic ash in the Laetoli area in modern-day Tanzania, preserving their footprints for us to discover as shown for one such footprint in fig. 3.12. As we have seen above, the earliest human species we know of, *Sahelanthropus tchadensis*, already seems to have had bipedal capability. We know that if needed, modern apes can walk as well, although somewhat clumsy. This is due to the build of their hip and the angle the thighbone connects into it, which makes their walk more like a swagger and energetically more costly compared to the way we walk. It is difficult for chimpanzees to run in this way any distance at all and if they really need to get away quickly, they do so by knuckle-running. The adaptation of knuckle-walking or running has allowed apes to keep the dexterity of their fingers and with that the climbing skills vital for their survival while at the same time being able to move on four "legs" on the ground. Apes evolving to maintain their finger dexterity while being able to move on the ground on all four limbs seems like the natural prerequisite for a later transition to walking. Therefore, we should not be surprised to find anatomical changes indicating some more advanced bipedal capabilities already in early humans such as *Sahelanthropus* and *Orrorin*. What triggered the transition from knuckle-walking to real human walking is still much debated. Some scientists see this adaptation to have been habitat driven. Changing climates and vegetation patterns could have forced our human ancestors to forage further into the extending grasslands as forest areas began to shrink. However, we now know that the climate change that could have produced this environmental forcing only happened later. Hence, our human ancestors more likely learned to walk in densely forested areas.

Another explanation sees he evolution of human walking closely tied to human tool making which really only could start once human hands were freed up and our human ancestors could put the finger dexterity they inherited from their knuckle-walking forebears to good use by making the first tools. Most recently, a new hypothesis suggests that human walking actually did not evolve first on the ground at all. A detailed study of the above mentioned almost complete 3.67-million-year-old *Australopithecus* skeleton of Little-Foot indicates that humans may have made their first steps – no pun intended - towards upright walking not on the ground but high up in the trees [53]. Presumably, only at a later stage was walking so-to-say perfected on the ground. Looking at how modern apes walk on tree branches upright this intuitively makes sense, but we can expect this explanation will not be the last word on how humans learned walking. For all we know today, human species may well have acquired the skill of walking more than once.

Paranthropus boisei
Nutcracker Man (OH-5) fossil

We do not know that yet for certain, but it may be that the transition towards becoming bipedal could have been made by not only one species but by several species independently of each other, separated by geography, time, or both. It just so happens that only one species has passed down its way of bipedal locomotion to us. Before we come to our genus *Homo*, we have to discuss briefly one more remaining genus, *Paranthropus*, also sometimes referred to as the robust australopithecines.[84] There are currently three different species assigned to this genus, two of them, *Paranthropus robustus* and *Paranthropus boisei*, originally designated under the genus *Australopithecus*. The third member of the *Pantarhropus* family is *Paranthropus aethiopicus*. The species designation *robustus* speaks for itself, *boisei* was named for the man who financed the expedition that found this species skeleton [85], and the species designation *aethiopicus* stands of course for the geography where this species was discovered. Reassignments from one genus to another or the need to add a new genus are an indication of our increasing understanding as more fossils of the respective species are found and studied. Of course, there is also a different way to interpret this and that is as an acknowledgment that we do not yet understand quite a few things about human species and therefore put what we cannot really make sense of in a new drawer. The defining characteristic of *Paranthropus* species members are their skull fossils that point towards very strong facial muscles, anchoring what must have been particularly large chewing muscles that powered massive jaws. Those jaws also held much larger teeth than modern humans have. It looks like the diet of *Paranthropus* included tough fibrous plants and roots which we as modern humans would not be able to grind down with our teeth. The fossils of *Paranthropus aethiopicus* discovered in East Africa date back to $\sim$2.7 to 2.3 million years ago; those of *Paranthropus robustus* found in southern Africa date to $\sim$2.1 to 1.2 million years ago. For this species the designation *robustus* specifically refers to tooth and face sizes and not to its body size. *Paranthropus boisei* fossils found in East Africa date back to $\sim$2.4 to 1.4 million years ago. The massive jaws of this species supported cheek teeth four times the size of those of modern humans, earning it the nickname Nutcracker Man. The brain volume of *Paranthropus aethiopicus* falls within the range found for the australopithecines. The cranial capacities of *Paranthropus boisei* and *Paranthropus robustus* are larger than within the *Australopithecus* genus and their brain sizes approach the lower end observed for Homo habilis. However, paleoanthropologists do not see species members of the genus *Paranthropus* as directly ancestral to our own human lineage. We just know too little yet to understand if and how they may be related to other contemporary species from the genus *Australopithecus* or the first species members of the genus *Homo*.

The first member of the group of species we refer to as the genus *Homo* is *Homo habilis*. The naming of the species, literally the "handy man", indicates that he once was thought to be the first maker of stone tools. We know today that the first primitive stone tools predate the oldest *Homo habilis* fossil by almost one million years.

Homo habilis
OH-24 fossil

The fossil record of *Homo habilis* stretches from ∼2.4 to 1.4 million years ago. Most *Homo habilis* fossils were found in East Africa and so were fossils of a species referred to as *Homo rudolfensis* dating to around 1.9 million years ago. The available fossil records of *Homo habilis* and *Homo rudolfensis* seem to indicate that the two species were quite similar. The cranial capacities for both species are significantly larger than for any member of the *Australopithecus* family with brain sizes for *Homo habilis* and *Homo rudolfensis* falling between ∼550–690 cm^3 and ∼530–700 cm^3, respectively. However, the slightly larger brain case of *Homo rudolfensis* does not quite fit into the size range observed for the *Homo habilis* species. This was reason enough, as it seems, for the experts to continue ranking *Homo rudolfensis* as its own species in the genus *Homo*. The working assumption of paleoanthropologists is that *Homo habilis* somehow evolved from the *Australopithecus* family. However, where and when that happened and which *Australopithecus* species could be ancestral to *Homo habilis*, we do not know. The earliest fossil record we have for *Homo habilis* is about 2.4 million years old. The youngest fossil records we have for *Australopithecus anamensis*, *Australopithecus bahrelghazali*, and *Australopithecus afarensis* are roughly ∼3.9, ∼3.3, and ∼2.9 million years old. Even the shortest gap between the youngest fossil record among these three species, the one for *Australopithecus afarensis*, and the oldest *Homo habilis* fossil is still about 500,000 years. This makes it not impossible that any of these species could be ancestral to *Homo habilis*, but quite unlikely. Therefore, if *Homo habilis* is part of our ancestral lineage, Lucy is likely not one of our distant direct ancestors, but rather a cousin, by whatever degree removed. However, paleoanthropologists believe that *Australopithecus afarensis* is ancestral to *Australopithecus africanus* and through this lineage possibly could be ancestral to genus *Homo*. Which brings us to the other three *Australopithecus* species, *Australopithecus africanus*, *Australopithecus gahri*, and *Australopithecus sediba*. The fossil record of *Australopithecus africanus* with its earliest fossils dating to ∼3.7 and the youngest fossils to just around two million years ago has a roughly 300,000-year overlap with early *Homo habilis*. *Australopithecus gahri* has very little overlap in time and none at all in the geography of its fossil sites with genus *Homo*. This leaves *Australopithecus sediba*. As we have just discussed, while geographically and in time its fossils do overlap with *Homo habilis*, it is actually too different for being so close in time, sharing not enough human-like features with *Homo habilis* to be its likely ancestor. Therefore, from the current fossil record, *Australopithecus africanus* would seem to be the most likely ancestral species to *Homo habilis*. That is, if we could place fossil records of the two species close to each other not only in time but also geographically. Which we can. However, there is still a significant gap between those two species in many respects. While this makes it not impossible for *Australopithecus africanus* to be a direct ancestor to the genus *Homo*, it however makes it unlikely, pending of course new discoveries. Therefore, it seems to be prudent for now to admit that the existing fossil record does not really allow us to arrive at any definitive conclusions yet.

Homo erectus ergaster
KNM-ER 3733 fossil

Which, if any, of the *Australopithecus* species could have been ancestral to *Homo habilis*? For all we know, we may not even have discovered that species yet. Both, *Homo habilis* and *Homo rudolfensis*, were quite different from the next human ancestor entering the stage which is *Homo erectus*. *Homo erectus* fossils date back to 1.9 million years ago as indicated by a small skull fragment found in East Turkana, Kenya. *Homo erectus* is the first human species we know of that migrated out of Africa. So far, the oldest *Homo erectus* fossils outside of Africa date to around 1.8 million years ago and they have been discovered near the town of Dmanisi in modern-day Georgia. These Asian *Homo erectus* fossils are surprisingly old, the oldest human fossils ever found outside of Africa. This led to a number of researchers arguing for some time that this species may not have originated in Africa after all. However, that debate is now settled, and the scientific consensus is that *Homo erectus* did originate in Africa. Paleoanthropologists refer to the *Homo erectus* species that remained in Africa as *Homo erectus ergaster* to differentiate it from its Asian cousin. Sometimes the African and Asian variant are even referred to as different species, *Homo ergaster* and *Homo erectus*, respectively. Here we will use *Homo erectus ergaster* when we talk about the African variant of the species specifically. The fossil record for *Homo erectus*, the "upright man", begins roughly 1.9 million years ago and continues in small pockets in Asia until very recently, maybe up to around 150,000 years ago, when this species eventually became extinct there too. That makes *Homo erectus* one of the longest surviving human species ever, certainly the longest-lived species member of our genus *Homo*.
For *Homo erectus ergaster*, the "upright workman" - a reference to the advances in stone tool making this African variant of *Homo erectus* is associated with – the fossil record is considerably shorter, stretching from ∼1.9 to 1.4 million years ago. While *Homo habilis* still had retained some ape like features such as longer arms and a protruding lower face and jaw, the anatomy of *Homo erectus* is in many ways moving closer to modern humans. For such a long-lived species as *Homo erectus*, it is not surprising that the fossil records show a high variability over time and space. *Homo erectus* was significantly taller than *Homo habilis* whose height is thought to have averaged around 1.3 meters. Until the arrival of *Homo erectus* all human species we know of were similarly small, essentially not much bigger than chimpanzees; and up to *Homo habilis*, with the exceptions of *Paranthropus boisei* and *Paranthropus robustus*, they also all had brain sizes fitting comfortably in the size range we know from chimpanzees. It is only with *Homo erectus* that humans up from the waist were less ape-like and their skeletons much more closely resemble those of modern humans. The famous Turkana-boy fossils found in East Africa and determined to be 1.6 million years old indicate that this roughly eight- to nine-year-old boy was already 1.6 meters tall. Even though scientists know that children of such archaic humans matured earlier than our children do today, Turkana-boy was likely not fully grown when he died. The assumption is that an adult male *Homo erectus* could have exceeded 1.8 meters in height. Interestingly, in Turkana-boy's time *Homo habilis* was still around.

Homo heidelbergensis
Kabwe 1 fossil

Hence, we have two contemporary species with a nominal overlap of almost 500,000 years, one smaller and more ape-like and the other one much more similar in size and skeleton to modern humans, living in practically the same area, East Africa. How interesting it would be to know more about how two such different early human species may have interacted. With larger bodies, we also find larger brain sizes for *Homo erectus* and *Homo erectus ergaster* ranging from ∼850 – 1,100 cm^3 for the former and ∼700 – 1,100 cm^3 for the latter. These large brain size ranges can be due to a strong sexual dimorphism, the male of a species being much larger than the female, or an evolutionary progression towards larger brain sizes over time, or both. The next milestone in our human species fossil record is *Homo heidelbergensis*. With *Homo heidelbergensis* we get closer to modern humans, however, the picture becomes more complicated. That is not because we have less information than for earlier human evolution. Quite to the contrary, the difficulty is that there is actually more information that needs to be mapped into a significantly more complex picture of human evolution than for example the one visible to us today for much earlier human species. In a sense, as we come closer to modern humans, the coexistence of several human species and how they relate or to what extent they may have intermixed becomes much more tangible than when we look at coexisting species in the much more distant past. While we do have more fossil finds and archeological data as we get closer to modern humans, this does not necessarily give us by itself a sufficiently more complete idea of what the human family tree may have looked like. There are limitations as to how much the analysis and interpretation of bones can help us to understand common ancestries of modern humans and other human lineages that lived alongside modern humans but are now extinct such as the Neanderthals. This became quite clear with the fossil discovery of a finger bone fragment, a pinky as it turned out, of a juvenile female in the Denisova cave in the Altai Mountains of Central Asia. The fossil find was dated to 40,000 years ago which would not have been so remarkable. The real surprise came with the analysis of the human DNA contained in the bone fragment. The DNA analysis gave scientists conclusive evidence that the young Denisovan female was related to Neanderthals and less closely related to modern humans than Neanderthals are. That scientists today can draw such conclusions from just a small bone fragment of a human living some 40,000 years ago is due to the remarkable progress made in DNA sequencing over the past twenty to thirty years. Only a generation ago, it seemed unlikely we would be able to decode the complete human genome. However, in the twenty-first century it became a reality with the decoding of the first complete human genome accomplished in 2003. At the time, DNA sequencing was still quite expensive, and most science institutions would not have been able to afford it. Amazingly, it took only a decade of technological innovation to make DNA sequencing much more affordable. Not just for anthropological uses such as analyzing ancient DNA but also for private citizens to research their genetic heritage and maybe much more importantly, for the development of new medical therapies. Ancient

DNA analysis is significantly more challenging than analyzing our own DNA. DNA does decay over time and microbial DNA frequently contaminates ancient DNA samples scientists can extract for example from archaic bones or teeth. In addition, researchers also have to be very careful not to cross-contaminate ancient DNA with their own DNA or other ancient human DNA handled in their labs. Scientists overcame all those obstacles and ancient DNA analysis has become a commonly available tool for researchers to use. The rapid progress of ancient DNA analysis enabled the complete decoding of Neanderthal DNA only a few years ago. This revealed two things that one could not learn from interpreting fossils: first, modern humans had inherited a small amount of Neanderthal DNA, a clear indication that at some point modern humans and Neanderthals interbred; and second, it provided a time frame for when the populations of Neanderthals and modern humans had separated. Neanderthals and modern humans had a common ancestor from which they diverged sometime in the distant past. When their descendants met again much later, they interbred, probably around 50,000 years ago. However, that may not even be the final word yet on modern humans interbreeding with Neanderthals. Modern humans were present in Europe much earlier than previously thought as we recently learned from an improved understanding of human fossils discovered in Greece in the late 1970's and dating back to some 210,000 years ago [54].

Until a few years ago, there was little doubt that the common ancestor of modern humans and Neanderthals was *Homo heidelbergensis*. The latter in turn may have descended from populations of the African variant of *Homo erectus*, *Homo erectus ergaster*; or alternatively, from *Homo antecessor*, another archaic human that is thought to have been the European form of *Homo erectus* or *Homo erectus ergaster*; or alternatively, an early form of *Homo heidelbergensis*. It is not clear yet how *Homo heidelbergensis* relates to either *Homo erectus ergaster* or *Homo antecessor*. However, a plausible scenario seems to be that some *Homo heidelbergensis* populations or their ancestors migrated out of Africa while others stayed behind. *Homo heidelbergensis* fossils have been found in Africa, Asia, and Europe and this species fossil record stretches from roughly 600,000 years to 400,000 years ago. *Homo antecessor* fossils have only been found in Western Europe so far and date from between 1.2 million years to 800,000 years ago. If *Homo antecessor* is indeed a European variant of *Homo erectus* than a *Homo erectus ergaster* population must have moved into Europe more than 1.2 million years ago. Alternatively, *Homo antecessor* could be an early form of *Homo heidelbergensis* but then again, a migration of this early ancestral *Homo heidelbergensis* population out of Africa must have occurred earlier than 1.2 million years ago. What we have here is an unfortunate gap in the fossil record that we only can hope future fossil finds will eventually close. That may then confirm one of the two alternatives given above or maybe we will find something completely different. Whichever it will be, one thing seems to be certain, we are looking here at a second large human migration out of Africa after *Homo erectus* moved into Asia at least 1.8 million years ago but we do not know which human species did the migrating or exactly when that happened. The assumed brain size of *Homo antecessor* is somewhere in the range of $\sim$1,100 - 1,150 cm^3 putting it indeed between that of *Homo erectus* and *Homo heidelbergensis*. However, that by itself does not yet tell us if *Homo antecessor* is really an intermediary human species between those two and with that potentially our own ancestor.

Before the discovery of the Denisovan fossil the hypothesis had been that a *Homo heidelbergensis* population staying in Africa became ancestral to modern humans while the *Homo heidelbergensis* population that migrated into Eurasia evolved into what became the Neanderthals. The analysis of Denisovan DNA and its comparison to Neanderthal DNA and modern human DNA showed that Denisovans and Neanderthals were closer to each other than either was to modern humans. Because of that the expectation was that both, Denisovans and Neanderthals, would be equally unrelated to populations such as sub-Saharan Africans that never interbred with Denisovans or Neanderthals. However, this turned out not to be the case as the result showed that sub-Saharan African DNA is more closely related to Neanderthal DNA than to Denisovan DNA. The only viable explanation for this seems to be that Denisovans interbred with an unknown archaic human population, quite different from sub-Saharan African or Neanderthal populations, which in turn never interbred with either of those latter two. Where and when this happened remains a mystery. Because the split between Neanderthals and Denisovans occurred after their common ancestor *Homo heidelbergensis* left Africa, this poses a real conundrum. This so-called archaic ghost population which interbred with Densiovans, must have migrated into Eurasia. It seems quite unlikely that some Denisovans migrated back to Africa to intermix with an unknown archaic population only to leave again and somehow spread the newly acquired gene diversity among the Denisovan population that had remained in Eurasia. Therefore, there must have been another out of Africa migration and in this case, of an archaic human species about which we know nothing else yet. The only indication we have for this happening has been hiding in Denisovan DNA. Now this is certainly something that scientists would not have been able to infer from the interpretation of human fossils only. This really was only possible through modern genetics applied to the exploration of the human family tree. Maybe with some more future fossil finds, DNA analysis will also help us shed some light on this archaic human population that left its genetic markers in the Densiovan DNA.

More recently, it has become increasingly clear that there is a problem with assigning *Homo heidelbergensis* as the common ancestor of Neanderthals and modern humans. In 2016, scientists succeeded in analyzing the DNA of a 430,000-year-old human fossil from the Sima de los Huesos fossil site in the Atapuerca Mountains of Spain [55]. The fossil is believed to belong to an ancestor of the Neanderthals and what scientists can read from its DNA analysis is that this Neanderthal ancestor was more closely related to Neanderthals than to Denisovans. This means that the latter two species must have split from each other before 430,000 years ago. However, that is not all of it. Importantly, this DNA analysis firmed up the estimate for when Neanderthals and our modern human ancestors may have split from a common ancestor to sometime between 765,000 and 550,000 thousand years. Now, the oldest known *Homo heidelbergensis* fossil dates to around 600,000 years ago. That there is no earlier fossil record of *Homo heidelbergensis* is difficult to reconcile with this species being the common ancestor from whom Neanderthals and our modern human ancestors split sometime between 765,000 and 550,000 years ago. This has made it more likely that *Homo heidelbergensis*, the ancestral species of Neanderthals and Denisovans, and our human ancestors split much earlier from another ancestral species. Based on the current fossil record, this other ancestral species could have been either *Homo erectus ergaster* or *Homo antecessor*.

Homo neanderthalensis
La Ferrassie fossil

As pointed out above, the latter could well have been an earlier form of *Homo heidelbergensis* and as such also a link between *Homo erectus ergaster* and *Homo heidelbergensis*. Our picture of how modern humans may have evolved is already much clearer and grounded in a lot more solid evidence than it was only a couple of decades ago. Unearthing the story of human evolution from fossil records is a work in progress. We need to remember that for *Homo antecessor* the fossil record we currently have to date is still quite incomplete. There is so much we do not know yet and so much more to learn about human evolution. We need to resist the temptation to construct our definitive lineage out of the existing fossil record. The latter remains rather incomplete and despite all our efforts, it may not yet support our desire to discover our true family tree. For now, the picture that emerges is that of an ancestral species or early form of *Homo heidelbergensis* becoming the ancestor of modern humans and *Homo heidelbergensis* itself, the latter eventually splitting into Neanderthals and Denisovans. In this picture, the split between the lineage evolving into modern humans and *Homo heidelbergensis* happened in Africa, where exactly we do not know and neither precisely when but scientists believe it must have happened more than 600,000 years ago. Whereas *Homo heidelbergensis* in turn split into Neanderthals and Denisovans in Eurasia sometime before 430,000 years ago after *Homo heidelbergensis* populations had migrated there. After the ancestral Neanderthal and Denisovan populations separated, the Denisovans encountered sometime later an unknown archaic human population, which had left Africa much earlier. Interbreeding with this archaic population then resulted in the Denisovans becoming genetically more different from sub-Saharan Africans than Neanderthals are. Denisovans have not been designated a separate species because they did also interbreed with modern humans and even more so than Neanderthals did. In a strict sense, both species are subgroups of *Homo sapiens* as they both interbred with our species. In the scenario just outlined, *Homo heidelbergensis* and the human lineage eventually leading to modern humans split from a common ancestor thought to be most likely an earlier form of *Homo heidelbergensis*; unless the latter is indeed *Homo antecessor* we do not have fossil records of this species yet. How can we know when all of this could have happened without the supporting fossil records? Well, it essentially is the molecular clock that helps us here as genetic differences between human species grow and accumulate over time. It is on such assumptions that the above hypothesis seems to be reasonable, and quite a number of paleoanthropologists believe it may likely be correct. Hence, from what we know today, the split of an early *Homo heidelbergensis* species into the "proper" *Homo heidelbergensis* species and the ancestral lineage of modern humans happened likely well before 600,000 years ago. The oldest modern human fossils found to data are around 300,000 years old. That leaves about the same time for the ancestral modern humans to evolve into the earliest form of *Homo sapiens*, the modern humans who would eventually migrate out of Africa and as they moved into Eurasia interbreed with both, Neanderthals and Denisovans. Until recently, we thought modern humans migrated out of Africa sometime between 75,000 to 50,000 years ago. Then, a reconstruction of *Homo sapiens* skull fragments from a 1970's Greek fossil site dated to ∼210,000 years ago pushed the first appearance of modern humans in Eurasia back in time by ∼150,000 years [54]. If the first modern humans indeed moved into Eurasia some 150,000 years earlier, they may have lived alongside Neanderthals or Denisovans for much longer than we previously assumed.

The Neanderthals, *Homo neanderthalensis*, or more correctly *Homo sapiens neanderthalensis*, evolved from *Homo heidelbergensis* populations migrating out of Africa. Its fossils have been found in Europe and the Near East, Southwestern Asia and in parts of Central Asia. Humans descended from European and Asian populations carry a small portion of Neanderthal DNA. When modern humans started to migrate out of Africa they interbred with Neanderthal populations whom they may have encountered as early as some 200,000 years ago but certainly around 50,000 years ago in the Near East as we know from ancient DNA analysis. The genetic heritage that modern humans received from Neanderthals may have been small but it could have helped them to adapt more quickly to the colder climates of Eurasia. Neanderthal populations were always small and there may never have been many more than 100,000 individuals throughout the regions of Eurasia which they inhabited. The combined challenges of environmental pressure, increased competition, and as it looks like, natural catastrophes in the form of large volcanic eruptions, must have been more than the gene pool of such a small and scattered population could cope with. Just after 30,000 years ago, practically all traces of Neanderthals vanish, except for the genetic heritage that a significant portion of the human population still carries today. Modern humans leaving Africa some 200,000 years ago would have given them much more time to interact with their Neanderthal and Denisovan cousins, both of whom they would have encountered in their migrations into Eurasia. However, the current DNA evidence traces the interbreeding of modern humans and Neanderthals to somewhere in the Near East around 50,000 years ago. There is no evidence that this happened earlier which could have a number of reasons. One of them could be that only very few modern humans left Africa some 200,000 years ago, none of whom mixed with Neanderthals or Denisovans and we were just lucky enough to find the fossils of one of them in Greece. We will have to wait and see if future fossil finds can substantiate a more widespread presence of modern humans in Eurasia some 200,000 years ago.

For a long time, Neanderthals were seen as the proverbial brutish Stone Age people. This long-held public perception, only corrected over the last few decades, says much more about us than the Neanderthals. Somehow, we, the members of the species *Homo sapiens sapiens*, seem to have this terrible urge to justify our own uniqueness and we commonly achieve this by denigrating those who are not like us. Maybe it is a sign of guilt, maybe not. It is an affliction, from which our modern societies still suffer much; in the darkest days of our history, it has led to the slaughter of millions. The inevitable consequence of demeaning others because they are different always has been murder, extermination, and genocide. As the above explanations suggest, this may have been different when modern humans and Neanderthals encountered each other. At least we can strongly wish that it were so. As for the Neanderthal's, the supposedly brutish beasts, their brains were bigger than ours with sizes ranging from ∼1,300 - 1,600 cm^3. This does not indicate automatically that they were smarter than we are – after all, we are still here, and they are not – but it should make us think. Neanderthal brains were different from ours not just in size but also in shape. Since we could not argue that a smaller brain made them less intelligent than our modern human ancestors, the argument now seems to be about the relative sizes of specific parts of the human brain, which seemingly differ between Neanderthals and modern humans. Given the little we know about how our brains really function and what we know today about human brain plasticity it seems rather disingenuous to find fault with Neanderthal brains because some of their brain areas were sized somewhat differently. Have we not learned anything? Maybe there will be a time when we understand our own brains to the extent that such statements would actually be more than the pretense of understanding, but we are not there yet. Without an actual working Neanderthal brain to compare to, how can we assert that we know where their brain functions may have differed from ours? Heavier use of certain cognitive abilities can result in relative size differentiation between brain areas. If we lose some brain areas due to an accident for example, we can to some extent train our brains to recover lost functions by using other parts of the brain. This capacity of the brain to reroute functions is what scientists call brain plasticity. We can look at Neanderthal brains as an evolutionary size differentiation, so-to-say the species equivalent of brain plasticity. Neanderthal brain areas were differently sized and shaped from ours. But that tells us little about what their brains were capable of. There is nothing we can say with any certainty about how their cognitive capabilities may have differed from ours, either way. Just imagine for a moment what we could learn about ourselves if another human species had been successful next to ours. How much better should we be able to understand our own evolution? The more we learn about Neanderthals the more we see that they were not so different from us. They could have been our neighbors today as they once were. Maybe looking a little bit different from us but then, maybe more neighborly than many of us are ourselves. However, we do not know that and never will. It is sad that we are the only human species left on Earth. With one human species having achieved the planetary dominance that we exercise today over life on Earth, it is quite unlikely that another large bodied intelligent species would ever have the chance to evolve in the future.

Around the same time as the Neanderthals, the Denisovans, *Homo denisova* or more correctly *Homo sapiens denisova*, also became extinct. Because Denisovans, just like Neanderthals, interbred with modern humans, they must have been still around when modern humans moved into Asia. We know less about the Denisovans than about the Neanderthals. The fossil record we have for the Neanderthals is much richer than the one we have so far for the Denisovans. Like the Neanderthals, the Denisovans left some of their genetic heritage with modern humans when the two species interbred. The fraction of Denisovan DNA that can be found in many Asian populations today is higher than what the Neanderthals left behind. Therefore, while the Neanderthals seemingly left us a richer fossil record, the Densiovans left many modern humans, specifically in Southeast Asia, with a richer genetic heritage. The fact that some Asian populations carry a slightly higher percentage of Denisovan DNA than Europeans carry Neanderthal DNA may point towards the Denisovans maybe managing to hang on a little longer than the Neanderthals. For the sake of simplicity we will assume here that the Denisovans vanished roughly about the same time as the Neanderthals.

Homo sapiens
Skhūl V fossil

In the grander scheme of human evolution it does not matter much if they vanished a little earlier or a little later than the Neanderthals. The genetic legacy Denisovans left behind seems more varied and just as Neanderthal genes may have helped modern humans to adopt to Eurasian climates and environments, something similar seems to have happened with Denisovan genes in Asia. An example for that are the Tibetans who are thought to owe their adaptation to living at high altitude to genetic inheritance from Denisovans. Natives of New Guinea share a particularly high fraction of Densiovan DNA, more than double what modern Europeans have inherited from Neanderthals. The Aboriginal population of Australia also inherited some Denisovan DNA and throughout much of Asia traces of Denisovan DNA can be detected, however, in smaller amounts and at varying degrees. Until recently, the emergence of modern humans, *Homo sapiens*, as a separate species had been tied to a fossil find dated to around 195,000 years ago. However, a *Homo sapiens* fossil found in what is today Morocco pushes the date that *Homo sapiens* entered the stage back to around 300,000 years ago [56,57]. There is still some discussion if these fossils discovered at Jebel Irhoud near the Atlantic coast in modern-day Morocco are indeed *Homo sapiens* fossils or belong to a more primitive form of *Homo sapiens*. However, as there are also indications of earlier forms of *Homo sapiens* evolving in other parts of Africa, it may well be that we have to date back the arrival of *Homo sapiens* to just before 300,000 years ago. It is certainly remarkable to have such an early fossil now for modern humans. But what stands out even more is just where these fossil were found: thousands of kilometers away from all other African fossil sites where archaic *Homo sapiens* fossils had been discovered before. This is another clear indication that the origin of *Homo sapiens* is certainly more complex than previously believed. The large geographical distance between the fossil sites in Morocco and East Africa makes one wonder what we may one day find in-between. It certainly shows that *Homo heidelbergensis* and its Neanderthal and Denisovan descendants not only populated much of Eurasia but that *Homo heidelbergensis* and most likely earlier human populations must also have reached into all corners of their continent of origin, Africa. We should not forget that Africa back then was a quite different place. Where today we find vast stretches of sand deserts in which only few animals can survive, there once was a time when much of that part of Africa was habitable for humans. We already encountered some species whose fossils were found in what today are essentially desert areas but which in the past must have provided a much more welcoming environment to early humans such as *Sahelanthropus tchadensis*, *Orrorin tugenensis*, or *Australopithecus bahrelghazali*. Who knows how many archaic human fossils, belonging to human species of which we have no knowledge, are hidden beneath the sea of sand stretching over much of Northern Africa? We have no idea how many other human populations *Homo heidelbergensis* or its ancestors may have encountered on their inner-Africa migrations. However, if they did encounter other species about which we know nothing yet, they may well have interbred, as human species who are first cousins are seemingly prone to do. Our genetic heritage may be much more colorful than we assume today and maybe the picture of our species now being the only remaining one with all others becoming extinct is plain wrong. Would it not be a much more wonderful picture if it turned out that modern humans became the vessel to preserve and continue the heritage of many different human lineages? As already discussed, the cranial capacity of modern humans is somewhat smaller than that of their Neanderthal cousins, ranging from ~1,100 - 1,580 cm^3. Interestingly, the average brain capacity of today's humans, *Homo sapiens sapiens*, is in turn somewhat smaller than that of *Homo sapiens*, seemingly an effect of domestication; or put differently, as we became more civilized our brains became somewhat smaller. Because *Homo sapiens* is our direct ancestor, we apparently have no problem with them having on average somewhat larger brains than we do because there seems to be little to no discussion about this. Maybe it is just very human to be partial with respect to one's own direct family ancestors when it comes to such matters as perceived intelligence; and maybe if Neanderthals had become the only surviving human species, they would do the same.

When *Homo sapiens* emerged, *Homo erectus* was still around and so may have been other human species. We know that *Homo sapiens* encountered such earlier species when migrating into Europe and Asia and we know that they interbred. The same likely happened also in Africa when two species that could interbreed staid long enough in contact. We are the descendants of *Home sapiens*, sometimes we are referred to as *Homo sapiens sapiens*.[86] However, we are also carrying DNA from earlier humans, humans who either remained in Africa or had moved earlier out of Africa into Eurasia. Human populations outside Africa continued to evolve and eventually the next waves of human species migrating out of Africa encountered these cousins at a later time. Something similar could also have played out in the vastness that is the African continent itself. The story of human evolution is much more varied than we ever thought. New fossil finds continue to broaden our view of human evolution. It is only a decade ago that fossils of a new species were discovered in southern Africa and given its own species designation, *Homo naledi*. The fossils date back to ~336,000 to 236,000 years ago. That makes this a rather young fossil find, placing it close to the emergence of *Homo sapiens*. However, in several respects this small-built species has also some *Australopithecus*-like features. The mixture of *Homo*-like and *Australopithecus*-like features in *Homo naledi* is seen as another indication that species interbreeding may have been a more important evolutionary path than previously assumed.

Another surprising species discovery was *Homo floresiensis*, a species that never existed in Africa, only outside of it. *Homo floresiensis* seemingly was confined to a small geographical area, the Island of Flores in Indonesia. Its fossils date from 100,000 to 60,000 years ago and stone tools that are associated with this species date back to around 190,000 years, with the youngest "only" about 50,000 years old. According to the status of the current fossil record, this recent date indicates that *Homo floresiensis* only disappeared about 10,000 years before modern humans arrived in Asia. While we cannot be so sure in other cases, here for once, modern humans are quite unlikely to have replaced another species. *Homo floresiensis* was of a small build, just under one meter or so, earning the species the nickname of Hobbit, and it is thought to be a separate species offspring of an early human and not descending from modern humans. The *Homo floresiensis* island population derived from an earlier human migration with *Homo erectus* seen as the most likely candidate for now. Once a *Homo erectus* population became isolated on the Island of Flores, so the hypothesis goes, it started to diminish in size due to the well-known phenomenon of island dwarfism. This phenomenon has been observed for many species, mostly larger mammals such as elephants, who when they became stranded on small islands were forced to adapt to their new resource poorer environments by becoming smaller. It is thought that the Hobbit was present on Flores as early as 700,000 years ago and may have been still present there some 50,000 years ago. Intriguingly, specific anatomical features of *Homo floresiensis* skeletons make it look more similar to some of the *Australopithecus* species than to *Homo erectus*. Maybe, just maybe, there has been an earlier migration out of Africa, preceding *Homo erectus* populations moving into Eurasia of which the only indications we have found until today is *Homo floresiensis*. For now, this is only speculation, but it highlights how a single fossil find could overturn long held views about such important milestones as the first migration of a human species out of Africa. Until we have more data, *Homo floresiensis* remains the descendant of an early form of *Homo erectus* rather than being the offspring of more archaic human lineages such as an *Australopithecus* species leaving Africa for which there is no evidence yet.

Fig. 3.13 summarizes the timeline of human species evolution as discussed here. By no means is this graph definitive in any way but rather, it is a snapshot of what our extended family tree currently looks like. It needs to be emphasized that the species concept such graphs use is tenuous at best as we know that several species have interbred in the past, including our own lineage. However, the species concept, and how we understand it today is the best we have. This graph also gives an indication of human brain size evolution. While an interesting metric, such data has its own problematic. For some human species, the fossils do not include complete skulls or no skull at all so in these cases we just cannot measure the respective cranial capacities. Brain size also correlates with physical size and smaller bodies usually come with smaller brains. Often this correlation is misinterpreted, and many times in the past, it has been misused to justify racial biases or discriminatory practices. Using brain size comparisons to assess potential intellectual prowess is a treacherous path best avoided. Just remember, it is often the smaller people in our lives who turn out to be the smarter ones, including of course our children. Staying close to home, none of us would ever dream of associating more or less intelligence with our spouses because they happen to be bigger or smaller than we are. In several cases the brain size data in fig. 3.13 represents limited sampling data, mostly so for early human species; sexual dimorphism is more pronounced for some species than for others. For the more recent human lineages, the number range shown mostly reflects dimorphism, like for our species and the Neanderthals. For earlier humans such as *Homo erectus* the number range shown reflects the evolution of larger brains over the long time this species was around rather as well as sexual dimorphism, both of which are present in the *Homo erectus* data.

As to what has been driving the evolution of our brains, we still are very much in the dark. It looks like that for much of human evolution, from whenever sometime between seven and five million years ago our ancestors split from the chimpanzee line until some 2.4 million years ago when *Homo habilis* enters the scene, brain size changed little.

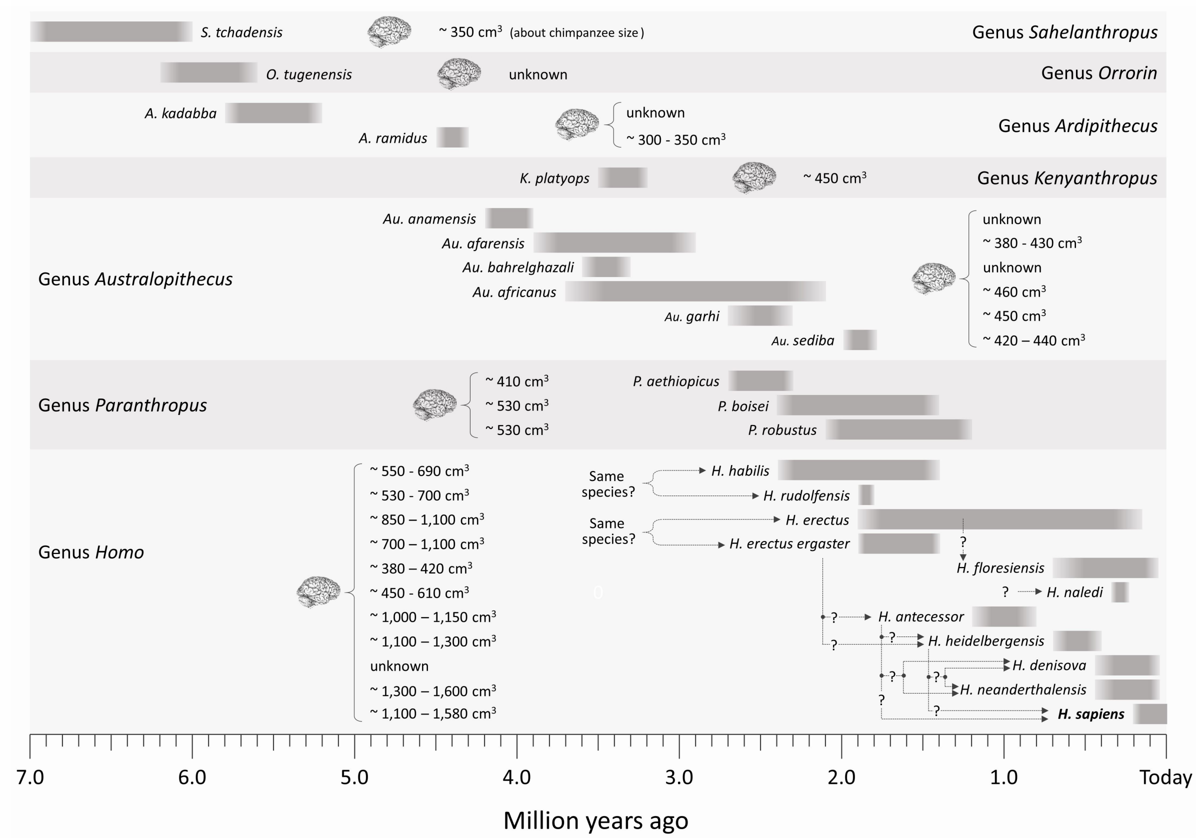


Figure 3.13: Our current picture of the human evolution timeline, known human lineages, and their respective species members. For the genus Homo, some of the hypothetical relationships discussed are tentatively indicated.

If we leave out the somewhat larger brains of *Paranthropus boisei* and *Paranthropus robustus*, brain sizes essentially remained within the range we know for chimpanzees. What drove the growth in brain size we see with *Homo habilis* we are not sure yet. For some time, adaptation to a significant climate change seemed the likely explanation to have triggered this evolution. However, today, scientists believe it is more the variability of the climate conditions themselves, which forced so-to-say *Homo habilis* to become smarter. What then drove the increased brain capacity of *Homo erectus* may be a different story as between *Homo habilis* and *Homo erectus* something else took place; the transition from small ape-like creatures to humans with practically modern skeletons. Humans discovering how to cook likely best explains a much-reduced gut relative to body size that characterizes *Homo erectus*. While many human species before *Homo erectus* were bipedal, it looks like *Homo erectus* is the first species built to run. Combined with another physical adaptation, the loss of most body hair, this is believed to have made *Homo erectus* a very efficient persistence hunter, running his prey to exhaustion for an easy kill. More meat in their diets in turn would have supported the evolution of larger brains requiring more energy. The other side of the argument is that this could also provide an explanation for the long stagnation in brain size in early human evolution.

Early humans may just not have been able to get the calories needed to support the evolution of larger brains and bodies. Maybe for similar reasons *Paranthropus boisei* and *Paranthropus robustus* had somewhat larger brains because from their dental records we know that their diet likely was different from other early humans such as the members of the *Australopithecus* genus. Much of what we believe is driving the evolution of the human brain is still educated speculation. We just do not know enough about the brain evolution of the human species. Given that brains do not preserve in the fossil record and brain casts of ancient humans can only tell us so much, it is unlikely we ever will know the full story.

Clearly, for some human lineages scientists have not made their minds up yet of how to best classify them. New fossil finds or discoveries combined with continuously improving gene analysis methods will most likely change our perspective of the human family tree. However, for now fig. 3.13 shows in a simplified way what we know about the time line of human evolution. Only a century ago, we practically had no knowledge of how humans evolved or how many ancestral lines there were to which we could trace our own roots. Compared to that, our current understanding, however incomplete it still is, already is a giant step forward; and hopefully many more such steps will follow.

The migrations of our ancestors out of Africa make for fascinating stories by themselves. Just a brief look into this topic will reveal to anyone that it is nothing less than amazing what modern genetics allows us to understand now. As we now know, our own direct ancestors were not the first ones to migrate out of Africa, they were just the most recent ones to do so and when they did so the last time between 40,000 to 50,000 years ago, they either replaced or absorbed other species they encountered on their journey out of Africa. There has been much debate about the various out of Africa migration scenarios and even more controversy between out of Africa scenarios and multiregional human evolution scenarios. In the early days of paleoanthropology, multiregional development scenarios seemed like a natural explanation for the fossil record. Until around 1920, archaic human fossil finds were limited to Asia and Europe. Prominently featuring among them were the Peking Man and the Java Man in Asia, both later apprised as representatives of the *Homo erectus* species; later in Europe, we have the *Homo heidelbergensis* and Neanderthal fossil discoveries. However, all that started to change with the 1924 fossil discovery of what came to be known as the Taung Child, the first discovery of *Australopithecus africanus* fossil remains in South Africa. Eventually, the increasing number of archaic human fossil finds would reveal Africa as the cradle of the human species.

Multiregional evolutionary scenarios in their hard form maintained, to put it simply, that once *Homo erectus* populations had left Africa some 1.8 million years ago, the respective regional *Homo erectus* populations evolved to eventually become the modern human populations we find in those regions today. This view was prone to racial biases, to an extent even invited them, inadvertently or not. The human fossil record we have today by itself sufficiently proofs this hard version of the multiregional evolution hypothesis wrong. And modern DNA analysis certainly has proofed it wrong.

The soft version of multiregional evolution, again put simply, allows for successive migrations out of Africa where the newly migrating people would intermix with the humans that had already evolved in the various regions from earlier out of Africa migrations. While this is a little closer to what may have happened, it still is not what the genetic evidence tells us. Modern human populations outside of Africa show between themselves significantly less genetic diversity than what we find today among populations that remained in Africa. This fact leaves only one sensible interpretation: the genetic make-up of human populations outside of Africa reflects their genetic inheritance from a rather small modern human population that left Africa in recent times. Scientists refer to this as the bottleneck effect where only a few thousand modern humans, some say even less, became the genetic stock from which all of modern humanity outside of Africa derives. That is if we disregard the migrations of our modern societies, which increasingly mix populations across all continents. While a substantial number of human populations outside of Africa also carry some inheritance from earlier human species which their ancestors encountered as they moved into Eurasia some 40,000 – 50,000 years ago – such as from the Neanderthals and Denisovans – these contributions are small. As we have seen above with *Homo floresiensis*, we may not even have the complete picture of early human migrations out of Africa yet. Therefore, we may not really know if the migration of modern humans out of Africa is the third or fourth migration of humans out of Africa; and there is always the chance that some human populations may also have migrated back into Africa during the million years of human evolution.

The "recent out of Africa" approach championed by the British anthropologist Chris Stringer (born 1947) seems to be the most sensible way to navigate the out of Africa versus multiregional evolution controversy. First, separating the most recent out of Africa migration from what came before recognizes the fact that this most recent migration defines the diversity of the genetic make-up of human populations outside of Africa today. Second, it also focuses the discussion on the dynamics between modern humans leaving Africa and the earlier human populations they encountered in Eurasia. Finally, it is an open concept as our understanding of the most recent migration out of Africa continuous to be modulated by new discoveries.

Today we can precisely trace back our ancestral roots to what are called the Mitrochondrial Eve and the Y-chromosomal Adam to around 195,000 years ago in East Africa. Mitrochondrial Eve is the most recent female ancestor from whom all living humans inherited DNA in an unbroken chain purely through their matrilineal heritage. Even though she is called Eve, she should not be looked at in any way as a single ancestor of all humans as it is rather chance events that somehow led to DNA fragments of this Eve to survive until modern times. And the very same is true for Y-chromosomal Adam whose DNA fragments can still be found in all living male humans today. The oldest *Homo sapiens* fossils predate Mitrochondrial Eve and Y-chromosomal Adam by more than 100,000 years. However, unlike the DNA fragments inherited from Mitrochondrial Eve and Y-chromosomal Adam, this earlier genetic heritage is not uniformly shared anymore by all humans. Even though Mitrochondrial Eve and Y-chromosomal Adam are not single direct ancestors, their DNA fragments in all of us clearly show that we are all related. While some of us have suspected this long ago, this is not good news for those who still are boxing themselves into racial categories. Race, for all we know today, is not a biological category in any respect. It reflects minor differences in appearance and climate adaptation due to long-term impacts of environments and not major genetic differences. Race is a social construct that reflects back the kind of ethics and moral codes various human societies value but tells us little to nothing about biology.

With that, we come to the end of our brief journey through human evolution as it has been and continues to be translated for us by countless researchers dedicated to the understanding of the evidence provided primarily by fossil and genetic records. Before turning the page, a word of caution is required. The picture given here is a snapshot in time. The story of human evolution is not static as it continues to change with every new discovery. Certainly, a species such as *Homo erectus* that was around for more than 1.5 million years was not static but continued to evolve and adapt. The late representatives of the species *Homo erectus* were quite different from the very first members of this species. The species concept, as used here and in many other places telling the story of evolution, seems static. However, it is all but that, for *Homo erectus* as well as for the other species characters that played their roles on the stage of human evolution. Most importantly, we may be looking, or better quite certainly, we are looking only at a much-reduced character set of this ancient drama we call human evolution. The original character roll for this play was likely much longer and we can only hope that over time some of the missing characters will reveal themselves. For what it is worth, we are only at the beginning of our quest to understand how we modern humans came to be who we are today. What we know today about human evolution is so much more than only a few decades ago. Our efforts to explore human evolution not tainted anymore by delusions of racial superiority or other inherent racial biases have become a true science in the best sense. It helps us understand our human species heritage, who we are, and maybe even what humans may become in the distant future. Can anyone imagine how different our descendants will be from us in a few million years given how different we are today from our most distant ancestors we know? There remains much to discover, many blank spaces in the story of our human evolution to be filled in, and the tools of modern science will help us to do just that. However, finding new human fossils will remain important. Behind the quest for new fossil discoveries, there still is likely a good part of "gold rush fever" and that may never quite fade. But that is just as well. What can be more exciting than to discover the first fossilized remains of some of our relatives in the human family tree, or for that matter, to be able to analyze their DNA and be the first to understand our common heritage?

The Human Endeavor

Last Rays

If nothing else, the previous chapters should have made it evident just how much our knowledge about the world we live in and where we come from has changed over the past few hundred years. Some may date the onset of this deeper understanding to the beginning of what we call the industrial revolution but that would not be correct. The scientific revolution that ushered in our modern age started earlier as Europe transitioned out of the Middle Ages into the Renaissance and we would be wrong to assume a single beginning at all. Human societies of the past have often progressed amazingly towards what to us would seem like enlightened, if not modern views of the world. However, always only in certain areas, never in a broadly encompassing way including all aspects of the human experience. Often when we read about these past human societies, we are looking at the results of cherry picking. Seldom, if ever, have human societies only progressed towards what we frequently endorse as the selective perception of our heritage. More often than not, there are the dark sides where practically we have to force ourselves to stare into the bottomless abyss of the evils, we humans are capable of. Our understandable natural reaction is to recoil in disgust and look away. However, as much as we would like to reject it, this also is part of our heritage. We must acknowledge that, as difficult as it is. Inevitably, our perspectives on humanity and its societies are often fragmentary, regional, and frequently biased. At times in our history, specific regions were seemingly able to control much of the planet; European colonialism and imperialism certainly comes to mind here. Ultimately, human cultures and traditions are deep seated and eventually they reassert themselves. Slowly but surely, we are moving towards a less Western Culture dominated perspective on history and on humanities achievements as well as its defeats since pre-historic times. Often, this struggle for progress is difficult at best. Mostly, because our own personal limitations prevent us from taking advantage of all available sources, regardless from which corner of the world they originate and in which language they were passed down to us. Many of us keenly feel this restriction, as does yours truly. Then there is of course also the fact that sometimes there are just no good sources at all. At which point we gaze over blanks in the records and must work with accounts less than trustworthy as best we can. How we as a species came to develop what we know as civilization is something that we do not fully understand yet. We all do have our own preconceptions of what defines civilization, which makes such an enterprise even more challenging. At the peril of falling into this trap, we still do have to look, at least to some extent, into some of the assured historical markers so we can put the discussion to follow into a temporal and geographical context.

Our foundation for building a lasting future for life on Earth must be a profound understanding of the world we live in as it manifests itself, in the small and the large. From the teeming life in a small puddle of water on Earth to the cosmic drama playing out in the vast spaces of our universe. We must understand our role in the evolution of life on Earth including our own evolution and acknowledge our responsibilities as the stewards of life on Earth or else there will be no future for life beyond Earth. At least not for the life on Earth as we know it, once our home planet becomes inhabitable, as eventually it surely will. However, our remaining time on Earth may even be shorter than that and we may go extinct well before Earth becomes inhabitable. There have been several mass-extinction events, each one wiping out much of life on Earth as it once existed. Ironically, we have to be grateful for the last such extinction event because without it, dinosaurs might still rule the Earth and mammals may have stayed small, not much bigger than shrews. It is not only external extinction events, which we must fear. We have become quite capable of destroying ourselves along with almost all other life on Earth. If some such event, externally caused or probably something we will bring on ourselves, should wipe out life on Earth over the next few million, tens of millions, or even a couple of a hundred million years, there will still be enough time for new life to thrive again before Earth becomes inhabitable. For what it is worth, maybe a kinder and smarter intelligent species will evolve on Earth, avoiding our mistakes. It is our singular responsibility to do what is required for life on Earth to thrive until we can find a way to spread the life of Earth among the stars. Not to say other life has not already evolved somewhere else in the vastness of our cosmos, it most certainly has. However, we can count on life on Earth to be a rather exceptional story. While we certainly should expect the existence of other intelligent lifeforms somewhere in the vastness of our cosmos, it is highly likely that no other human beings just like us are anywhere out there. Earth has lost its position at the center of our universe quite some time ago, so has the Sun and our solar system. Our Earth and Sun are not unique, there are more stars in the universe than grains of sands on all of Earth's beaches and many of them may have planets not to dissimilar from Earth orbiting them. However, for all we know, life as it has evolved on Earth may be unique. Therefore, if we ever should be so lucky to encounter alien civilizations – again, the chances for that happening are close to zero - it would certainly be a most interesting encounter, and hopefully a peaceful one. In what follows we will first take a brief walk through human pre-history, taking tool development as an example to illustrate the complexity of our interwoven human biological and culture evolution.[1] This brief historical narrative will end with the introduction of agriculture. After that, we will set out to explore some of the really hard questions regarding us as a species where there may not be any good answers at all but where it is so important that we ask those question if for nothing else but to realize how tenuous our human condition is.

The Dawn of Humankind

There is still much debate around when humans developed the capability to communicate through speech and when and how tool making started. Some claim both of these paradigm-changing events occurred simultaneously, others say tool making came much earlier and communication by speech may not have developed until some hundred thousand years ago. Determining when and how human speech evolved is one of the grand challenges to solve in understanding human evolution and we never may get a satisfying answer. For tool making, we can hope that new archeological finds will tell us when that happened for the first time; or at least close to that. Over the past decades, we have learned much about tool use in the animal kingdom, and we keep being surprised by what some animals are able to do using tools of their own. While tool use is now well established for a number of animal species and not just among primates, tool making is a different story. Finally, tool making not just for oneself but also for other members of an animal group adds a completely new quality to this discussion. It is this last kind of tool making that we should have in mind when we think about what really differentiates humans from primates and other animals.

Understanding the origins of speech and tool making are but two, though admittedly two of the most important aspects of understanding human evolution. As we acknowledge this, we are stepping over an invisible line as we move from talking about biological evolution into the domain of culture evolution. When we talk about human evolution, we need to be very clear that we have a shared evolutionary ancestry with other lifeforms that continues with us as we keep evolving. However, human species evolution is also unique, as no other lifeforms have developed in any way even remotely similar to us, ever, at least on this planet. The challenges we face with understanding human culture evolution are manifold. First, there are the discontinuities we face in both the human archeological fossil records and in the artifact records that we scrutinize for signs of human culture evolution. The picture we have of human species evolution, even though much richer than only a few generations ago, unfortunately, is still quite spotty. We do not have the records for a human species continuum we could match up against the archeological records for early human tool use or for other indications of human culture evolution. The artifact record for human culture evolution we have so far is an incomplete patchwork at best. In essence, we have no clear picture as to which species should receive credit for what in early human culture evolution. Just consider our own species, which may go back some 200,000 years or maybe even further than that as recent paleoanthropological studies in the Near East as well as a reexamination of human skull fossils found in the late 1970's in Greece indicate. We now have to assume that the first *Homo sapiens* arrived in Eurasia some 210,000 years ago. This is more than 150,000 years earlier than what the oldest confirmed paleoanthropological *Homo sapiens* record previously supported for this species in Eurasia [54,58]. What this shows us is that any conclusions we draw about human culture evolution on spotty fossil and artifact records will often be tenuous at best and more often than not, we have to correct our conclusions frequently with new fossil or artifact finds. For the more recent past of human evolution, say for the last few hundred thousand years and specifically for our own species, we can hope that this will produce an ever more consistent and less spotty picture of human culture evolution. Not so for the much longer time span stretching back millions of years when humans first started to use and make tools. Human culture evolution should be perceived as a complex process carried by multiple human species, some of which we may not even know about today. In the million-year-long relay of human culture evolution, our human species merely is the most recent and for now last carrier of the torch.

Another challenge we face with understanding human culture evolution is that the parts that usually do not preserve in fossil records are the ones that are most important, the ones we could learn from most. Brains are not preserved and the same is true for vocal cords. Therefore, we have next to no evolutionary records of those human body parts so critical to our evolution as a species. With respect to human fossil and archeological data we only have a rather incomplete record to provide us for example indications for the onset of speech. Not surprisingly, there are many different estimates for when humans may have evolved the capability of speech. Until quite recently there were still discussions whether some of our cousins such as the Neanderthals even had the capability of speech at all. Today we know the answer to this question. Neanderthals did have the capability of speech. However, the controversial debate around this topic should be a warning to us. The voices of our Neanderthal cousins are silent now for some thirty thousand years. We as the only surviving human species may have an inadvertently biased view regarding the capabilities of other human species whose voices we cannot hear anymore.

The speed at which humans have evolved over what is in evolutionary terms a very short time period is unprecedented. No other lifeforms, not on planet Earth, have ever evolved to a point where they would be able to ponder the origins of the universe and can reflect on the origins of life as their species builds ever more complex and technologically advanced civilizations. How could this happen, how did this happen? Can we expect biological evolution by itself to eventually provide the complete answer or is there something else we are missing? Why has something like human evolution not happened earlier and more often than it did? We need to be clear-eyed that culture evolution is not something special and just part of biological evolution like anything else in nature. We are and continue to be part of the natural world and culture evolution is part of the broader biological evolution, it just so happens to kick-in only once certain biological pre-conditions are met. What makes it seem so exceptional is that culture evolution builds on itself and rather than taking place over hundreds of millions of years it can happen in a few million years or even in hundreds of thousands of years. Culture evolution follows a trajectory that looks more like an exponential progression and seemingly leaves biological evolution behind, but it does not. Culture evolution imprints on biological evolution and not the other way around. For those species lucky enough to enjoy it, culture evolution may make it seem like they are different, the chosen ones so-to-say. However, we are still from this Earth and the hubris we can see in our past perceptions of nature and our place in it should be a stern warning to us.

Understanding human culture evolution continues to be a work in progress and we should not pretend that we have all the answers yet. Far from it. There are so many things that we do not know yet about human culture evolution as we only have started asking those questions recently; and in an interesting twist, as our culture evolution continues, we may eventually be able to find more satisfying answers to what actually underlies it.

A Brief Walk Through Human Pre-history

With human pre-history, we generally reference the long stretch of time from when the first humans started to use stone tools until the time they became capable of recording their own history in writing. This bookmarks a time window opening up roughly 3.4 million years ago and closing about 5,400 years ago, when we have evidence for the first human writing system appearing in Mesopotamia. For almost all of human culture evolution we have no written records and what we know about this time span called pre-history comes exclusively from our interpretation of paleoanthropological and archeological finds. The more than three million years of human pre-history are not characterized in any way by a unique theme other than us still knowing very little about most of it. We can assume with certainty that beneath this single label for such a long period in our species culture evolution, our pre-history hides an almost unimaginable diversity of human evolution. Naturally, archeologists and anthropologists have thought to structure this long stretch of human pre-history once they began to realize that the human story reaches much further back in time than the Bible would have made us believe. The use of the term pre-history to refer to primitive human societies before the invention of writing systems did start in Darwin's time. It is only some two hundred years ago when scholars began to realize that human history reached much further back than biblical times. Even then, they were certainly not thinking in terms of millions of years but more like in tens of thousands of years at best. It was not until Darwin had put forward his hypothesis of humans and apes sharing a common ancestor that it became clear that human history must stretch back in time much further.

Dividing human history into ages based on tool materials has its roots in antiquity. Historically, western scholars have viewed the time before humans began to record and study their history as belonging to three distinct ages: the Stone Age, the Bronze Age and the Iron Age. The Danish scholar Christian Jürgensen Thomsen (1788 – 1865) pioneered the use of such a scheme to provide a chronology for dating artifacts. Widely adopted and refined by nineteenth-century scholars it proved useful throughout the geographies European scholars focused on at the time. But its application to characterize human history in other parts of the world can be problematic. For example, human pre-history in the Near East ends 5,400 years ago when we find the first writing system. In the Near East the end of human pre-history and the beginning of history almost coincide with the end of the Stone Age and the beginning of the Bronze Age 5,300 years ago; the beginning of the Bronze Age and the first records of writing systems are practically contemporary. That is different in other parts of the world where writing systems come later, like in Asia where this occurs towards the end of the Bronze Age there. The three-age system continues to be useful, but it has its limitations. This is most evident for the middle and younger Stone Ages and the Bronze Age as there are large overlaps in time with one stage continuing longer in some areas while in others the next stage was already established. The three-age system also makes no sense for cultures that had not developed metal working technologies by the time they made contact with cultures that already had done so. The simple fact is that there is no strictly synchronized linear development of human cultures throughout pre-history. This is most evident for more recent times, when we see some cultures developing metal working technologies while others in more recluse areas of our planet would not develop them until the twentieth century, when it was much more likely that they were exposed to these technologies by our modern societies reaching into every corner of the Earth. We have to be careful not to become judgmental here, as a few thousand years are really nothing when we look at the millions of years humanity spent in the Stone Age. It just so happened that some cultures developed metal working technologies earlier, which as it turned out would greatly accelerate their capabilities to develop other new technologies and, in a sense, helped them leapfrog cultures that had not made this transition yet. During the Stone Age, various human cultures also moved at different speeds. However, the relative advantage gained by being first to make the next step throughout that time was comparatively smaller than the one gained by moving for example from the Stone Age to the Bronze Age.

A linear progression, taking place in a coordinated way throughout human history, synchronizing human cultures in their developments across geographies, is an illusion. We need to be aware of that when we look at chronologies of human pre-history and realize that there always was a spread of human cultures present. An earlier time of human pre-historical development could easily be contemporary with a somewhat later one if the respective cultures were geographically too distant to interact. This seems to have been the case specifically for the later Stone Age. Only with the Bronze Age and Iron Age do we see the development of human cultural and technology development becoming more synchronized as humans increasingly were capable of bridging geographical separation on a regional level. Such was the case for much of the lands around the Mediterranean and Asia Minor, China and parts of Southeast Asia, as well as large regions in the Americas, mostly in Middle- and South America. Motivated by the above discussion we will restrict our brief walk through pre-history to the Stone Age. The Stone Age is divided in three main sections, the Paleolithic or Old Stone Age, the Mesolithic or Middle Stone Age, and the Neolithic or New Stone Age.[2] Fig. 4.1 shows the timetable for the Stone Age alongside the major tool cultures and the several human species that existed during the more than three million years the Stone Age lasted. Taking the first indirect evidence paleoanthropology and archeology currently can provide for any human tool use as the beginning of human pre-history, the graph starts at 3.4 million years ago. Using this starting point, the first period of the Stone Age, the Lower Paleolithic lasted for about 3.1 million years from 3.4 million years ago to roughly 300,000 years ago. The Middle Paleolithic starting at around 300,000 years ago was much shorter and lasted for about 270,000 – 260,000 years, until 40,000 to 30,000 years ago. As can be seen in fig. 4.1, the end of the Middle Paleolithic overlaps with the Upper Paleolithic, the last period of the Paleolithic, starting between 50,000 and 40,000 years ago, depending on the archeological evidence found in various geographies.

Human evolution, biological and cultural, is not a straight line. For most of human pre-history, we see more than one human species present. As we have seen in the last chapter, this is a rather recent discovery. More than one species is the rule for most animals, *Homo sapiens* as the only surviving human species of its kind left is the exception. Human population densities, if one can use this term in a meaningful way for pre-historical times at all, were very low. If some human populations migrated off to distant lands, contact may have been lost for a very long time and maybe never regained. Low population densities and several human species at any given time seem to be typical for human pre-history.

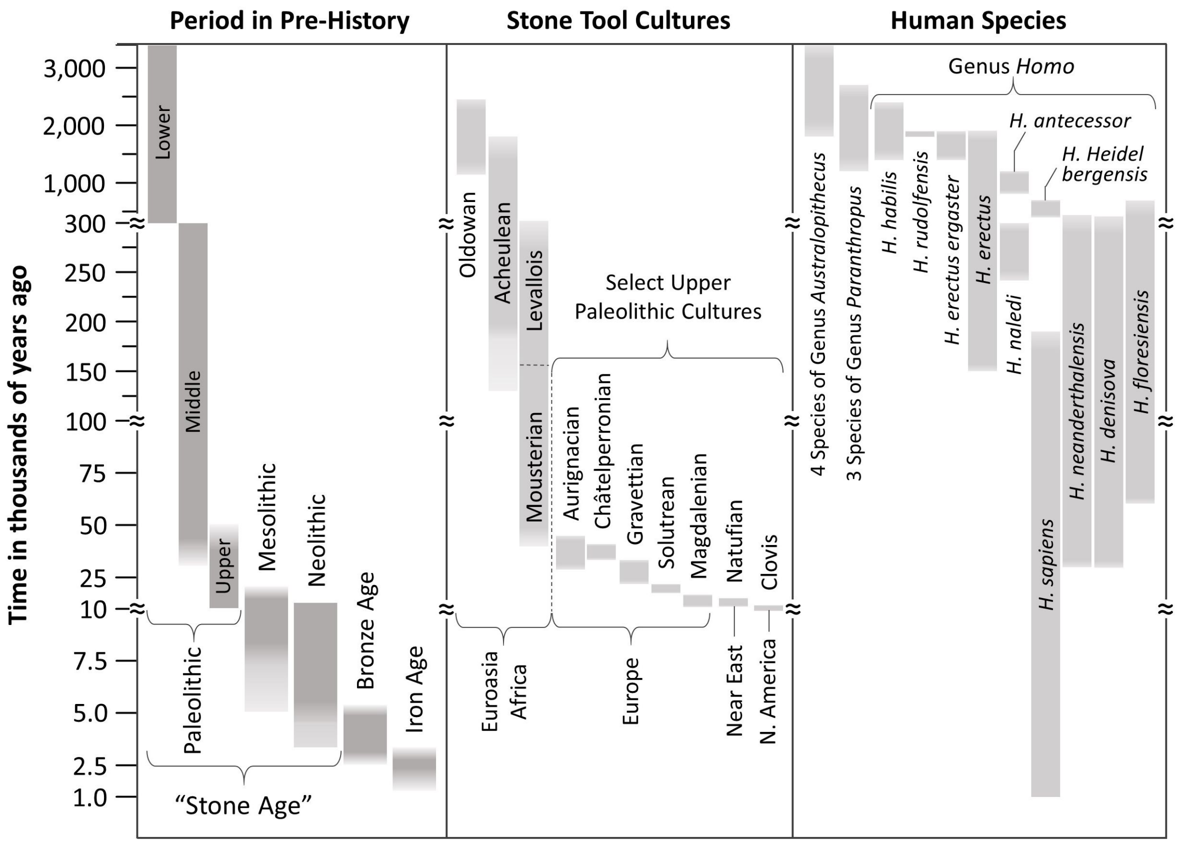


Figure 4.1: The periods of human pre-history with some of the associated stone tool cultures shown alongside a simplified perspective on concurrent human species walking the Earth during pre-historical times. Pre-history first ends in the Near East 5,400 BCE, just before the Bronze Age begins in this region. Please note that the vertical time axis is broken into four segments, each with its own linear scale showing increased resolution for recent times.

This in turn should make us expect the archeological and paleoanthropological record to reveal a richly differentiated development of tool cultures, geographically and in time. We can fairly assume that all human species and sub-species – and there may have been many more than we can detect now looking back into the distant past – were capable of tool making. The tool cultures listed in fig. 4.1 represent a selected fraction of the varied tool cultures that likely existed at different times. Specifically so for the Upper Paleolithic, however less so for much of the Middle- and Lower Paleolithic where all we can distinguish are three major tool cultures: the Oldowan, the Acheulean, and the Levallois-Mousterian tool cultures. All three are named after the respective locations where tools belonging to what later became these recognized tool traditions were first found.[3] In Europe, the Neanderthals were the carriers of the Mousterian tool culture, while in North Africa modern humans were the Mousterian type toolmakers. That we find two human species contributing to the same tool culture should give us pause. It may seem intuitive to associate one given tool culture with one specific human species as its carrier. However, at least for the Mousterian tool culture that seems not to be the case. Therefore, we should not exclude that more than one human species may also have contributed to earlier tool cultures such as the Oldowan and Acheulean.

The oldest stone tools ever found date back to about 3.3 million years ago, some 100,000 years after the oldest dated indirect evidence for tool use [52,59]. The evidence for this first tool use found in modern-day Kenya consists of scratch marks left by stone tools on animal bones when some of Lucy's kin, members of the genus *Australopithecus*, likely butchered meat for dinner. Clearly, the genus *Homo* from which we derive was not the first human species to use tools. However, the first widespread tool culture, the Oldowan tool tradition is seen as anchored to the arrival of *Homo habilis*, the first known member of genus *Homo*, about 2.4 million years ago. Typical start dates for the Oldowan tool culture are between 2.6 and 2.5 million years ago, which is a little earlier than the oldest fossil record we have for *Homo habilis*. We do not know yet how the much earlier stone tools produced by genus *Australopithecus* and dated to 3.3 million years ago relate to the Oldowan tool culture. It would be surprising not to find one human species learning from another, but we may never find the evidence confirming it.

In the time window of fig. 4.1, the only human species we know of before 2.7 million years ago were members of the genus *Australopithecus*. Hence, we must associate all stone tools dated from about 3.3 to 2.7 million years ago with this genus until we find evidence for another human genus using stone tools before 2.7 million years ago. At that time, the genus *Paranthropus* joins the paleontological record. Currently, there only seems to be evidence for the genus *Paranthropus* using bone tools but not stone tools. As discussed in the previous chapter, there has been quite some discussion as to whether *Paranthropus* represents a separate genus or constitutes a more robust branch of genus *Australopithecus*. If the latter should be the case then the question needs to be asked: why should this branch of genus *Australopithecus* not have used stone tools since by the time *Paranthropus* arrived, members of genus *Australopithecus* had done so for presumably some 700,000 years? If genus *Paranthropus* represents a branch of human evolution distinct from genus *Australopithecus* then we need to ask ourselves: why should we assume that *Paranthropus* did not use stone tools given that this species cranial capacity was comparable to *Australopithecus*? More often than not we tend to underestimate what members of such long extinct species might have been capable of. Maybe we should give genus *Paranthropus* some credit for stone tool use until we definitely can prove otherwise; after all, the absence of evidence is not evidence for absence.

From 2.4 to 1.8 million years ago we have three known human genera as potential stone toolmakers: *Australopithecus*, *Paranthropus*, and *Homo*. Then, from 1.8 to 1.2 million years ago there remain two genera: *Paranthropus*, and *Homo*; and after 1.2 million years ago, only the genus *Homo* remains as the potential originator of stone tool artifacts younger than that. Of course, any new fossil or tool finds or insights into human evolution could change this picture as it frequently has happened in the past. As mentioned above, the emergence of the Oldowan tool culture is associated with the arrival of *Homo habilis*. What we do not know is how *Homo habilis* became to be the Oldowan stone toolmaker. Did he just invent it from scratch, or did he benefit from long existing tool traditions going all the way back to the first stone tools we know of, some 3.3 million years ago? It seems to be much more likely that the latter must have been the case. How exactly such a transition, if it then happened, played out, we do not know yet. But there is an implication: *Australopithecus*, *Paranthropus*, and *Homo* must have interacted over a prolonged period for tool traditions to be copied and to be subsequently improved into what we now call Oldowan tool culture.

Stone tool artifacts associated with the Oldowan tool tradition continue until up to 1.2 million years ago. By then another tool culture, the Acheulean tool tradition, had already been flourishing for several hundred thousands of years. The first artefacts associated with the Acheulean tool culture are dated back to about 1.8 million years ago and its origins are thought to be tied to the arrival of *Homo erectus* roughly 1.9 million years ago. Oldowan tools were mostly scrappers of different kinds that had been produced by flaking off pieces from a stone core using a heavier hammer stone. The sharp edges of these various sized flakes were useful for multiple purposes depending on flake size. Heavy-duty Oldowan tools called choppers complemented these smaller and versatile tools. These heavy-duty tools were produced by skillfully flaking off pieces from one side of a larger stone core, thereby forming a cutting or hammering edge, resulting in tools that could be used for chopping, hence the name. The Acheulean tool culture continues these traditions but at a refined level and importantly adds versatile hand-axes to the tool set. The archeological record tells us that the carriers of the Oldowan tool culture made their stone tools on site just when they needed them. They did not carry them along when moving to new locations, not the heavy ones for sure; they just would make new ones wherever they moved.

With Acheulean tools, specifically the hand-axes, our ancestors or distant relatives increasingly made tools to last longer, which they maintained and cared for to maximize their use. This transition is understandable as more time needs to be invested to produce improved tools and there must be a point where the investment to make a good tool becomes just too substantial to not maintain it for longer use; and maybe also keep heavy tools when moving to a new site as long as they are still fully functional. While no fossil sites with Oldowan tool traditions have been found yet outside of Africa, Acheulean tool traditions are not only widespread in Africa but also in Eurasia. This should not be surprising as the initial carrier of the Acheulean tool tradition was likely *Homo erectus*, the first human species to migrate out of Africa. But it was not only *Homo erectus* spreading this tool tradition. *Homo antecessor*, *Homo heidelbergensis*, and later, early Neanderthals and Denisovans also contributed to the wide spread of Acheulean tool culture. Spread it did but not quite everywhere, as it seems that in parts of Asia the sharp edge of a cut and dried bamboo provided a technological alternative to the Acheulean tool set. About 200,000 years ago, evidence for Acheulean tool culture artifacts disappear in Africa and by about 130,000 years ago, the same happens outside of Africa.

Already about 100,000 years before Acheulean tool traditions disappear in Africa a new tool culture had started to develop, the Levallois tool tradition. With the Levallois tool culture, the early part of what is frequently referred to as the Levallois-Mousterian tool culture, human stone tool production moves to the next stage. Producing stone tools through flaking is not difficult. With some trial and error, one can produce a simple scrapping flake from a stone core quite easily. However, producing a stone tool of preplanned shape and size through flaking from a larger stone core is a quite different achievement. Envisioning the final stone tool shape in the raw stone core requires abstraction, something that was not needed to produce tools in the Oldowan or Acheulean tool traditions. With the beginning of the Levallois tool technology, our human ancestor had started to master this skill.

Until very recently, the earliest fossil records for *Homo sapiens* seemed to indicate that our species appeared no earlier than some 200,000 years ago. However, this picture has now changed. While it is still being debated if the human fossils from just before 300,000 years ago discovered at the Jebel Irhoud site in Morocco belong to an earlier form of *Homo sapiens* or represent the fully evolved version, increasing evidence points to *Homo sapiens* being around for quite a bit longer than we thought. That early members of a species are somewhat different than members of the same species coming much later should not surprise us; we have seen this before with other long-lived species such as *Homo erectus*. More recently, the reconstruction of a ∼210,000-year-old *Homo sapiens* skull from Greece and new discoveries in the Near East reset the clock for this species even more dramatically [54,58]. What is intriguing here is that these earlier dates for the arrival of *Homo sapiens* now seemingly could coincide with the appearance of Levallois tool techniques.

Before these discoveries, a *Homo sapiens* arrival date of about 200,000 years ago separated it safely from the emergence of Levallois techniques, which originated in Africa some 300,000 years ago. Therefore, the assumptions had to be that *Homo sapiens* so-to-say inherited the Levallois tool traditions from his ancestors. Not so now. While a similar date for the emergence of the Levallois technique and the appearance of the earliest known type of *Homo sapiens* may still be a coincidence, it certainly is suggestive. But the picture is even more complex. Originally, the assumption was that the Levallois technique originated in Africa some 300,000 years ago and from there migrated into Eurasia as archaic humans moved out of Africa, which as we now know included earlier members of *Homo sapiens*.

However, there seems to be increasing evidence suggesting that the development of Levallois type techniques may have been the result of independent stone knapping improvements over a wide geographic region. Rather than being migrated out of Africa by archaic humans, the geographically widespread parallel development of Levallois tool traditions, in instances seemingly preceding its 300,000-year origin in Africa, may have been the logical outcome once tool making had progressed to a certain point during the Acheulean tool culture [60]. Maybe, it were even members of several human species making the first steps towards Levallois type stone tool making. There is evidence of Levallois type tool techniques associated with fossil finds at the Jebel Irhoud site in modern-day Morocco [61,62]. Initially these fossil finds were attributed to *Homo sapiens* but are now seen as belonging to a pre-modern form of *Homo sapiens*. The latter eventually replacing this earlier form of modern humans similar to what happened to other archaic human species in Eurasia once *Homo sapiens* had migrated there.

Human species including archaic relatives or ancestors of *Homp sapiens* were much more widely dispersed than previously assumed and so were Levallois tool cultures. However, eventually it would be *Homo sapiens* in Africa and the Neanderthals in Eurasia to carry the Levallois tool culture. As for Asia we probably need to give similar credit to the Denisovans. The Neanderthals never made it this far east, and modern humans arrived there too late to be the makers of the Levallois stone tools recently discovered in this area [60]. Late members of *Homo erectus* in Asia can likely be excluded as well as *Homo erectus* disappears in Asia around 150,000 years ago while the Levallois tools discovered in southwest China are dated to 170,000 – 80,000 years ago [60].

The Mousterian tool culture represents a perfection of Levallois tool techniques, enabling the creation of more specialized flake tools which in turn seem to have replaced the use of hand-axes in a number of ways. Mousterian tool culture sites are located in Europe, western Asia, and northern Africa. Mousterian type stone tools start to appear around 160,000 years ago and practically vanish with the disappearance of the Neanderthals. Because of that and their geographical spread largely overlapping with the spread of Neanderthals, Mousterian tool culture for a long time used to be associated only with Neanderthals. However, we know today that modern humans and their archaic relatives and ancestors were present in North Africa as well. The recent discovery that the first modern humans migrated into Eurasia significantly earlier than we thought only confirms this. Therefore, paleontologists think that modern humans and pre-modern ancestors of *Homp sapiens* were the carriers of Mousterian tool traditions in northern Africa. We also know that modern humans and Neanderthals were in close contact throughout the Near East for a long time before modern humans migrated out of Africa. We can assume that the two human populations lived in close contact for extended periods and sharing tool cultures would likely have been a natural consequence of a long neighborly relationship. After all, we know today that eventually they would interbreed somewhere in that area maybe even before the first exodus of some modern humans into Eurasia; or maybe not. The current DNA evidence only supports the intermixing between Neanderthals and modern humans some 75,000 to 50,000 years ago.

Homo floresiensis, the human Hobbit species, thought to have evolved from a *Homo erectus* population stranded on the Indonesian Island of Flores, represents in some ways an interesting evolutionary experiment. The oldest stone tools found on the island are dated back to about one million years ago, which must have been when *Homo erectus* arrived there. We do not know exactly when rising sea levels isolated Flores to the extent that the *Homo erectus* population on Flores could not leave anymore. However, once that happened, paleoanthropologists believe that island dwarfism eventually forced the evolution of the original *Homo erectus* population into what became *Homo floresiensis*. The stone tools found next to fossil remains of this species date to between 190,000 to 50,000 years ago, providing solid evidence that *Homo floresiensis* remained a prolific tool maker of tool sets similar to Lower Paleolithic tools found in Asia or Africa. However, the species making these tools not only had shrunk to the size of a hobbit, it also had shrunk its cranial capacity to fit the smaller body size. The brain size of *Homo floresiensis* shrank back to less than half of the average brain size of *Homo erectus*. In fact, the brain size of *Homo floresiensis* was very similar to the brain size of an *Australopithecus*. So, while its brain shrank by half, *Homo floresiensis* continued to be a skilled toolmaker. Should that not make us wonder about the tool making capabilities of much earlier humans with similar brain sizes?

The Mousterian tool culture was followed by a number of new tool cultures and traditions developed during the roughly 30,000- to 40,000-year-long Upper Paleolithic. That a culture follows another one is often more a figurative speech than reality and so it is here, where we see the end of the Mousterian overlapping with new tool cultures.

Modern humans migrating into Europe in the Upper Paleolithic are associated with the Aurignacian tool culture. We know the carriers of this tool tradition and of the other Upper Paleolithic tool traditions that would follow it as the Cro-Magnons, members of the *Homo sapiens* family named after the place in southwestern France where their fossils were first discovered. However, it looks like these modern humans were not the exclusive producers of improved Upper Paleolithic tools. Several archeological sites of the Châtelperronian tool culture actually also seem to be associated with Neanderthal fossils. Given how our picture of the Neanderthals has changed over the last decades it is not surprising to see Neanderthals master similarly advanced tool techniques as the Cro-Magnons. However, by shortly after 30,000 years ago Neanderthals became extinct and if there were still some small pockets of Neanderthals left that we do not know of, by around 20,000 years the last members of their species had vanished for good. After that the Cro-Magnons and their descendants would be the only carriers of the tool cultures to follow. The Denisovans shared the fate of their Neanderthal cousins though we do not know if they may not have become extinct somewhat earlier or later. The Denisovan fossil record is much less rich than the Neanderthals and we know even less about the tool cultures the Denisovans practiced. The Denisova cave in the Altai Mountains of Russia is associated with Acheulean, Levallois, and Mousterian type tools; different human species including Denisovans occupied the cave over much of the past 300,000 years. All we know today is that Denisovans inherited the Acheulaen tool culture of their ancestors and that they are associated with a variant of Mousterian tool traditions. Both, Neanderthals and Denisovans, left some of their genetic legacy with us as they interbred with modern humans when the latter expanded into Eurasia. Interbreeding between human species has likely happened during human pre-history more often than not. We should look at *Homo sapiens*, the only human species populating Earth today, as a hybrid species. Similarly, we should be open-minded with regard to human cultures also being hybrids as most of them, at least to some extent, always have been. Occasionally, cultures are just so adept at integrating external cultural influences that it is hard to tell that they ever came from the outside at all.

Fig. 4.1 lists some of the tool cultures *Homo sapiens* developed during the Upper Paleolithic. Human pre-history in Europe and the Near East has been studied for much longer and in much greater depth than the contemporary periods in other regions of the world; a bias that fig. 4.1 reflects. It needs to be emphasized that humans across time and in different regions must have been equally capable of developing tool technologies, as we know from the archeological evidence. To mind come the peoples who first settled Australia and the many Pacific islands, all corners of the Americas, or also much of sub-Saharan Africa, just to name a few. We know much less about tool cultures of peoples first settling in regions that are not researched and cataloged yet to match up with records for Europe, North Africa, or the Near East. But that does not imply they are in any way inferior. Far from it. It just so happened that the archeological records of pre-historic peoples were first established by scholars in Europe based on what they discovered so-to-say in their own backyards; or in nearby regions with cultural affinities to the broader European context which framed their perspective on human pre-history. During the late Middle Paleolithic and Upper Paleolithic humans spread to all corners of the Earth except Antarctica and the North Polar Region. As they did so they continued to evolve their tool cultures albeit driven by different environmental constraints and at different speeds. This is why we see the Mesolithic and Neolithic stages of the Stone Age overlapping quite significantly as some regions transitioned out of the Mesolithic into the Neolithic earlier than others did. So speaking of a Mesolithic Age in a world-wide context really does not make sense as in some areas it already gave way to the Neolithic while in other areas it would persist for much longer. Cultures need to be looked at locally, as cultures for much of human history always have been local. Certainly, what local really meant at any given time could be changing and mean something different in South and East Asia as compared to for example in Western Europe. We need to keep in mind that well into the last century, Stone Age cultures managed to survive in a few remote areas of our world. So technically, one could maintain that the Stone Age was not over a few thousand years ago but that instead it lasted up until a few decades ago. Clearly, that would not be a very meaningful statement, but it illustrates how cultures at different stages of development could exist in parallel for a long time.

Human tool development is a good indicator of human culture evolution and this is why we took this brief excursion through tool development in human pre-history. There are of course also numerous other achievements but some of them are so-to-say one-time achievements such as for example the first use of fire while others are tied to materials that do not preserve as well over millions or hundreds of thousands of years such as clothing. We can safely assume that early humans must have used fire in an opportunistic way long before they were able to make fire at will. There is a fear of fire deeply engrained in all life. We can see that with animals, and we feel it ourselves. However, if one can overcome the fear of fire, one can use it for one's own protection. At some point, early humans must have realized that they could carry away burning flames from where they naturally occur, as for example started by lightning. Humans likely still ran away from large wildfires just as we and all other animals still do, but smaller contained fires could have been an opportunity. Naturally, we cannot know when that first occurred, but quite likely, we can assume it must have been sometime before humans learned how to kindle fire themselves. The latter may have happened as early as one million years ago as indicated by archeological evidenced unearthed in the Wonderwerk Cave in South Africa [63,64]. There are older sites with evidence of burning fires, but they are all in the open and there we really do not know if the fire had been kindled or not. Not so for the find in the Wonderwerk Cave where we can be quite assured that the fire was indeed kindled. Given the date, roughly one million years ago, this makes *Homo erectus* the likeliest species to have been doing the kindling. Once humans had learned to control fire, the invention of cooking was likely not far behind. We do know from archeological sites in modern-day Israel that humans were cooking some 790,000 years ago [65].

It should not have taken humans 200,000 years from learning how to kindle fire to figure out the advantages of cooking food; and it very likely did not. We have to remember that we are reconstructing human pre-history from the evidence we can find, which we know will always be incomplete. Fortunately, in this case, in addition to the earliest archeological evidence for the kindling of fire by humans and of humans starting to cook their food, there are also other indicators that humans may have made fire and cooked their food long before those earliest dates for which we have definitive evidence of either. This brings us back to *Homo erectus* and the evolutionary change he represents when compared to *Homo habilis*. Not only is *Homo erectus* as tall as modern humans but his skeleton and body build are close to modern humans. Importantly, this includes a much smaller relative body volume for a gut than what *Homo habilis* still possessed. Scientists believe that this change is a very good indicator that *Homo erectus* must have cooked his food. Cooking food breaks down cell walls and releases many nutrients that then are easily digestible by a smaller gut. Raw food is less nutritious than cooked food and takes longer to digest. Therefore, *Homo habilis* had to consume large amounts of it, which is why his gut required a much larger relative body volume. That should give us confidence that *Homo erectus* must have known how to make fire and cook his food, which moves the date for both of those cultural achievements to some two million years ago; much earlier than what we can infer from the archeological evidence alone.

There is an important lesson here. Not only should we give early humans some credit for being smart enough to discover how to make fire and how to cook food much earlier than what we can glean from the archeological record. We must also remember that in exploring pre-history, as some have succinctly pictured it, we are really like explorers in a dark cave. Occasionally, by chance or perspiration, some of our flashlights fall on something interesting and we get excited and draw conclusions; but it is then that we must not forget the sheer vastness of this dark expanse we are trying to explore. Almost all of what we would like to read off the endless walls of this cave remains hidden to us, so we only ever get to see small samples of past realities from which we try to deduce what really happened. We may sometimes come close to the truth but then we have to realize that new evidence often forces us to acknowledge that several of our earlier conclusions were quite off the mark. That will not change. We will continue to iterate our way towards a better understanding of human evolution with persistence by adding new findings, but our sampling space is so large that we only can hope to get a good approximation at best; if we are lucky. While that is regrettable it does not distract from the fact that even the few data points we have and to which many researchers in different fields of science continue to add, already give us a much better view that helps us appreciate what we know and what we do not know.

There is no direct evidence for when humans may have first started to wear clothes because skins and organic fabrics or textiles leave no fossil evidence. With respect to clothing, early humans already had an advantage, just as all mammals did. The fact that mammals were hairy was likely one of their key advantages in being able to spread to every corner of Earth, as they could inhabit warmer, temperate, as well as colder climates. However, with humans something interesting happened: they lost much of their body hair and evolved the capability to sweat through their skin. This unique adaptation made humans very effective predators even though they had no claws or teeth as their fierce contemporary animal predators did. By wielding a spear and being able to run long distances without suffering heat collapse as animals do which can only cool by panting, they could practically hunt their prey to complete exhaustion for a comparatively easy kill. However, humans once still had enough hair for a single louse species to colonize our bodies, what we know today as the human head louse. That changed once humans lost most of their body hair and the original human louse retreated to our head hair becoming the head louse. When did that happen? It looks like it must have been before 3.3 million years ago at which point humans acquired a second louse, the human pubic louse.

Interestingly, humans must have acquired this pubic louse from gorillas; the speculation is from sleeping in abandoned gorilla nests. At the time humans acquired this pubic louse they must have already lost most of their body hair. Therefore, we can assume that humans were running around in their naked skins starting sometime around 3.3 million years ago. The next question is how long did it take them to cover their new nakedness with clothes? For modern humans the answer comes from a third kind of louse, the human body louse living in our clothes; that is, if we let it. In essence, it seems that humans seemingly created new habitats for lice by wearing clothes. Science tells us that the human body louse which lives in clothing is thought to have diverged from the human head louse some 170,000 years ago [66]. From that we can infer our modern human ancestors began to wear clothes around that time. But it does not tell us anything about other human species wearing clothes earlier than that or not. Given for example the cold climates that our Neanderthal cousins lived in it seems unlikely that modern humans were the first human species to wear clothes. There could also have been cultural and not climatic reasons to start wearing clothes; maybe along the lines of the narrative of the original sin where Eve and Adam covered each other realizing in shame their nakedness after being kicked out of paradise. Let us hope humans started wearing clothes for a much better reason than that; my personal favorite is vanity.

Finally, there is pre-historic art, if we want to call it that. It is doubtful if the “artists” who roughly forty thousand years ago began to leave us imprints of their world and their minds in the cave paintings we admire today, saw themselves as artists. However, we do know that something must have changed for humans to capture what must have mattered most to them, the essence of their world, on cave walls. The famous cave paintings of Altamira in Spain and Lascaux in France have been known for some time. Caves in El Castillo in northern Spain have revealed rock art from the Upper Paleolithic dated to about 42,000 years ago; including the now famous hand stencils virtually leaving us hand prints of our human ancestors. Similar ancient cave art, including animal paintings and hand stencils, has been found on the Indonesian islands of Sulawesi and Borneo. Rock art has been discovered in many other places, some of them even farther away from Western Europe and others quite unexpected. We can admire it in Australia with its rich and unbroken traditions of rock art going back some 40,000 years. And we find it in the middle of what is now part of Africa’s largest desert, the Sahel, which once was more of a lush garden than the endless stretch of sand it is now. The Upper Paleolithic also left us the first stone and clay figurines of animals and humans as well as combinations of both; like the famous lion-man figurine found in Hohlenstein-Stadel, Germany. Many human figurines are associated with fertility such as the so-called Venus of Willendorf, found near a village of that name in Austria. It seems that our distant human ancestors likely had a similar self-awareness to what we experience. We know that customs of self-decorating or painting each other, mostly with natural clay earth pigments such as ochre, go back for tens of thousands of years. These pigments were not only used on people but also on objects as we know from archeological finds such as the one at the Blombos Cave site in South Africa dated to around 75,000 years ago [67]. From all we know today about human pre-history, it looks like it is only in the Upper Paleolithic when we finally see humans beginning to express themselves in many different ways that all speak to us very directly across this vast ocean of time that separates them from us.

Frequently, the flowering of human imagination in pre-history is referred to as the *Cognitive Revolution*. If we associate its first manifestation for example with the artifacts from the Blombos Cave or with a somewhat later date when we find the first cave paintings around 40,000 years ago, its beginnings date back to the late Middle Paleolithic. The exact time does not matter that much here. What really matters is the assertion that the *Cognitive Revolution* was specific to modern humans, distinguishing *Homo sapiens* from other human species. To be upfront, yours truly is deeply skeptical of any notions that ordain our species with an intrinsic superiority that made the outcome inevitable; it often enough just sounds like we were the chosen ones, and the other species were somehow inferior. It may well be that *Homo sapiens* had special capabilities manifesting themselves in behavioral traits traceable in the archeological record for which we do not find analogues in human species contemporary to modern humans, such as the Neanderthals. Having acknowledged this, we must remember that all we have is a very selective archeological record in which we keep searching for traces that can help us understand how we came to be who we are today. In addition to looking for traits of a *Cognitive Revolution* in other human species such as the Neanderthals, we have to ask ourselves a very different question. Who is to say that what we became is in any way superior to what the Neanderthals or other species could have evolved into if they had been dealt a superior hand that predestined them for survival and not us? As many a paleoanthropologist has experienced, even when it is still possible to conduct field studies of archaic cultures surviving to this day, it is extraordinarily difficult to understand cultures that are completely different from our own. For that matter, what if Neanderthals rejoiced in singing their perceptions of the world rather than painting them on cave walls?

Neanderthals must have encountered examples of the Cro-Magnon's cave art and they certainly would have had the capability to copy some of it. They did not. We do not know why. But can we exclude that they had other means of expressing within their culture what Cro-Magnons expressed through cave paintings? We know that in many ways Neanderthals were not so different from modern humans, in some, they even may have been superior. Neanderthals seemingly cared for the weak among themselves and at some point they started to bury their dead. They also developed quite sophisticated technology such as the glue that they used to fix stone spear points to their spear shafts. Why do we even consider Neanderthals not capable of similar cognition as Cro-Magnons? Why do we not assume that they could just have expressed themselves differently? Rather than making a plea for Neanderthals, the above is really a reminder that we are prone, often all too easily, maybe subconsciously, to employ a lopsided approach when assessing other human species. Specifically, when the unspoken but anticipated outcome is that our species was somehow always superior.

As a warning, we just need to look at how our perception has evolved with respect to many native cultures, ploughed under by colonialism and modern civilization alike. We can take the Aborigines in Australia as an example, *Homo sapiens* just like us. It is simply appalling, how for a long time they were pictured as not only being primitive and brutish but in addition as incapable to ever become members of our seemingly so advanced societies. This widely held perception was likely motivated as much by sheer greed as by racial prejudice; it is easier to steal land from someone who is deemed inferior to you, someone who should not really be entitled to hold land at all. It took a long time and Aborigines had to suffer much until this perception changed. We know today that the native people of Australia have a no less rich culture, than other members of *Homo sapiens sapiens* have, and in many respects actually a more humane one. There are many other examples that could illustrate how our species has demeaned and exploited those who were deemed ethnically and racially inferior. We know much less about other ancient humans than we do about any of the people that have been enslaved and exploited by our ancestors not too long ago. But no enslaved people were ever inferior to any other people. They were just unfortunate in being the victims and not the masters in what is frequently a deeply racist human history. So what makes us so sure that other ancient human species were somehow inferior so we can deem them incapable of having had their own *Cognitive Revolution*?

At the same time, this may be the right and the wrong question to ask. Clearly, there are indications that our late Neanderthal cousins were on a path towards a *Cognitive Revolution* of their own. Unfortunately, they became extinct before it could manifest itself in a similar way as we now in retrospect can see it for our modern ancestors. There may be a simple answer for why we do not see a blooming *Cognitive Revolution* in the archeological record for Neanderthals; or any other archaic humans for that matter. The explanation that has been proposed is intriguing and so far convincing at the same time and it has nothing to do with modern humans being smarter than Neanderthals; for all we know it was the Neanderthals who had the larger brains. The key to understanding this was the realization that a species population size needs to cross a certain threshold for the kind of *Cognitive Revolution* to emerge that modern humans went through, starting some 40,000 years ago. Earth climate favored modern humans in that respect as the foothold Neanderthals had in the much colder northern latitudes was always tenuous and never allowed them to grow a large enough population for the seeds of a *Cognitive Revolution* to fully sprout. Modern humans just got lucky and if the roles would have been reversed it may well be that Neanderthals could be wondering today how they managed to pull off a *Cognitive Revolution* while modern humans only made it to the early beginning of their own.

The end of the Paleolithic is marked by the beginning of the Holocene epoch, the warm climate of the interglacial period that followed the last of the ice ages in Europe, which lasted from about 115,000 to 11,700 years ago. With the end of the Paleolithic and the climate warming into the Holocene, we see the beginnings of agriculture. Humans had supplemented their diets with wild grains long before that and they must have kept animals. Agriculture was not appearing out of nowhere, rather humans were ready to try it once circumstances enticed them to do so. The Fertile Crescent, frequently referred to as the cradle of civilization, is the area where we find the first evidence for agriculture and domestication of animals. As the name suggests, it is a roughly crescent shaped area that starts in the Nile Valley of Egypt to rise towards the northeast through lands that are today part of Israel, Lebanon, and Syria into Turkey where it crests at its most northern point to swing in a southeast curve into Mesopotamia, now part of modern-day Iraq. In the late 1920's, the British archeologist Dorothy A.E. Garrod (1892 – 1968) excavated caves in the Near East that she thought may have been occupied by humans in pre-history. In 1928, the Shuqba cave located in the western Judean Mountains of what is today the West Bank revealed just such a presence of pre-historic humans. This was the beginning of the discovery of a people that we know today as the Natufians, the originators and carriers of what archeologists call the Natufian culture. It was Garrod herself who coined the name. Since we cannot know how these pre-historic people referred to themselves, she named them after the Wadi an-Natuf on the northern banks of which Shuqba cave is located.

Natufian settlements spread over lands that are part of today's Israel, the West Bank, Jordan, and Syria. We find their first villages for example in places such as Ain Mallaha and the HaYonim cave in modern-day northern Israel and in the West Bank at the site of the modern city of Jericho. A similar culture excavated in Tell Abu Hureyra in the Euphrates River Valley in Syria seems to mark the farthest northern extent of Natufian settlements. Natufian culture is likely to have existed in this Near East region starting from around 12,500 up until sometime around 9,500 years ago when it seemingly vanishes from the archeological record. Further discoveries may change those dates to an earlier arrival or later departure of these people in our historical records. The important fact is that the time window for the presence of Natufian culture includes the closing of the Paleolithic period and the beginning of the Neolithic with the Mesolithic period sandwiched in-between. Archeologists refer to this time window as the Late Epi-Paleolithic because the Mesolithic phase of pre-history, as pointed out earlier, is not synchronously observed when comparing archeological records in the Near East, Europe, and Asia. For example, in the Near East, traces associated with the Mesolithic occur earlier than in Europe or Asia.

What makes the Natufian culture so remarkable is that it provides a window into humans transitioning from leading hunter-gatherer lifestyles into becoming agriculturalists. The lands the Natufians occupied, probably in groups not exceeding a few families, were rich in wild barley and wheat. Estimates suggest that a few weeks of harvesting these wild grains could produce enough grain to sustain a family of four for up to one year. Large amounts of grain are not easy to carry around and this is likely one reason why Natufians started to build settlements. This brings us to the question of how humans becoming sedentary, and the beginnings of agriculture relate to each other, if at all. Over the past century, archeologists have debated this topic intensely. For a long time, it had been assumed that the arrival of agriculture provided the incentive for humans to move together in more or less permanent settlements. However, that was not so. Today we have the archeological evidence that hunter-gatherers already lived sedentary lives to some extent at least in some areas. The archeological site of Ohalo in Israel's Rift Valley demonstrates that hunter-gatherers some 23,000 years ago could choose to lead sedentary lives in that region. There are also examples for hunter-gatherers building houses and essentially living communal sedentary lives in places outside the Levant. Lepenski Vir, a site located at the Danube River within the Iron Gates marking the boundary between modern-day Serbia and Romania, was occupied as early as 12,000 years ago. As such, this settlement of hunter-gatherers and fishermen represents the oldest European village we know of, many thousand years before agriculture would arrive in Europe.

Therefore, when we look at Natufian settlements and their potential role in early agriculture we have to keep in mind that humans had been sedentary when it suited them for a long time before our species ever started to cultivate grains. The earliest Natufian sites found to date indicate for such settlements only a few single room huts built into the soil, made of stone, and likely covered with branches and grasses. Interestingly, they also seem to have buried their dead in the floors of their huts; something we also find in earlier hunter-gatherer settlements. Among the tools found at those sites are sickles made of flint blades that were set straight into carved out bone handles. Trace analysis of the material swapped from the edge of the blades indeed shows that these sickles were used for harvesting grains. Large stone pestles and stone mortars used for grinding grain found at the sites were certainly too heavy for hunter-gatherer groups to carry around. The rich sources of wild grain and the food security that they must have provided is likely what enticed the Natufians to settle. In this context, we need to remember that wild grains over time may have become an increasingly important food source and humans have been braking bread long before they started agriculture; the same is incidentally true for making beer. However, Natufians did not give up their hunter-gatherer lifestyle yet as they still supplemented their diet by hunting, using spears and slingshots as can be seen from the round polished stone artifacts found in their dwellings as part of tool kits. It looks like that at this early stage they made no effort to cultivate grains. That started to change as the climate turned colder again.

The time from around 12,700 right to the beginning of the Holocene around 11,700 years ago brought a return of glaciation and with that colder climates. For the Near East, this last glaciation phase of the Pleistocene, the so-called Younger Dryas, brought a drier climate, the result of which was that wild grains became much scarcer. At that point, some of the Natufians seem to have taken the plunge and instead of moving out of the area and going back to a hunter-gatherer lifestyle doubled down on agriculture. They are thought to have done so by relocating to sites that provided access to water and to soils that could be used to grow grains. We do not know how difficult it was for them to accomplish this, but they succeeded as we can see today from the archeological records of the first granaries. The grains they were sowing and harvesting which were found at these archeological sites are larger than the wild ones they had been opportunistically harvesting before; a tell-tale sign that agriculture had taken root.

Parallel to the development of agriculture in the Near East we also find evidence of animal domestication, including practically all traditional farm animals as we still have them today. Of course, at that time, dogs had already been living and hunting alongside humans for thousands of years. The Younger Dryas lasted for about one thousand years. Once the climate warmed again at the end of the Younger Dryas, Natufian agriculture started to spread to the Northeast across the Fertile Crescent. It is not clear when exactly grain based agriculture and animal husbandry found together as some of the typical farm animals such as pigs seemingly were domesticated earlier in other parts of the Fertile Crescent. We also do not know if the descendants of the Natufians or different people spread agriculture across the Fertile Crescent.

Archeologists frequently refer to this transformation from hunter-gatherer to agricultural societies radiating out from the Fertile Crescent as the *Neolithic Revolution*, a label first used in 1923 by the Australian archeologist Vere Gordon Childe (1892 – 1957); sometimes archeologists also call it the first *Agricultural Revolution*.[4] Labeling this transition, which archeologists approximately date to some 12,000 years ago, the *Neolithic Revolution* may sound confusing because this time frame seemingly should belong to the Mesolithic or Late Epi-Paleolithic. Indeed, it is confusing, and part of the reason is that Childe thought of the Mesolithic as a useless concept and ignored it. His scheme for human pre-history went straight from Paleolithic to Neolithic so there could only be a *Neolithic Revolution* as the Mesolithic was for him out of the question. However, it is not as bad as it sounds as some of the most important grains such as barley where indeed domesticated in the Neolithic as archeologists define it in the Paleolithic, Mesolithic, and Neolithic sequence. We should also not forget that while revolutions are usually a better sell in any scientific discipline and likely more helpful to a scientist's career than gradual transitions, there was a long period where humans experimented with grains and agriculture that preceded the onset of the *Neolithic Revolution*. Among the place names associated with the beginning of this *Neolithic Revolution*, we find for example Jericho, Abu Hureyra, and Jarmo in modern-day Palestine, Syria, and Iraq, respectively. All of which provide evidence for the earliest domestication of grains.

From the Fertile Crescent agriculture spread through modern-day Turkey up into Europe. As it did so, new centers of agriculture arose such as Çatalhöyük in the Anatolian plain. Sometimes we just do not know if people move and thereby transport culture or if culture spreads by itself without people moving. About 6,000 years ago, farmers from the Near East, their agriculture, or both emanated from the Fertile Crescent and started to make their way up into Europe. Likely, it was a bit of both. Moving into Europe from Asia Minor was a very different challenge back then because Europe was still heavily wooded. Therefore, these farmers moved along rivers and coasts. Europe was already populated and compared to the size of indigenous European populations the number of farmers that moved into Europe was likely small. Farming was new in Europe. The soils were different and the seasons in northern latitudes were certainly more pronounced than in the Fertile Crescent. The transition from hunter-gatherer lifestyles to agriculture happened therefore slowly. Rather than an agricultural frontier washing over Europe, agriculture percolated through Europe, conquering small areas that looked promising for farming, one at a time. Clearing the forest and preparing the soil so it would be fertile ground for grains to grow was hard work. In the best case, woods could be cleared using fire but more often than not stone axes had to be used to cut down forested areas tree by tree to clear the ground. Because European soils were much heavier, it took the invention of the animal drawn scratch-plow for farming to conquer Northern Europe and move into Asia.

While up to this point it may have sounded as if agriculture developed only in one place, the cradle of the Fertile Crescent, nothing could be further from the truth. The Fertile Crescent was particularly important for the spread of agriculture into Europe and parts of western Asia, but East and South Asia, the Americas, as well as Africa developed their own crops. The cultivation of wild Asian rice occurred sometime between 13,500 and 8,000 years ago which is about the same time we see the cultivation of grains in the Near East. The tools of genetics seem to suggest that all varieties of domesticated rice we find in Asia today can trace their origins back to this time window. The most likely places archeologists have identified for the earliest cultivation of rice are the middle Yangtze and the upper Huai Rivers. Millet, another important early cereal crop seems to have also been cultivated first in the same areas not long after Asian rice itself. Evidence points towards soybeans being domesticated in northern China around 4,500 years ago. Wild African rice was cultivated later, sometime between 2,000 to 3,000 years ago in

West Africa. In modern times, the transatlantic slave trade would bring domesticated African rice to the Americas. Similar to the domestication of grains, the domestication of other plants also started in the Fertile Crescent among them lentils, peas and flax. Figs and olives were domesticated as well, figs somewhat earlier and olives by 6,000 years ago, although there seems to be still some debate as to where olives were cultivated first. Many of our most important food crops were cultivated first in the Americas. This includes such basic food staples as corn also called maize, which may have been cultivated as early as 7,000 years ago and potatoes even earlier than that, sometime between 10,000 and 7,000 years ago. Among other gifts, the indigenous people pioneering agriculture in the Americas also gave us new varieties of beans, peppers, squashes, and of course tomatoes. By the middle of the Neolithic, around 8,000 to 7,500 years ago, the transition to agriculture was mostly complete. In some cases, it was grains and husbandry combined, in others it was focused on grains only and then there were areas where nomadic pastoralists would retain a powerful foothold with their herds for millennia to come and a few are still living such a life today.

This brief walk through human pre-history from the first use of tools some 3.4 million years ago to the development of the first written language that followed soon after the transition to agriculture has certainly left out many things. The most amazing aspect of pre-history is the quickening pace with which our human ancestors became more and more like us. For almost two million years after the dawn of pre-history, progress seemed to be slow as judged from the artifacts we can find. However, we have no window into what likely were the most important changes of all, those that happened in our brains. We can only try to reconstruct the process of becoming human, as we define it in our own terms today by judging the artifacts that our early ancestors left behind. Archeology tells us that humans frequently choose to live together in village-like settlements long before the *Neolithic Revolution*. Before the *Neolithic Revolution*, ushered in by the rise of agriculture, there must have been another equally important transition where humans learned to live sedentary lives with all that comes along with that. That earlier transition likely played out over a much longer time, preparing our ancestor for a life in the more closely-knit communities that we see emerge with the *Neolithic Revolution*. It was with the transition to agriculture, that a few thousand years later the first real cities emerged. Sometimes archeologists refer to this rise of the first real cities in the modern sense as the *Urban Revolution*, just like the *Neolithic Revolution*, a term coined by Gordon Childe. As it stands today, until recently, the oldest city on record was Uruk in Mesopotamia, which traces back its origin to about 6,000 years ago. Settling in cities became possible with agriculture providing the means to feed larger numbers of people. Toiling the land and provisioning cities with grains meant counting became important. The need to record what goes in and out of granaries and to keep track of land ownership led to the development of the earliest writing systems some five thousand four hundred years ago, the Sumerian cuneiform and the Egyptian hieroglyphs.

This is where roughly 3.4 million years of human pre-history end and our brief ∼5,000 years of recorded history begin. To put this period in perspective, the evolution of higher life on Earth began some 540 million years ago and our earliest evidence for life on Earth goes back some 3,800 million years.

The Enigma of Human Culture Evolution

In a simplistic way we could look at our biological evolution has the evolution of the human hardware and differentiate it from the evolution of the human software, what goes on in our brains so-to-speak. Unfortunately, we cannot explore what thoughts our early human ancestors had or when they started to have thoughts in the common meaning of the word at all. The evolution of what we just referred to as the human software does not preserve so there is no archeological record documenting it. When was the light switch turned on, when did the first human refer to her- or himself as a distinct person? We most likely can never know that. However, it is reasonable to assume a gradual process rather than a switch turned on. Human consciousness must have been evolving over a long time. At some point, humans must have realized that they had consciousness; they must have become increasingly self-aware in a process that could have taken a very long time; or not, as we just do not know this. Are self-awareness and self-consciousness uniquely human qualities or do other animals possess them as well, at least to some degree? We know that all animals must have self-awareness to the extent that they otherwise could not function in their environments. For some animals living in social groups, this goes even further as they play the specific roles that exist in their given social orders. Where it gets dicey is the next level which is what we would refer to as full self-awareness in the sense that you know that you are conscious of yourself, know inwardly and outwardly who you are. We usually refer to this as our self-consciousness. The jury seems to be still out if and to what extent animals are capable of this but many more animals than we would have thought just a few decades ago are now passing some version of the famed mirror test, which is still widely used to test for traits of self-awareness. However, for all our good intentions to test for such self-awareness traits in animals, we seem to be testing for specific articulated awareness in a strictly human sense. Mirrors are just not as much a part of the animal world as they are of our human created world. Increasingly, biologists have started to realize that it is very difficult and maybe even impossible to test for the presence or absence of certain qualities that we see as essentially human in a human created test setting. What Werner Heisenberg observed for physics is not restricted to this science discipline but a realization of a much bigger challenge and in its own sense quite true for biology as well:[5]

> *"We have to remember that what we observe is not nature herself, but nature exposed to our method of questioning."*

It is perfectly normal for human beings to react to any kind of human administered stimuli in human provided settings. All of us do so every day in a more or less natural way but certainly always in a human one that we can relate to or understand. Not so for other animals. Putting them in a human environmental context and testing for human type responses is fraught with difficulties. Ultimately, we could ever only hope to understand animals in their own environments. However, even then our human perception will still be different from how fellow chimpanzees for example would perceive the behavior of any individual in their social group. Biologists have realized that, at least many of them have, the result of which is the emergence of ethology, a new discipline of biology focused on the study of animal behavior as much as possible under their natural conditions.[6]

This new approach could not be further removed from how animal behavior was still studied only a few decades ago when the dogma was that animals are nothing more than biological machines. Biological machines had of course no self-awareness or feelings and their reactions to stimuli were testable under whatever cruel schemes some of the leading scientists back then could conceive of. If one tortures animals in such tests, it is of course convenient to assume that they are nothing but machines. Times have changed. Today it is not heresy anymore for biologists to consider if animals have emotions. Instead, most biologists today would acknowledge that animals do have emotions and good careers await those in biology looking into studying the depth and breadth of animal emotions. As we find more and more evidence for animal emotions that also sheds a much starker light on human cruelty towards all kinds of animals but on none more so than the ones we raise for slaughter.

It is hard to reconcile our human species perceiving itself as a moral species while we continue to torture and slaughter animals that we know have emotions and feel pain; not to speak of any self-awareness that we are not aware of yet. Personally, most of us presumably believe our dogs and cats always knew who they were but they likely would all have failed the mirror test. We just can see them standing in front of a mirror not realizing what they are looking at, can we? However, that would be irrelevant as our animal friends and family members always knew that we knew that they knew we just love them for who they are. Therefore, we will stick here with human self-awareness only. When did this specific human characteristic emerge? How did it emerge? Unfortunately, we may never be able to answer these questions. However, we can look towards how and when some of the specific traits that make us humans human emerged and hope that this will help us understand how we became who we are.

What is it that over the past few million years made us so different from our closest living animal cousins in our family tree, chimpanzees and bonobos, with whom we shared a common ancestor some 5 to 7 million years ago? What are these traits that resulted in humans evolving into what we are today while chimpanzees, in evolutionary terms, seemingly are still pretty much the same species they were several million years ago. There is no way that a Darwinian type of evolution could explain this and Darwin knew it. Does that mean that we should resort back to biblical or mythical accounts of how we became who we are? Certainly not, but maybe deep down, humans always had a hunch that they were somewhat different from other animals. Many human cultures have mythical accounts of how humans were formed out of clay and animated into the human beings we are today by some divine force breathing life into the clay, or by some other similar symbolic means. In these mythical accounts human cultures, inadvertently or not, have framed the most important of all questions. Translated into modern terms the moldable clay represents biological evolution and the act of breathing life into clay becomes the emergence of human culture evolution. The currents of human culture evolution and biological evolution somehow joined to produce this incredible human species transformation, from ape like creatures living in small groups only a few million years ago to who we are today. How did this happen?

Anthropologists have tried to answer this question for quite some time but with limited success. Looking back into our more recent past, we can understand how for example human societies a few thousand years ago evolved into the societies of today. There are seldom straight lines connecting the past to the present, if ever, but we clearly can see this more recent past as once being part of us. Throughout all the complicated processes that shaped this more recent history, we can still discern the major threads that provide a continuity from the there and then to the here and now. Yes, some of the facts may have been lost or may not be completely accurate but overall, we get a good idea of what life could have been a few thousand years ago; complete enough that we may picture ourselves living in those past times. As we look back further, this connection becomes weaker, but we still can imagine ourselves to some extent as our ancestors, interacting in ways we could understand. However, there comes a point in our past where this breaks down. Once we pass this threshold and move into the even more distant past, we are not able anymore to imagine ourselves in this past. Of course, there is no clear delineation in the past in terms of a threshold where this happens. In reality, it is more like a diffuse zone, a nebula separating the recent past to which we can still connect to, from the more distant past where we become strangers to ourselves. If this nebula stretches over thousands, tens of thousands or hundreds of thousands of years we cannot really say; but it is there, that we can say with certainty. How could we imagine ourselves as humans who for example cannot speak or do not have a shared communal memory yet, accumulated over generations? We most likely could not as it would mean to imagine oneself as being a member of a small primate group such as chimpanzees still live in today.

Humans are not the only species capable of learning, other species can do this as well but not quite so well. Chimpanzees can learn a lot of things if provided the right incentives but in most instances, training or conditioning may be a better description than learning. A good number of other species also have shown capabilities to perform remarkable feats if trained properly. The real question is not if we can make animals do certain things in human shaped training set-ups, because we can. The question rather is how animal learning works in their natural environments. As it turns out, there are two ways to learn, one risky and one less risky. Put simply, the risky approach is to try something new out yourself, which is referred to as asocial learning. The less risky approach is to observe others doing something new and then, if it turns out to benefit them, to copy their behavior in a process referred to as social learning.

Many animal species are capable of asocial and social learning and not just the ones where one would expect to observe these behaviors, such as monkeys and apes; fish and even some invertebrates are capable of it too [68,69]. From a species perspective, it looks like as if a mixture where social learning dominates with a small complement of asocial learning added to ensure continued innovation seems to be the best strategy. There is one aspect of social learning that matters greatly and that is to ensure high fidelity copying. The more complex the observed behavior to be copied, the more important it is to copy exactly. As it looks like, the evolution of social learning and the mastery of copy exactly could also be at the heart of human culture evolution [69].

Clearly, the success of more complex social learning, such as for example required for learning tool-making techniques, very much depends on what can be communicated. Looking again at the animal world we can see that what can and cannot be communicated limits the complexity of behaviors that animals can learn through copying. Take the example of a chimpanzee who has learned how to expertly fish termites from tree bores by following a certain set of rules. What other chimpanzees can copy of that method is what they can see. If successful fishing for termites involves difficult to observe rotations of the stick that other chimpanzees cannot discern from a distance, how can they successfully copy this method? What chimpanzees can copy from each other is what they can observe each other doing and what they can communicate in other ways; and it is with the latter where the limitations quickly become clear. It looks like that this limitation may shed also some light on human tool culture development.

However, things are not that simple. Under the umbrella of social learning, we can find a very broad range of different types of learning; that is one side of the coin. The other one is our capability to decipher what exactly is being learned and how it is done. Clearly, social learning among fish has to be different from social learning among primates. For primates, social learning among the great apes or hominids also varies as most likely our ways of social learning differ quite a bit from those of chimpanzees or bonobos to whom we frequently compare ourselves. Voluntary or involuntary communication is a key ingredient in all types of social learning. For fish, there is no vocal communication, and most would likely agree that fish have no facial expressions either that another fish could interpret; or that we could read in any way for sure. That leaves two other avenues: observation of behaviors that lead to a certain result and the observation of the result itself. A fish observing other fish could copy their behavior or could copy something similar to the observed behavior of other fish that leads to the same result.

While primates add to their social learning repertoire the observation of vocalizations and of facial expressions, both of these still serve the same end: to copy the behavior that leads to the desired result or to use their own variation of the copied approach to reproduce the result. As it looks like, primates seemingly less copy behaviors but rather try to reproduce the intended results of observed behaviors however they can. For chimpanzees, it is more the observation that by sticking something into a rotten tree one may be able to catch a termite rather than exactly fashioning the same type of instrument and using it in the exact same way as they observed another chimpanzee do it. They are not simply copying the method but instead seem to adapt the idea and make it work with what tools are available to them or can easily be fashioned. Would we talk about human behavior here we would speak about copying of intentions and ideas, in this case the fishing for termites, rather than copying the techniques themselves. Why should the same description not be adequate for chimpanzees behaving in the same way?

Increasingly, science is indeed moving in this direction, giving credit where credit is due, regardless of whether it concerns human or chimpanzee behavior. However, humans do add significantly more complexity, in essence a completely new dimension to the palette of social learning that we can observe throughout the animal kingdom. This extension of social learning beyond what we find in the animal kingdom is uniquely human and best called culture learning. At the heart of culture learning is teaching and the kind of teaching culture learning requires is not possible without much improved communication. Social learning as embodied by human culture evolution, where learning is shared broadly and available to all members of a group to improve on, is the mechanism that enables cumulative culture evolution through a process often described as the ratchet effect.

When looking at fig. 4.1 we can see that the early tool cultures extended over much longer time spans than the later ones. Almost a million years separate the first evidence for tool making from the earliest records of the Oldowan tool culture. It took around 500,000 years for the next phase in tool making to emerge, the Acheulean tool culture; and it took almost another two million years before the Levallois tool culture started to replace the Acheulean tool culture. Why did this take so long? One could rightly object to this question that we just do not have any idea of how long such processes are supposed to take. There seems to be nothing even remotely similar to human culture evolution in the evolution of life on Earth, nothing to which we could compare it. However, it still makes sense to ask this question in a relative context as the archeological records show that the speed of human culture development as seen through the lens of stone tool making eventually increased. Stone tool development seems to have been going quite slow for a long time and then for some reason started to accelerate. Human culture development certainly does not progress linearly.

Nothing shows this better than our most recent past when we compare the pre-industrial age to our modern societies. There seem to be thresholds in human culture development that once crossed, open the path to much faster human culture development. Why should something similar not have been true for our human species distant past? It just may be that humans were stuck so long in the early phases of stone tool technology because in order for humans to get better at copy exactly at a more advanced stone tool technology level, they first had to learn how to communicate better so they could achieve copy exactly for more complex tool making processes. Therefore, a plausible environmental force fostering the development of language could have been the need for transmission of tool making knowledge to cross the threshold to improved tool technology and making. Any improvement beyond the vocal communication our chimpanzee relatives are capable of, would have resulted in an improved human ability to more efficiently transfer tool making skills.

To draw an analogy we can look at the beginnings of the industrial age, as in this instance a different kind of language enabled human cultures to make the transition from agrarian societies to highly industrialized modern societies. It was the language of science that humans had to develop first in order to cross the threshold that separates pre-industrial from industrial societies. In essence, the rise of modern science exemplifies nothing else than the ratchet effect of cumulative culture evolution at work. Once a language started to emerge, the language of science, enabling the communication of new scientific discoveries in a reliable and transparent way, it became much easier for scientists to build on each others achievements, regardless of how far apart they may have lived. All what was required to participate in the new human science endeavor was to learn this new language, which we continue to teach in schools today with the objective to enable our children to understand science as well as participate in and contribute to it.

In a similar vein, humans had first to develop the primitive roots of language to cross the threshold from Oldowan towards Acheulean and then into Levallois-Mousterian type tool making. Surely, other group needs or factors within a group could provide additional incentives for developing communication beyond what chimpanzees could and can accomplish. However, putting the necessity to teach more complex matters at the center of human language development and culture evolution, rather than for example where to forage for food, makes sense [69]; after all, if foraging for food would be a key driver we could expect many other species to also have developed higher communication skills.

With the onset of language and the ability to learn a shared set of limited vocal expressions, not only the capability to teach would improve but it also would stimulate the development of shared communal memories. Culture after all, is first and foremost socially inherited so it needs to be internalized and memorized to be passed on. Only when we look into how every infant human first has to learn our culture and how long human adolescence learning lasts, can we begin to appreciate how different human culture learning is from the social learning that we find among primates. Becoming human is a very different process than becoming a chimpanzee. While there are clear parallels between becoming human or chimpanzee, there are many more significant differences. In a sense, though not literally, infants recapture the essence of what it means to become human. The process of becoming an adult member of a human group must have evolved over a long time and the time it took to become a human adult in the very distant past was much closer to the time it still takes to become an adult chimpanzee, that is, it must have been shorter. This is actually something we can observe with ancient humans where the skeletons of young adults tell us that they matured earlier. To phrase it colloquially, programming a modern human to become a fully functional adult takes longer than for example it took to program a young *Homo heidelbergensis* to become a mature adult; and in turn, that took longer than to program a young chimpanzee to become an adult. The skill-sets that humans acquired over the few million years since chimpanzees and humans split from a common ancestor make us who we are today, and in their current modern form, they need to be relearned with each and every generation.

As human culture evolution progressed, it would have provided an edge in the Darwinian evolutionary struggle of the fittest and it would have been selected against, thus favoring populations that evolved towards more increased social complexity. Social complexity needs to be seen here as a relative measure as it is not supposed to convey anything close to what we see in our modern societies. But compared to the social interactions we see for example in chimpanzee groups today, human groups with rudimentary language and teaching skills would indeed be much more socially complex. The process whereby genes and culture interact over long periods of time in such a way as just described for language and communication is referred to as Gene-Culture Coevolution, a concept first introduced in the 1960's.

If the interaction of genetic evolution and culture evolution sounds more Lamarckian than Darwinian that is no coincidence because this is exactly what it is. There is still a Darwinian selection process at work but the dominant aspect in Gene-Culture Coevolution is clearly the culture part. We are not so different from humans that lived a long time ago than we would like to believe. There is a deep truth behind the saying that civilization is only skin-deep. Culture is a social concept, a learned concept. We only have culture because we inherit huge amounts of knowledge as well as implicit rules and norms through the societies, we as individuals are members of. Without such transmission, we never could be full members of our societies and all of us would be quite different individuals than the ones we are today.

There are of course alternative explanations that have been put forward as to how language has developed. Prominently among them is language becoming a necessary replacement for the social function of grooming in larger groups. As the argument goes, with larger group sizes individual group members would have to spend increasingly more time on grooming each other to reassure themselves of their respective standing within the social hierarchy of the group. Because of that, we supposedly figured out that gossiping could do the job as well while freeing up much time for more essential group needs that otherwise would have gone into grooming; foraging being one of them. The argument has some charm but not enough to buy it. While grooming serves a fundamental social function among our primate relatives it also has the very basic function to rid ourselves mutually from pests that invested the hairy skins of our earliest ancestors no less than they still plague primates and sometimes humans today; we are talking about lice. Therefore, we can safely assume that groups of early humans continued to groom each other for getting rid of lice as well as for social cohesion and comfort.

We do not know anything about the group sizes early humans lived in. However, we can assume some continuity and that it may have taken quite some time before larger human groups became viable. Not so much from a social cohesion perspective but from the simple fact that group sizes must have been constrained by what habitats could support in terms of foraging. It is reasonable to assume these constraints were not too different from those that our primate relatives still are subject to today. That may have limited groups size and only when humans had developed new hunting and foraging techniques relying on tool use, could larger groups sizes likely have been supported.

A turning point, literally, may have been humans figuring out how to turn tools into weapons that allowed them to defend carcasses of larger prey animals they either had claimed from another predator or had developed ways to successfully hunt and kill on their own. We certainly cannot verify this, just as we cannot verify if language developed from the need to replace grooming to be able to live in larger groups. However, the above arguments make it more likely that humans had to develop efficient tools first before they could support larger group sizes. That in turn could have been done most effectively if humans had developed ways to ensure tool-making skills could be passed between generations such as through copy exactly; and as tool making became more sophisticated, learning through observation would have to be assisted by learning through communication. The primitive language that could have met the need to communicate tool making and other skills vital to group survival may have persisted for a long time, similar to the early tool cultures persisting for long time spans.

Language may have evolved only very gradually but also may have seen faster development spurts in-between, and increasingly so as human technology progressed, human group sizes grew, and maintaining the social fabric of such larger groups required more complex communication. Because of all that, communication between human groups may have become more important. There are many different language development scenarios that can be drawn up, some more speculative than others. Unfortunately, critical pieces of the puzzle are missing, such as an understanding of what kind of vocalization early humans were capable of. Our primate relatives do communicate vocally but in a very different way than we do, and this is first and foremost due to the lack of a vocal apparatus that would allow them to articulate themselves differently than they do today. Therefore, even if primates possessed a richer inner language to express themselves, they could not articulate it. The speed of human language evolution must have depended on the speed with which the evolution of the human vocal tract progressed. Unfortunately, we know little to nothing about this aspect of human evolution for the early members of the human species.

What we know is that the human capability to develop speech requires the presence of a specific mutation of the gene FOXP2. This gene is sometimes also dubbed the language gene but that is an oversimplification as there are other genes involved in speech development. However, the presence of this gene mutation is still a useful tracer because if it is missing in a human species genome, it is unlikely that the respective species had language capabilities in any sense that we would recognize. Neanderthals possessed the same FOXP2 gene mutation we do [70]. Now we can speculate if modern humans and Neanderthals inherited this gene mutation from their common ancestor - *Homo heidelbergensis* or an earlier form of this species for which we have no fossil record yet – or if they developed this FOXP2 gene mutation independently. Clearly, the odds that the very same mutation happens to occur independently in two species descended from a common ancestor are far less than those two species inheriting the gene mutation from their shared ancestor. As discussed in the previous chapter, the latter would imply that this shared gene mutation had been around for maybe 600,000 years or longer. However, there is more. Denisovans also had the very same FOXP2 gene mutation, which should not be surprising because they, like the Neanderthals, also descend from *Homo heidelbergensis*. Hence, a good argument can be made that *Homo heidelbergensis* already possessed the FOXP2 gene mutation. With that it gets even more interesting. As hypothesized in the previous chapter, *Homo heidelbergensis* proper – the one we have a fossil record of - and the ancestor of modern humans may have split off from an earlier version of *Homo heidelbergensis*. If so, that would make it likely that this early *Homo heidelbergensis* already possessed the FOXP2 gene mutation. As we have seen in the previous chapter, a likely candidate for the early form of *Homo heidelbergensis* could be *Homo antecessor* who is seen as related to *Homo erectus ergaster*. We cannot really know this yet, but what becomes clear is that the FOXP2 mutation critical for developing our human speech capabilities may have evolved very early.

Following the arguments above, it could have been that the FOXP2 gene mutation may have evolved with *Homo erectus ergaster* or *Homo antecessor*. If that was indeed so it could actually have been part of the human package that some 1.9 million years ago enabled the evolution of Acheulean tool cultures. Alternatively, in the case that modern humans split off from *Homo heidelbergensis* proper at a later time it may be linked to the emergence of Levallois tool cultures. As indicated above, Levallois tool making required the toolmaker to actually envision the final shape of the tool inside the stone core from which it needed to be released. Clearly, such a skill is much easier to communicate and learn if the "how-to" can be communicated through language assisted copy exactly. The evolution of language in some way seems to be tied to the evolution of human tool technologies that we can track in the fossil record. However, it likely is linked as well to societal developments as interaction among humans and between human groups must have become richer and more complex over time.

With respect to special language genes, a word of caution is necessary. Humans are not the only animals where a version of the FOXP2 gene enables in a broad sense what one could call communication. We know today that songbirds without their FOXP2 gene version cannot sing. Some mammals also need it for vocalized communications such as whales or mice for example. If the FOXP2 gene in those animals is in some way blocked or its function impaired or restricted, their communication capabilities suffer accordingly.

Therefore, when we ask the question "Do animals communicate?", the answer should be a resounding "Yes" through analogy to how the FOXP2 gene is critical to human language capabilities. Admittedly, this analogy oversimplifies matters, but that is not relevant because we are not interested in the specifics here but rather in what questions it raises once we concede that animals communicate using their versions of the gene that is also key to human communication. The logical follow-up question is what do animals communicate about? Is it all about utility, like where to best forage for food, or does it also include more personal notes. Dangerous territory for any biologist to explore only a few decades ago but not so anymore. If animal communication includes content not related to food or advertising one's genetic makeup for procreation but simple things that confer for example a sense of security or fear, comfort or interest, how can we not ascribe them their equivalent of emotions? While human language as a communication tool seems to be unique in the versatility and differentiation it allows us to communicate with each other, we should acknowledge that there are other languages animals use to communicate among themselves. Yes, humans are special in the sense that we have used this capability to build societies, cultures, and technologies, something no other animal ever did or ever was even in a position to try. However, we are only different in degree in how we can communicate and cannot claim to be the only species that talks. We just have not managed to understand yet in detail how animals talk to each other and are as far from deciphering their languages, as they are from understanding ours.

Frequently, human culture evolution was set in contrast to animal evolution, specifically the evolution of our closest primate relatives. In many ways, this has helped us understand where, as a species, we are different. However, at the same time it also has brought us closer. Closer in the sense that it is becoming increasingly clear that we may not be the only species in the animal kingdom capable of emotions and being able to articulate them. Maybe no less than a generation ago it would likely have been heresy to suggest any animals had emotions. Animals were deemed to have instincts and behavioral patterns, essentially biological machines, but no emotions as those were reserved for humans. Not so anymore. It is safe to say that primates most certainly do have emotions and many other animals as well. That of course poses a problem. How can we treat animals as we do, rather cruelly to put it diplomatically, when they are not machines but living beings with emotions, maybe with a different complexity than ours but emotions just as well? It looks like we have to take another step in our culture evolution and acknowledge that we are not the only animal endowed with feelings but that other animals also can be happy or sad, can feel joy and pain. Today we know that this is the case for our primate relatives. However, other animals may have feelings as well, most of us certainly would ascribe them to our animal friends who keep us company. Is it not interesting how we humanize animals, readily endowing them with feelings in the stories we like to tell our children? However, the same adults taking their children to movies featuring animals which clearly have emotions, seemingly have no problem with our industrialized societies torturing animals by the millions. Many aspects of human culture evolution we do not understand yet. For example, how humankind acquired its dark side including our treatment of other animals and our general disrespect for other life. Maybe we really are not too keen to understand where this comes from as we may be somewhat reluctant to look into the mirror and acknowledge that those human traits are also part of us.

In a sense, the roughly three million-plus years of human pre-history can be looked at as the prolonged kindergarten of humankind. Like our children learn many of the social skills individuals and societies depend on alike, our human ancestors became the first social animals capable of building large societies that eventually enabled the development of our modern world. Of course, other animals live together in large numbers as well, such as ants or bees. However, while their states are certainly complex, they are not social constructs such as we perceive human societies to be. An ant state can work only in one way; not so a human state. A worker bee can never become a queen; not so for humans where on occasion workers have risen to rule societies. The dynamics of an insect state are strictly ruled by biology, not so in the human case where it is as much the social strength and cohesion of different parts of a human society that determine which rules are acceptable and which are not. True or not? Do we need to question as to what extent this is actually true, not so much in the case of the insect states but in our case? We keep finding more and more evidence that many of our complex behaviors, individually and as societies, have their roots not so much in what we like to call our free will or our human rational decision-making capability, even if we like to believe so. Rather, it looks like the biological underpinning of those concepts is much stronger than what we would have freely acknowledged not too long ago. In a sense, our cultural constructs resemble our biology much stronger than what our sophisticated societies seemingly suggest. We will come back to that when we take a closer look at our tribal instincts.

Regardless, there is no analog for human culture and societies in the animal world. Human culture evolution, the development of our civilizations and the impact this had and continues to have on life on Earth is a non-repeatable experiment. This has not always been so and there may have been different outcomes. There could have been scenarios where one of the other human species succeeded and replaced *Homo sapiens* or where several human species build civilizations. However, that is not what happened. This evolutionary outcome and not any divine plan made us humans the very special animals we are. But this is where we need to be careful. We have nothing that we can compare human evolution against; we have nothing we can compare human culture evolution against, nothing that we can benchmark our societies against but the values and rules that we put in front of ourselves. Instead of being able to learn more about ourselves by comparing our past and continuing evolution against that of other human species, we are doomed to learn about ourselves through the species equivalent of introspection. We are stuck with us.

The Promise of Civilization

According to ancient Greek mythology, the Titan Prometheus created man from clay, stole fire from the Olympian Gods to give it to humanity, and by doing so kindled human civilization. Other cultures have their own stories about the emergence of civilization but inadvertently it tends to be either a voluntary or an involuntary gift from gods or from nature and animal spirits, or the heaven in general. These creation myths tell us that humans were marveling at civilization from its earliest beginnings and deemed it an accomplishment that could not have been a uniquely human creation itself, it had to come from a higher power. We have come a long way since then but even though, many of us may still have similar feelings. That is understandable. What humans achieved through civilization leaves us in awe and in horror at the same time. From the towering peaks of our greatest achievements to the bottomless depravities human civilizations have committed many times throughout our species history. We struggle to understand how we are seemingly capable of both, how the one can be so close to the other. How can that be?

It is with modern humans emerging from the mists of pre-history that we witness the beginnings of what we commonly call civilization. This happens in the Bronze Age in several places as evidenced by the archeological records and the histories of the ancient civilizations of Sumer, Akkad, Babylon, Egypt, the Indus Valley, and the Yellow River; to name just the most prominent ones. Later followed the civilizations of antiquity, all of them borrowing to a greater or lesser extent from the ancient civilizations they were familiar with while adding their very own contributions. These civilizations of antiquity, as we refer to them here regardless of their geographic origins, include among others the Mediterranean civilizations of the Greek city-states and the Roman Empire in the West, as well as in the East the empires of the Maurya on the Indian sub-continent and the Han Dynasty in China. For many of us, much of this is familiar terrain, remembered from high-school history. These civilizations represent the cultural traditions and values that many of us are still rooted in today, in the West as well as in the East.

There were of course other early civilizations, but their traces are harder to find today as they were consigned to oblivion because they succumbed to internal strive or were conquered by forces more ruthless and powerful than they could muster themselves. The civilizations of Middle- and South America, only rediscovered in modern times, come to mind. As with human evolution itself, there have been winners and losers among civilizations. Similar to extinct human species contributing to our biological heritage, once flowering but now vanished civilizations have left their marks in our cultural heritage, sometimes they are just hard to see as we often tend to perceive them as our very own cultural heritage to begin with. Throughout history, all civilizations have borrowed or inherited some of what they eventually came to perceive as their very own unique cultural heritage and contribution to human civilization.

Like any other quest for knowledge, the study of our own past continues to produce new and unexpected results. Therefore, it should come as no surprise that our picture of ancient civilizations and of our human heritage is getting more differentiated as we learn about bygone civilizations which followed different trajectories and for reasons, we do not fully understand yet, vanished from history as they were replaced or merged into other civilizations. The ancient civilizations we are all more or less familiar with from our school days have one common characteristic. If we look to the civilizations of Sumer, Akkad, Babylon, or ancient Egypt, they were all elite dominated societies. These elites demonstrated their power and entitlement to rule through buildings and monuments, proof positive of what they may have seen as their unquestionable leadership as quasi-gods, thereby quite visibly preserving remnants of their civilizations through the ages well into our time. The visible heritage of these civilizations is still part of our modern civilizations. It took archeology a bit longer to discover other civilizations that did not leave us lasting monuments such as the pyramids but who may be no less important for understanding the possibilities of human civilizations and cultures. It is only in recent years that we are learning more about such civilizations like the Danube Valley civilization, sometimes also referred to as the Old European civilization [71].

The Danube Valley civilization existed between the sixth and third millennium BCE and its area covered much of the Balkans and Greece, extending eastwards well into what is today Ukraine. Scholars have come to understand that this agrarian civilization seemingly was essentially egalitarian with no evidence for dominating elites. Their burials indicate no differentiation between rich and poor; no traces or artifacts were discovered that could be associated with elite status either individually or by belonging to an elite group. This Old European civilization seems also to have been characterized by a genuine gender equality rather than being a male dominated civilization like the ones we know for example from Mesopotamia or ancient Egypt. There are indications that this agrarian society had an evolved division of labor. While this included a partitioning of chores between what most would even today consider more male or female activities, there seems to have been no dominance of males over females or the other way around. Interestingly, this being an agrarian society in the fourth millennium BCE, we also see large cities in what is today western Ukraine and northern Moldova that were home to likely more than ten thousand people, essentially the largest cities we know from this time, including the city-states of Mesopotamia [71]. While these cities seemingly had some multi-storied buildings, they were cities without monumental palaces or temples and because of that their traces were harder to find than for example the ziggurats of Mesopotamia.

Today we know that this Danube Valley civilization was highly innovative and sophisticated in its technology use like metal smelting and ceramic making, including the invention of the pottery wheel. Excavated art objects include beautiful and expressive figurines, vases, and cult objects, the latter indicating religious beliefs in natural deities but with no evidence for a powerful or organized priesthood. While still being debated, the script of this early European civilization which has not been deciphered yet, could be the oldest writing we have from a human civilization with some clay tablets dating back to around 5,500 BCE.

To some extent, we can count the Indus Valley civilization of the early third millennium BCE as one of these more recently discovered civilizations. It also does not quite fit the mold of the elite led societies of the much longer known Mesopotamian and Egyptian civilizations. Different from the Mesopotamian and Egyptian civilizations, the Indus Valley civilization, like the Danube Valley civilization, did not build a unified empire but much more resembled an integrated economic and trade network. The people of the Indus

Valley civilization built more than one thousand settlements along the fertile planes of the Indus River Valley. We know today a lot about their sophisticated trading methods including the prospering trade between the people of the Indus Valley civilization and trading centers all around what we today call the Persian Gulf. The latter included the trading ports of the famed land of Dilmun, another civilization we only started learning about more recently. For the Indus Valley civilization, we know it vanished because of a succession of pro-longed dry periods. These resulted in the degradation of what were once fertile agrarian plains between the Indus and Saraswati rivers as the latter dried out completely. Unlike for the Danube Valley civilization, we also have some indication as to how the heritage of the Indus Valley civilization still lives on in the rich cultural traditions of today's India, mostly in how the religious traditions of this ancient civilization have influenced both Hinduism and Buddhism.

In the second half of the last century, archeologists made a number of discoveries in the historical lands of Anatolia in modern-day Turkey, which very much have reshaped our perspective on human pre-history. Sometime between the tenth and the seventh millennium BCE, hunter-gatherer cultures built what historians interpret as a major temple complex in Göbleki Tepe, a place in eastern Anatolia [72]. What makes this so remarkable is that this happened at a time in human pre-history where there were no cities yet. The most massive stone pillars used in building the Göbleki Tepe complex weigh up to fifty metric tons. In the past, these kinds of engineering feats where only associated with high cultures such as those of ancient Egypt or Mesopotamia that succeeded in state building, but not with hunter-gatherer societies. All of that tells us that we have to rethink our perceptions of what we believe were the first cities of humankind and what they looked like. From what we know today, our ancestors have been building cities well before the city of Uruk in Sumer which is still on record as the first real city dating to about 6,000 years ago and in its prime time in the early third millennium BCE may have counted several ten thousand inhabitants.

We already mentioned the pre-historic city of Çatalhöyük in southern Anatolia. The British archeologist James Mellaart (1925 – 2012) and two of his fellow British archeologists discovered its location in 1958 and Mellaart led the team that started to excavate the site in the first half of the 1960's. Before the discovery of Çatalhöyük, archeologists knew of no Neolithic artifacts in Anatolia. The Neolithic seemingly had bypassed Anatolia, which posed a problem for archeologists as they struggled to explain how the *Neolithic Revolution* had spread to Europe. This lack of archeological evidence for the spread of Neolithic cultures into Europe was explained away by the hypothesis that Neolithic people had moved there from the Fertile Crescent along the shores of the Mediterranean. However, this was not a good explanation.

While it is possible that the spread of Neolithic culture to Europe could have happened in this way, it always remained a riddle why Neolithic culture and people had not moved into Europe through the Anatolian plane. With the discovery of Çatalhöyük it became clear that the latter indeed had happened. Unfortunately, excavations at Çatalhöyük stopped for more than thirty years for reasons we cannot go into here. In the 1990's archeologists started to dig again at Çatalhöyük, now under the leadership of the British archeologist Ian Hodder (born 1948) and excavation continues to this day. Since then, Çatalhöyük has certainly lived up to its reputation as a Neolithic treasure trove.[7]

The beginnings of the Neolithic city of Çatalhöyük date back to 9,500 years ago and archeologists believe it has been continuously occupied as a city for about two thousand years until about 5,600 BCE [73]. It was a different city than Uruk for sure with no streets as houses budded up against each other sharing walls and having no entry doors as they were accessed through openings in the rooftops, which seems to have been the only entrance these buildings had. Çatalhöyük, populated by a few thousand people with an estimated peak population of about 8,000 inhabitants, also had no palaces or any other outstanding monuments, as all buildings pretty much looked alike. This of course would indicate a more egalitarian society with no pronounced elites and given the way it was built it must have resembled more of a beehive than a city. But it was a city of thousands of people predating the city of Uruk by some three thousand years.

There is still a debate whether the people of Çatalhöyük may have worshiped a female deity or not. Such views arguing for evidence of goddess's worship and matriarchal societies started with Mellaart himself who thought he found evidence for what he called a mother goddess and later have also been suggested by other archeologists, such as the Lithuanian-American Marija Gimbutas (1921 − 1994). However, we have to see such debates in the context of their own times. In the final decades of the last century women started to more forcefully assert their equal rights that our morals and laws tell us they do have just as men do but that societies to some extent even today continue to deny them. It is not surprising that a reevaluation of the role of women in society also would encourage a reassessment of gender roles in our more distant history. Of course, a natural part of that is the search for matriarchal societies and goddess worship in pre-history. However, since then archeologists have come to realize that it is very difficult and sometimes impossible to infer cultural practices from the artifacts they find in pre-historic sites. Today, it is likely more accurate to speak about a prominent representation of female artifacts that the people of Çatalhöyük have left us and acknowledge that this leaves a number of interpretations, none of which we can be certain of. Among others that may include the possible use of such figurines as toys as well as for worship. However, there seems to be little if any evidence now to support a mother goddess as central to the beliefs that the people of Çatalhöyük may have held.

Çatalhöyük has redefined in many ways how we perceive this time of pre-history, just a few thousand years after the onset of the *Neolithic Revolution* in the Near East. Among the treasures it has passed down to us are the very first paintings on human made canvases. Archeologists had always wondered about the large gap in time between the spellbinding cave paintings our ancestors left us in the Upper Paleolithic in places such as Altamira and El Castillo in Spain and Lascaux in France and when we find humans in Eurasia painting again in the Bronze Age. The frequent explanation was that in those regions humans somehow had shifted their mode of expression from painting to other art forms while their cousins in Australia for example continued with their rock art. The beautiful mural drawings the people of Çatalhöyük left us on the plasters of their house walls have proofed this view wrong and we now can justly assume that humans in Eurasia likely continued to paint throughout the Upper Paleolithic into modern times. It is just very difficult to find the evidence for that as human paintings are fragile and without the protection of a cave climate or from burial under successive layers of civilization such as in Çatalhöyük, they are easily lost forever.

There are many more examples of early human civilizations which archeologists and historians are uncovering, all pointing to a much greater diversity of human societies. The elite and male dominated societies of our historic period date only back a few thousand years ago. As they are pretty much the only ones that continued into our times, much reinforced we can add by the monotheistic and clearly paternalistic religions arising in the Middle- and Near East, it is no wonder that for a long time we took this kind of civilization as the de facto human society model. That this is a biased perception of the evolution of human civilization is now becoming clearer. As scientists discover that the past had quite different models for human civilizations aside from these elite-based ancient societies we are so familiar with, we are also witnessing our modern societies reshaping, becoming more egalitarian in many ways, like finally including gender equality; however, elites are not giving up their presumed prerogatives easily. More likely than not, there is a reason for civilizations evolving into elite based societies. In the best case, this should be meritocracy at work; however, more often than not, this is not the case. As we will see a bit later, holding elites accountable continues to be a major problem that plagues practically all modern civilizations to a greater or lesser degree. However, for now, back to the promise of civilization.

Each one of us is anchored in a civil society that likely can trace its roots many hundreds or even thousands of years back in history. Science has shown us that humans across the globe are much more alike than we ever knew. Unfortunately, too many among us still cannot accept this and keep being misled by our different appearances which are superficial. Something similar seems to be the case for human civilizations and cultures. Outwardly, civilizations and cultures can express themselves in very different ways; this is what we justly perceive as the richness of human cultural heritage. Then, on closer inspection, we discover that much of the fabric woven through our civilizations and cultures is part of a shared heritage that belongs to all of us. We are often reluctant to acknowledge this as it seemingly takes away from the uniqueness of our specific kind of civilization and the culture through which we express it. However, that is a misguided perception as it really does not diminish our own specific civilization, our culture, or our very own uniqueness as one of its members. All it does is to point towards something greater we all share in, a fabric of human societies connecting us to the very essence of what it means to be human but which we usually only notice when we look at it in a larger context.

Part of this fabric are fundamental human values and ideas that deep down we all believe in. They may have originated in one specific society first or they may have emerged in several societies, eventually to become part of the cannon of values we all believe civilizations should embody. Maybe not all to these values carry the same weight in all societies. Some of these values occasionally must be balanced against each other and not all societies chose the same balance. Prime examples for such balances are the rights of the individual versus the needs of society at large. There is not one single solution for how to adjust this balance but a range of different outcomes depending on the choices made by those governing society. Sometimes with the consent of the governed, sometimes not, but usually grown out of cultural traditions going back many generations. Western societies shape themselves much more around the rights of the individual while Eastern societies place a higher value on the needs of the community. As long as neither of them compromises fundamental human rights, whatever balance the freely elected representatives of a society choose should be respected. Countries where citizens are free to stand for election and can vote freely are still a minority. However, even in quite a few of these countries which technically meet such requirements, powerful interest groups have shown themselves well able to co-opt politicians and governments and thereby put their interests above the will of the voters. Seemingly, the struggle is never over and social contracts within and across nations must be continually renegotiated and reinforced.

The quest for improving ourselves as humans and for creating better civilizations has spawned religions, utopias, and has cost the lives of countless millions. Many of the utopian visions remained confined between the covers of a book. However, others became realities, often quite perverted ones; and still others are at the very heart of our worlds major religions. Among the classical utopian visions, we can count some of the most important works of antiquity. This certainly includes *The Republic* by the antique Greek philosopher Plato. It also includes several later works of the Renaissance such as *Utopia* by the English cardinal and statesman Thomas More (1478 – 1535) published 1516. Other still later works that come to mind are *The City of the Sun* by the Italian Dominican friar Tommaso Campanella (1568 – 1639) published in 1623, and *The New Atlantis* by the English philosopher and statesman Francis Bacon (1561 – 1626), published in 1627. The first important Renaissance work on the philosophy of state since antiquity, *Utopia*, gave the whole genre its name. The utopian visions cited above seeded many later works, usually categorized under utopian socialism, right up to the utopic visions of the English writers George Orwell (1903 – 1950) and Aldous Huxley (1894 – 1963). The French philosopher Jean-Jacques Rousseau (1712 – 1778) also stands in this tradition with his *Du contrat social; ou Principes du droit politique* (On the Social Contract; or, Principles of Political Rights), published in 1762. To a lesser extent this is also true for the much earlier work on social contracts and statecraft by the English philosopher Thomas Hobbes (1588 – 1679) with his *Leviathan, or the Matter, Forme, and Power of a Commonwealth, Ecclesiasticall and Civil*, published in 1651, which conjures up a much darker vision of an utopian state. Translating utopian vision into reality is fraught with all kinds of dangers and most such attempts do much more harm than good, but not all of them.

One vision of a better society and an important milestones in the history of modern human civilization is the constitution the thirteen North American British colonies gave themselves as they were fighting for independence. In a sense, it is likely the most perfect imperfect constitution a social body ever gave itself. Imperfect because it was not all-inclusive as the rights of the native population, colored people, or women's rights were clearly not part of it. Perfect, because it provided the framework to achieve eventually all of these. Amended versions of this constitution eventually rectified its birth defects. However, the spirit of the constitution prevailed against all the dark currents opposed to these changes only after many generations. With some justification, one could maintain that even today, the promise of this constitution has not been fulfilled. Interestingly, at the time of the American War of Independence the country the Americans were fighting against, already had become a constitutional monarchy. Voting was still very much restricted as the landed nobility was holding on to is privileges but over time, and for the most peacefully, the British parliament became more representative of its population.

However, in Britain like in many other countries it would take until after the First World War before women received the right to vote. On the European continent, the French revolution in 1789 marked the beginning of the end of absolute monarchies in continental Europe and Russia and the rise of modern imperialism, nationalism, and extremist ideologies. The result were numerous wars throughout the nineteenth century right up to the Balkan Wars in 1912 - 1913, before dragging Europe and the rest of the world into the abyss of two World Wars. Only after living through the atrocities and mass murders by national socialist, fascist, and communist regimes would most of Europe eventually enjoy the freedoms of democracy, roughly two-hundred years after the French revolution. Human civilizations, at least the vast majority of them, have gone through greater transformations since the industrial revolution than over the many millennia before it. The speed at which human technologies have and continue to change our societies is seemingly overwhelming. So are the consequences of these changes, manifest in the convulsions our societies have gone and continue to go through, driven and enabled by technology. The seeds of racism and nationalism must have been always there. Before the industrial revolution, on average, the furthest people could expect to travel in their lifetimes would have been the next village or so. For most back then it would have been very unlikely to encounter people markedly different from themselves. Technology has very much shrunk our world while it has done little to change our innate fears. Ideas always had the power to change our world. However, with the changes brought about by the industrial revolution they could now spread much quicker and wider than ever before. The rampant spread of ideologies across the globe that we can witness in the twentieth century would not have been possible before the industrial revolution. In a sense, as the maladies of early capitalism and imperialism spread around the globe so did the ideologies that supposedly were to provide the cure. During the so-called Cold War that followed the Second World War, two very different ideas of civilization faced each other. We know the outcome. What we do not know is how long it will last.

By the looks of it, most of our world is already in the grips of a voracious consumer culture, well on its way to it, or longing to join it as soon as possible. Over the past three centuries, groundbreaking scientific discoveries, technology development, and mass manufacturing on a scale never seen before, have completely transformed the societies we live in and with it the planet that we call our home. Before the invention of steam powered technology, the only way to scale up production was to put more people to work. Until then, the simple limits to expanding production of goods were determined by how much a single worker could produce per day and how many workers could practically be put to any given job. The quantity of goods a workshop could produce was limited and product quality varied between workers according to skill level and proficiency as well as between individual products sequentially made by any single worker. All of that started to change with the practical inventions of the eighteenth and nineteenth centuries. Steam powered machines allowed not only much more product to be made in a given amount of time, but they also enabled the production of large amounts of goods with more consistent quality. Before the age of the steam engine, aside from the waterwheel, adding more horsepower to a factory literally meant that. With steam-powered machines, mankind had started to learn how to harness energy in a scalable way with practically no limit; although in hindsight we now know that there are limits.

Mass production requires faster and affordable modes of transport to bring those many goods to consumers, eventually worldwide. Before the age of steam engines, the fastest mode for overland transport was the horse. This could be a rider transporting something small, a horse drawn cart bringing produce to market, or a barge with heavier goods pulled by horses alongside a canal. Not counting homing pigeons or information transferred through light signals, nothing moved overland faster than a horse. Steam engines revolutionized transport as much as they did production, giving birth to the railroad age. Sea transport always had been faster than overland transport and for a while, sailing ships continued to rule the seas but eventually the clockwork reliability of steam powered transport also prevailed on the seas. Harnessing energy to scale production up to levels that we now associate with the industrial age was the first step. Using the same energy to provide fast and reliable transport to almost any market with consumers having the purchase power to afford those goods was the second step. The third and final step would be to find a way to let consumers everywhere know that there was something to buy that they either really needed or could be convinced to need. It was these early times of marketing and distribution efforts by the industry and its associated logistics that built the first large scale information network. Few today know about the great sales catalogs that people marveled at and from which they would select at their hometown stores what they fancied. After national came international marketing efforts to develop new markets for an ever-increasing number of products. Today's commerce and advertisement on the world wide web we call the internet is in many ways the culmination of an effort that started with the delivery of the first sales catalogs a few generations ago.
How much the modern industrial revolution and the technology revolution of the second half of the twentieth century have reshaped human civilization is sometimes difficult to comprehend. Personally, what always illustrated this change most dramatically for yours truly is a look back at my grandparent's lives. They were born in a world where automobiles were still scarce, and airplanes did not exist or were just on the verge of being invented. Back then few people had telephones, if any, and radio was mostly still unheard off; no television of course, and on the practical side no running hot water and in rural areas no electricity and no toilets inside the home; to say nothing about medical services. I never knew my father's father as he had passed away before I came around, but my other three grandparents all lived a long life. When they passed away, all of the above existed and people could afford their own cars if they wanted to, had telephones – land lines though because the cell phone invention was still a few years away - and television too; but likely more important than any of that was general health insurance providing access to affordable quality medical services for everyone. There were many more things available that a generation earlier were just unheard of. Not only the first personal computers but an educational system that had provided their children and grandchildren, including yours truly, a solid and rounded education; something they ever had only limited access to when they grew up themselves. What my extended family experienced in Europe over the past few generations has happened since then in many other parts of the world and it still continues to unfold in many countries, though largely out of view. For most of us, what we believe to be the known facts about our world is still stuck in many ways with what we learned in school. However, things have changed and even if our facts were correct then, they are most likely not so anymore.

We have to thank the late Hans Rosling (1948 – 2017) for opening our eyes to what really is happening in our world, the stuff that really matters but that somehow never makes it into the evening news. If there is a single book from the ones listed in the bibliography at the back that I would like you to read more than the other ones, it is Hans Rosling's *Factfullness*. You will not only enjoy reading it, but you will likely learn more important facts about the world you live in than by watching your evening news every day. One of my favorite facts has to do with world population. We all have learned to stare in disbelief at the exponential curve describing the growth of Earth's human population. However, that is quite outdated knowledge because from what the facts show, in many countries, decade-long efforts to limit population growth are starting to pay off. Through economic progress, access to birth control, a decreasing need for parents to have many children to support them in old age, and a number of other factors, population growth has already been limited. Instead of a continued exponential rise from today's roughly 7.8 billion, we can now expect world population to peak towards the end of this century somewhere around eleven billion people after which it will level off. Even more interesting than that is how our world's population will be distributed across regions. You can memorize the distribution of the world population today by using what Hans Rosling called the PIN code of the world. In 2011 the world's population crossed the seven billion mark and at the time our world's PIN code was 1-1-1-4: one billion people each living in the Americas, Europe including Russia, and Africa, and four billion people living in Asia including Australia. In 2100 the world's PIN code will most likely read 1-1-4-5: still one billion people in the Americas and Europe each, four billion in Africa and five billion in Asia. By the end of this century, we will have added another Asia in terms of population size. Critically, in 2100 more than 80% of the world population can be expected to live in Africa or Asia, up from about 70% in 2011. This 10% difference amounts to an additional three billion people in Africa and one more billion in Asia. The Americas and Western Europe will become an increasingly smaller percentage of our world's population. These huge demographic changes that we can expect to play out over the next three to four generations will have very significant consequences. Largely stagnant populations in the Americas and Europe, including Russia, an already large but now slowly growing population in Asia, and a quadrupling of the population of the African continent will require many adjustments on all sides. The economic and political frameworks we take for granted today are likely to change in dramatic ways too. Would it not be smart for all sides to start planning for this?

We have managed to talk about civilization now for quite a few paragraphs without giving a thought as to what we understand by it. How is this possible? The only sensible explanation seems to be that we all have an innate understanding and agreement what the word civilization stands for. That by itself is no surprise, as we all usually talk about civilization in plural form; and that is because there are many different kinds of civilizations. We need to reflect on that. The word civilization is rather a newcomer as this noun, coined in eighteenth-century France so it seems, is only a few hundred years old. Of course, there are many definitions out there trying to capture its meaning. Most simple definitions of civilization will agree that it describes a society whose development is driven by the governing needs of dense, urban environments that require more complex sets of agreements among its population in the form of laws, civil procedures, and accepted customs than what was needed before urbanization. The rules by which a civilization's social body governs itself require a consensus among its elites if it is to achieve any kind of stability. However, within that consensus, these rules can be subject to change and continued interpretation. For the Romans, to whom we owe the Latin word, "civitas" was the social body of its citizens, the "cives", bound together by the rule of law. That was actually how the word civilization first originated in eighteenth-century France, as a legal term. While there is certainly nothing wrong with this definition, it all sounds a bit dry and for most of us, it likely falls short of capturing the rich diversity of human achievements wrapped up in what we refer to as civilizations.

We need to focus for a moment on how we actually use the word civilization. Clearly, there seems to be a hierarchy of civilizations. Not because some are in any way superior to others but in the sense that they all share with and often inherit from each other. So maybe a good start would be to look at the lowest common denominator, which of course should be human civilization or human civilizations; even at this level, we do tend to use the plural form. When we use the expression human civilization, does that imply that there could be non-human civilizations? As we know, there is no such thing as a non-human civilization on Earth. So, what else could it imply? The most interesting implication is that we may see civilization as something of a shared characteristic of self-conscious intelligent lifeforms that have evolved to a level where they leave their earlier and more primitive cultures behind, just as we humans did. Therefore, it is essentially a species characteristic and that by itself marks another delineation. When and where do we refer to human societies as civilizations? This question brings us back to the dawn of recorded history, to Sumer, Akkad, Babylon, Egypt, the Indus Valley, the Yellow River, or the Danube Valley civilizations, all representatives of what we frequently refer to as river civilizations; we really do not refer to human societies before that as civilizations. Right at the beginning of civilization we already have the plural, multiple human civilizations evolving along the Nile, in the lands irrigated by the rivers Tigris and Euphrates, along the Indus and Yellow Rivers, or the Danube and its tributaries. From the very beginning of civilization geography has differentiated civilizations and that has changed little since then. Just think about the Mediterranean civilization, or the Atlantic civilization, or the civilizations of the West and of the East. Civilizations come and go but seldom vanish completely. Much of what we inherited from the Greek and Roman civilizations is still very much an active ingredient in Western civilization today while the same is true for Chinese and to some extent other Eastern civilizations with respect to the Han civilization in the East. Geography and cultural content put civilizations in contrast next to each other while science and technology seem to unite them.

Today, most of our world's societies are part of a shared industrial civilization that originated in Europe but over the past two hundred years quickly spread to the far corners of the Earth. Of course nowadays, seemingly everything is also more or less colored by what we casually refer to as nation-based civilizations; but that notion is a mistaken point of view. Most people would have no problem in describing what someone speaks of when talking for example about something like a Japanese civilization or a French civilization. However, with such notions it is really the color of culture more than anything else that we infer or refer to. So how does culture differ from civilization? In the not-too-distant past, both terms were used interchangeably but not so anymore.

Incidentally, today we refer to what came before civilizations as cultures. Some 43,000 – 28,000 years ago, we have for example the Aurignacian culture of the Cro-Magnon's. Later we have other such examples as the corded ware culture that we see emerging after the invasion of the Yamnaya pastoralists into Europe some 5,000 years ago; or the Hallstatt culture of the Celts about 3,200 to 2,500 years ago; and we also have still primitive societies today that anthropologists refer to as cultures but not as civilizations. Culture predates civilization and civilization sub-sums culture. Another factor that can be a divisive as well as an inclusive one in defining civilizations is religion. All major religions include various national civilizations as defined by their specific cultures, they transcend geographically separated civilizations, and they include industrial civilizations as well as barely industrialized or pre-industrial civilizations. At the same time, they also introduce new fault lines, religious ones of course that can split what otherwise is a homogeneous civilization with respect to culture or geography. Interestingly, politics seems not be a factor we usually associate with civilization. We all understand intuitively, more or less, what is meant by Roman or Han civilization but we will likely be at a loss to explain what a republican, aristocratic, or imperial civilization could refer to; and that continues to this day. Have you ever seen the phrases democratic civilization, socialist civilization, communist civilization, or fascist civilization being used? Nobody ever seems to have seriously referred to the now defunct Soviet Union as a soviet or communist civilization. As previously, that is before 1918, nobody likely ever referred to a czarist or imperial civilization. It seems that people in either case would rather have talked about the Russian civilization; this unique molding of Eastern and Western cultures that over many centuries eventually became what we today recognize as such. In many ways, we identify by the various shades of civilization the threads that connect the evolution of our ever more complex societies through time and space. Civilizations are what unites and divides us at the same time and at their deepest levels connect us all. Maybe this is why usually we do not have to explain to each other what we mean when we use the word civilization. However, there is another equally reasonable explanation. In a profound sense, we may not even have to explain what we mean when we use the word civilization as it will stand in stark contrast to what it clearly is not. The ancient Greeks would have used the term barbarism to define what they were not, or as we would say, what their civilization was not. We have come a long way since then and nowadays we contrast civilization with the bottomless depravities that humankind has itself proven capable of over the past several thousand years since the ancient Greeks. Deep down, we all have a keen sense and fear of that which is the antipode of civilization.

Civilization has shaped us as a species over the past several thousand years and continues to shape all of our lives as individual human beings today. However, it has not erased the millions of years of human and primate evolution that came before it. Tribal instincts still underlie much of human behavior. For us to be rulers in our very own home – of ourselves as individuals – we must acknowledge and understand our tribal instincts or risk being at their mercy. Like with most things, there is good and bad in tribalism. As much as we may want to, we cannot get rid of it, nor should we. However, tribalism must serve us and society at large and not the other way round.

From Tribal Cultures

Throughout history and to the present day, religious laws often have bound many societies more closely than secular laws. In western as well as in eastern societies, secular power was for much of history conferred by divine right. Be it the divine rights of European kings or the mandate of heaven that Chinese emperors based their power on. In a looser sense this is certainly also true for much of antiquity where a common belief was that the gods would favor those who were considered just rulers, although the definition of what is just may have been different from how we see it today. Even more so in ancient times when rulers frequently were god-kings as we can see it with the Pharaohs, or at least god-like, as the Mayan rulers were. At times, religion became synonymous with the state itself in as much we can talk about states in these early days of human civilization at all. An example for this in our more recent past is the Papal States, dominating much of Italy for hundreds of years and in a sense surviving to this day in the little enclave of the Vatican City. Islam was from its very beginning always walking a narrow path between religious and secular forms of government. On the one hand, there were outright theocracy or political systems completely infused and dominated by Islamic religious law. On the other hand, Islam also produced models of religious tolerance, allowing scholarship and science to flourish in the early days of Islam and thus saving much of the ancient Greek cultural and scientific heritage for humanity while adding its own achievements to it. Unfortunately, today's practice and interpretation of Islam is in many places, though certainly not everywhere, not one of tolerance and open mindedness. As for Christendom, its modern tolerance is still often limited and was only advanced by much of its leading clergy under duress and not freely. Sometimes one has to wonder what would have happened in Europe without Renaissance, Reformation, and Enlightenment. Maybe instead of democracies, we could still have clerics directly or indirectly ruling the lands as it continues to be the case in much of Islam. Unimaginable, how much poorer our world would be in this case; and how much richer our world could be if Muslim societies would also have found a way towards modernity within a reformed Islam. However, there is hope. Just like Christianity, Islam is not monolithic, and some Muslim societies are well on their way towards finding their path to modernity.

Whichever way we may look at it, from the very first time one human being tried to raise himself and his family over his fellow human beings, more likely than not the justification was rooted in religious or mythological beliefs.[8] Asserting power over others was justified through belief systems. However, the basic human motives and motivations behind the grab for power may not have been much different from today. Frequently, the divine ruler and the high priest were one and the same person though over time that changed. Before humans became herders, pastoralists, and eventually with the start of agriculture settled into larger communities, the role of a leader was likely quite different. Human hunter-gatherer groups that staid continuously together were small, maybe not bigger than a few extended families in size. These groups were, as far as we can understand, more egalitarian than what would follow later but with close to certainty, we can assume that they had their own hierarchies. Only those who were best able to ensure the survival of a group would have been likely to gain leadership. We must assume that those groups also had deeply felt belief systems to make sense of the world around them. From burial rites that go far back in human societies, we know that connecting to the ones that came before must have been a central part in the lives of early human family groups. These small hunter-gatherer groups had their own intermediaries to help them connect to the world beyond; we would call them shamans today. Shamans were not just mediators but could assume good or evil divine characters themselves and wield the powers associated with that for the good or the bad of the people. Fortunately, to some degree we can still learn about the power of shamans from cultures who only in recent times entered modernity and where the memories of shamans are still part of a people's narrative and oral tradition, passed on from generation to generation. Shamans were respected and feared but were far from being the powerful priests of later times because in those early days, the realm of the here and now and that of the spirit world was in all likelihood much more connected for people in their everyday lives. No priest class had yet separated human belief systems from people in such a way that made them believe they needed a priestly class to satisfy their spiritual needs. It is only with organized religions, that eventually priest classes managed to insert themselves between the people in the here and now and the next world these people believed in. It was only natural that those seeking secular power and those seeking control over people's beliefs to further their own agendas became and remained close allies throughout known history.

What motivated some people to raise themselves above others? The answers may be quite simple. With agriculture comes surplus and regardless of how meager that may have been in the beginning, it did not stay this way. The start of agricultural society is the start of redistribution. The first peasants devoted to plowing the fields - metaphorically speaking, as the plow would not be invented for another few thousand years – would need protection. Stored harvest and seed corn would attract the attention of outsiders, maybe invite pillaging by other tribes. This need for increased protection arising with the arrival of agriculture likely marks the beginning of a much more pronounced division of labor in human societies. While there certainly was some specialization according to skills in hunter-gatherer groups, each member of such a group needed to master the comprehensive skill set required to survive. The loss of a skill in the group because of a hunt gone wrong that all of a sudden left the group with nobody to make critical tools anymore would have been unacceptable. Division of labor has its cost and the protection from robbery by invaders opened peasants up to exploitation by their protectors. The latter being the nascent ruling classes of aristocrats and priests, in a sense, the first elites. Maybe that is a too simplistic picture. For once, peasants would have retained the ability to defend themselves and often would have been the actual soldiers, frequently called on to defend the community. It would take a long time before full time soldiers became a majority of fighting forces. Maybe it was more the ability to organize the resources of a community in accordance with what most saw as reasonable or just, along with meeting religious needs that provided the avenues for ruling classes to establish themselves.

Alliances first built back then between secular rulers and religious authorities remained a constant throughout history and still flourish today. Gaining and retaining power is of course very lucrative as one can have others work for oneself. However, any ruling class, secular or religious, represents a hierarchy in itself and many that were part of one most likely believed that they were truly serving their people and their gods. However, as the saying goes, the path to hell is paved with good intentions. Unfortunately, the human capability to convince oneself that harming others is justified in the name of a greater good seems to be limitless. More thoughts on this later, right now we need to get back to the subject at hand, the beginnings of human civilization. What seems to be clear is that once humans started to live together in the much larger groups that agriculture could support, social differentiation must have become much more pronounced. We know that in the early days of agriculture, because of a much less varied diet, early farming populations actually were less healthy than their hunter-gatherer ancestors were. A single bad harvest could have devastating consequences. The average body sizes that we find associated with early agricultural settlements are smaller in build than those we know from hunter-gatherer groups before. This evidence leads archeologists to a simple conclusion. While agriculture overall was more predictable in supplying food it must have provided a less well-balanced diet and was essentially a high-risk investment as one failed harvest or pillaged grain storage could mean starvation for many. Hunter-gatherer groups also led a high-risk lifestyle. Failing to secure sufficient rations of meat for themselves and their families, they would not last for long. Only the successful hunters would survive and live to father children. Because of this natural selection and the varied diet of hunter-gatherers, their fossils typically indicate healthy adult individuals with little indication of disease. Of course, there are exceptions, some of which clearly show that individuals in such groups cared for each other, a trait shared across human species as we know for example from the Neanderthals.

In early agricultural societies, the average person may have lived a more miserable life with a much less varied diet, and many may have died from diseases unknown before. However, in a society that could provision for the future, even weaker individuals would still be able to have offspring. The result seems to have been stunted growth as evidenced by the diminishing physical stature observed in the archeological records, that is, were we have them from such early times. A different argument sometimes given for this less healthy physique of early farming populations is the supposed impact of domestication; not the domestication of animals but of humans themselves. As the argument goes, agriculture and civilization are the human equivalent of being domesticated, resulting in a physically weaker populace and in a stronger ruling elite. While there certainly is something to the domesticating effects that civilization has had on humans it looks a bit far-fetched to account for the extent of these changes we see so early on. It would seem that any such effect ought to have run through many generations to take hold. As the human reproductive cycle is not as quick as that of our domestic animals, any domestication effect should certainly take much longer to manifest itself. More relevant is here the observation that modern humans are again significantly taller than our early farming ancestors were. But then, compared to our Cro-Magnon ancestors, our average brain size is smaller. We also show a number of other features usually associated with domestication such as a reduction in teeth size or more rounded heads. To which extent does domestication also mold us mentally and spiritually? We do not really know. However, seemingly, civilization does make us more civil, albeit ever so slowly.

Using religious or mythological beliefs can help some people gain power over others. It can also help a ruling class to bind a greater number of more or less homogeneous groups tighter together thereby building larger groups. Religious belief becomes the mortar that binds them together and ties them to their ruling class. It is what all belief systems have in common, not just religious or mythological ones. What are the more or less homogeneous groups in this context? Tribal groups living in the same areas for a long time may have evolved shared religious beliefs and thereby slowly merged into a larger tribe, pledging allegiance to shared secular and religious elites. Tribe, as this word is most often used, has a strong ethnic connotation today, but it is much more fundamental to human behavior than that. It certainly is true that frequently tribes provided a ready instrument for those believing in the superiority of their own close kin. That could have been just to further their own and their tribes' economic interests at the expense of others; in cases it also could be the pursuit of power at the expense of everyone else; and in the worst cases it would not stop short of ethnically motivated mass murder. We can see this throughout human history from the killings of a few to the mass murder of millions, the genocides of the modern age.

However, tribal associations do not require an ethnic foundation and frequently they do not have one. While some ethnically more homogeneous sub-groups have been the driving force behind the birth of some civilization, more often than not civilizations largely have been ethnically egalitarian. One can point to such past examples as the Roman Empire for an ethnically diverse civilization governed by secular law or for that matter to the Umayyad or Abbasid Caliphates governed by religious law. In modern times, the United States of America provides probably the best example, so far from perfect, for an ethnically diverse and egalitarian democracy. However, it is only one but large representative of what we could refer to as members of the Western or Atlantic civilization that all have laws on the books ensuring equal rights for all ethnic groups, at least on paper. While before the law everybody is supposed to be the same, we can clearly not assert that Western civilization is ethnically egalitarian yet, far from it. This is not only a problem that Western civilizations have to face. As we look for example towards China, we would expect this problem to be less of an issue because a large part of its population is ethnically homogeneous, the Han people. However, that seems not to be the case as ethnic minorities are still discriminated, even though, just as in the West, Chinese citizens are supposed to be equal before the law.

Ethnic discrimination is not the prerogative of a specific culture, society, or ethnicity. It is rather more a question of who does the discriminating, which usually is determined by which ethnicity holds power. Ethnicity and race are two concepts often confused. Race refers to human differences in physical characteristics whereas ethnicity describes various expressions of human cultural characteristics. Sometimes racial and ethnic differences go hand in hand and sometimes they do not. It is often hard to tell if people are discriminated because of their race or their ethnicity, maybe it is one or the other, maybe it is both or maybe it is just because they are different. The need for common defense or offense to secure or acquire critical resources such as arable land, natural resources or trade, was and continues to be a strong motivator uniting different groups and interests in a common purpose. Something we can observe throughout history as well as in modern times. It looks like an external threat has to become just existential enough and all of a sudden, we all become brothers and sisters, regardless of how different we are. We can ascribe similar uniting forces to writing, language, law, religion, art, and learning. Modern genetics shows us that the concept of racially differentiated humans is very frail with little biological substance and ethnicity of course is a cultural concept only. Racial and ethnic differences between humans are superficial at best when we hold them against what we all share as humans, but nevertheless, some of them are quite visible. Maybe someday this will not matter anymore, but unfortunately, most of our societies today are still deeply xenophobic, rejecting anything that does not conform to what they mistakenly believe to be their racially and ethnically pure stock.

The frequently used definition of civilization given a few paragraphs above is set in contrast to looser and usually smaller tribal or nomadic groupings whose rules have a narrower focus on group survival and often remain unchanged for many generations. For most of the latter, it is more appropriate to speak of tribal cultures, as anthropologists do, instead of referring to tribal civilizations. Even today, there are still a few small tribes left, living by rules that may not be too different from those their ancestors lived by millennia ago. Today, the words tribe or tribal are often used with a negative connotation, invoking backwardness, primitiveness, and lack of social development or learning over long periods of time, just to name a few examples. Sometimes it seems as if there is a need on civilization's side to color tribe and tribal much darker than they deserve so civilization - to which most of us would proudly subscribe to - shines even much brighter in comparison.

However, things are a bit more complicated than that. As civilizations evolved humans did not just shed their tribal nature like an old skin to morph into civilized woman or man. Looking back at human history, as we know it today, our earliest bi-pedal ancestors, if we give *Homo erectus* the honors, roamed the savannas or hid from predators in Africa about two million years ago. For most of our history, our social fabric developed to ensure the survival of very small tight knit groups of humans, essentially groups of only a few families in size. This very much is still our inheritance today, if we like it or not. Our ability to have meaningful social connections with others still reflects that early history, anyone trying to maintain social relations with many more than one hundred people is likely to confirm this quickly. Social media may suggest to us that we can have thousands of friends and connections, but most will find their friendships to be poorer and the ones they have online will come at the expense of the real relations they could have close at home in their communities. There is no denying that the fabric woven into our ability to make social connections rests on almost two million years of human evolution if we take the appearance of *Homo erectus* as the start date. While one can argue about this dating, one thing should be clear, who we are as social animals is deeply ingrained into the evolutionary oldest parts of our brains.

In evolutionary terms, two million years, even so we have culturally evolved quite a bit during that time span, is not a very long time period. The truly fascinating aspect of human evolution to hold in front of us in disbelief is the incredibly rapid development that differentiates human lives today from that of early modern humans some two to three hundred thousand years ago. This amazing acceleration of human cultural and technological development into ever larger and more complex societies has given humans an ever-increasing command over their own destiny and that of their fellow human beings. It has resulted in the complete domination and subjugation of the animal kingdom and of habitable Earth by our species. One has to be careful when using words such as evolution and development as there is a dependency. Development by its very nature is reversible but evolution is not.[9] Evolution has endowed our species with a capability for development unlike any other but in evolutionary terms we are still in large part our primate ancestor's cousins. Human civilization itself is only skin-deep and a minor scratch will easily pierce it exposing our evolutionary heritage, which has deep tribal roots.[10] Every human that was ever born had to re-learn civilization and that is unlikely to change for a long time to come.

There is some justification to interpret many of the great achievements of human development in the context of containing the demons of our human evolutionary heritage. Arguably, the cultural constructs that humans have developed over the past several millennia were indispensable for us to evolve as the most socially and technologically advanced species, dominating seemingly everything on planet Earth. It is useful to look at those constructs as corsets that keep us straight, as training wheels that keep us from tumbling over, as crutches that allow us to keep going until in some very distant future our cultural behaviors leave an evolutionary imprint and provide us with a much thicker skin. Alternatively, science may be able to provide an earlier relieve by engineering our tribal traits out of our evolutionary inheritance. However, for the foreseeable future, we have no other choice but to make the best of our evolutionary heritage. At the same time, we must make sure that as a species we survive long enough to either biologically evolve further or to develop the capability to engineer our own evolution. For a long time, human history has been written and told from the vantage point of those who managed to come out on top. More recently, historians also began to explore it from the perspective of those who had no say in making it, those that were on the receiving and not on the shaping end of it. Admiring all the great achievements civilizations have left us we often still know little about the mountains of vanquished they have been built on. We cannot know how civilizations are typically built by intelligent lifeforms that may exist somewhere out in the distant universe. Human civilizations are the only kind we know. Are mass murder, genocide and war just part of civilization building, are they just part of the evil that we have to endure in order to get all the good that civilization undoubtedly has brought to us? At least to the many of us who are part of the human species, because other life may not have fared so well under human civilizations. We have to ask, where is all this evil coming from and will we ever be able to put an end to it? How much of this is rooted in the human self and how much of it is learned? *Gnotis auton* - know thyself - the ancient Greeks were well aware that there are things that cannot be explained and understood from the without but need a deep understanding of the within. Only very slowly do we get a better understanding of what we call the human self, and how much or how little we are in control of it. The key to this improved understanding lies in our brains. Today, we know so much more about the human brain and how it works but even though, much of it remains elusive. As it stands, we can only acknowledge that we do not really know how much of our tribal behavior is contingent on group interactions and essentially learned behavior and how much is evolutionary hard-wired in our brains. For now, we cannot explain, we can only observe and learn.

We humans are still tribal, and we are suffering from this on an almost daily basis. The number of people we can meaningfully connect with is limited. Tribal associations allow us to connect within larger groups and most of us do need this for comfort, security, or both. While tribal nature often still reveals itself in its original context centered on an enlarged family or sets of families it readily can mold itself to serve the purpose of any cohesive societal group. Thinking in terms of them and us, this comes to all of us quite naturally. We tend to look at societal developments through the lens of interest groups but often forget that those interest groups are underpinned by our tribal heritage; and we underestimate how moldable our tribal urges are to serve different purposes. There are good and bad sides to our tribal nature, and they are more closely related than we may like them to be. Before we go there, it will be good to take a closer look at tribal manifestations in today's societies. There, let us start with the familiar, the seemingly easier ones such as fan clubs, specifically in sports.

As with many things, one has to go back to the Romans to find the first great example of what in our language we would refer to as fan clubs.[11] The late first-century Roman poet Decimus Iunius Iuvenalis (c. 55 – 130 or later), better known as Juvenal, is credited with penning the quote: *"Bread and circuses, that is all the common people want"*. Even though most of us will likely not agree with Roman circus games being spectator sports in today's sense, they truly were for the Romans back then. Given the increasing popularity of bare-knuckle fighting contests with real spilled blood for all to see on high-resolution television screens, we may actually circle back much closer to Roman tastes for blood sports. During the Roman republic, chariot races featured four teams differentiated by color, the Reds, the Whites, the Greens, and the Blues. Each of them had their fan base, to use the modern term. By the time of the Byzantine Empire, only the Greens and Blues remained, and their fanatic fan bases were responsible for many bloody riots; we are talking thousands of people killed. Some of today's soccer hooligans would have felt right at home in Byzantium. For the ruling Byzantine elites those rivalries served a purpose, at least for some time, as the Greens and Blues were massacring each other and not directing their violence against the state and the ruling elite.

Channeling testosterone - we are typically talking males when it comes to sport fan clubs with aggressive behavior - by funneling it into a tribal activity where it can be discharged without endangering the state or other fellow human beings presents us with a very useful tribal function. The more so since, unlike Byzantium, we do not tolerate murder as part of enjoying a game anymore.[12] The danger lies in controlling emotions, or better the emotions of masses. As the Byzantines had to learn the hard way, it takes only a small step to transform fan bases like the Greens and Blues into political bases. Interestingly, we still do see this aspect today with much of the "hooligan base", many of whom are frequently right wing nationalists, racists, or both, moving with ease between cheering their sports team and beating up perceived political opponents or other fellow humans, who in their perspective just happen to belong to the wrong tribe. Few would disagree that many social activities are very much of a tribal nature, even if not as glaringly as those of the Greens and Blues in old Byzantium. While for some of these tribal actors it will always be "them against us" and "we belong, they do not" many of them are actually good tribes. Good tribes renounce intolerance and violence and its members very much rejoice in the social comfort and the sense of belonging that a close tribal association can provide. Good tribes build on a positive belonging that does not need to negate or discriminate other fellow humans to define itself. In a sense, some religions have evolved into that direction too, but we are getting ahead of ourselves.

There is accepted tribal behavior in our societies and the tribes who follow such conduct are looked upon kindly by the state or civilization at large; and there is unacceptable tribal behavior. Not surprisingly, different societies define acceptable behavior differently. As a rule, it is safe to say that democracies or in general open societies take great concern to ensure the freedoms and liberties as well as the safety of their citizens. Therefore, their rules regarding tribal activities are usually generous while seeking to safeguard everybody's safety. Liberties such as freedom of assembly and speech or the freedom to form partisan associations and organizations are rightly considered great achievements and they are to be enjoyed by all societal participants equally, whether one likes their speech or not. Other government regimes often find themselves in some sort of tribal conflict with large parts of their population, think dictatorships, oligarchies, or theocracies as examples. They seek preservation of their rule, which translates into preservation of their tribal association above all. Therefore, the ruling elites in those governments behave themselves as tribal actors in a negative sense seeking to align all tribal activities with their own tribal colors. They may still allow tribal activities but if the governing tribe for example wears gray colors other tribal activities better come in shades of gray or else. Interestingly, bad tribal actors seeking to secure increased rights for their tribal members at the cost of diminished rights for other parts of society usually have a much better platform to advocate their beliefs in democracies or open societies. In oppressive societies, one of the bad tribal actors has already succeeded at the cost of everybody else and that includes preventing all other potentially bad tribal actors from gaining a platform.

Tribalism of the bad kind is the cancer that threatens to destroy our democracies from within. Democracies certainly have a good number of deadly enemies, foreign and domestic, who would like nothing better than for them to pass into history. Increasingly, more and more entrenched tribalism of the political kind may ensure that the enemies of democracy only have to sit back and wait for us to destroy democracy ourselves. Politicians once saw themselves as the elected officials of their communities to represent the interest of these communities within what was acknowledged as the greater interest of society at large. Maybe this was always closer to an ideal than it was reality. However, there seems to be a clear difference between the sense of integrity and duty that we can see with many politicians in the not too distant past and with a select few even today and what has practically become the norm. Becoming a politician in our day and age seems to be more of a career path to power and wealth and less a calling to serve society. As it is, an increasingly large number of politicians in today's day and age represents less the interests of their communities and more the vested interests that helped them win elections by funding their campaigns. This can and will not go on forever because it destroys the very fabric holding a democratic society together.

History as the saying goes, does not repeat but it does rhyme. The Roman Empire did eventually fall, but not because it succumbed to the onslaught of barbarians. Rome was consigned to pass into oblivion much earlier when internal strive driven by special interest groups ripped apart what had been the very strength that had kept external and internal enemies of Rome at bay – a political class that lived up to its responsibility. What precipitated Rome's downfall was not the might of barbarians but a pervasive and prolonged failure of its elites. Our modern democracies may not face the threat of barbarian invasions but the increasingly rampant failure of elites in many of our democracies should deeply concern us.

To Species Civilization

Human civilizations go back a few thousand years, so are we a civilized species yet? What is a civilized species and when do we know that we have achieved the level of species civilization? We can assume that these are questions arising quite naturally to any unbiased and halfway astute observer of our earthly affairs. Think of the proverbial alien. Not that we are ever likely to see one of them here on Earth, unless they find a way around the laws of physics. However, everything is possible in a thought experiment. What impression would we make on an alien observer who staid long enough on Earth to understand our species? Let us assume that this astute alien observer stayed long enough to have learned to put her or his – we cannot really say which it is, it could also be both - observations in contexts and languages essentially conforming to ours so if we happen to get a copy of the aliens report we could make sense of it. You get the picture. Just ask yourself before reading on: would you expect such a report to look favorably on the human species? Specifically, considering what it has done to life on Earth and to itself? A reasonable guess it that after due consideration, you likely will come to the conclusion that the report will not be favorable at all; but that hopefully it would point out the potential for betterment inherent to the human species.
As a species, we have mercilessly fought our way to the top of the food chain and have become the supreme rulers of Earth. Animals we domesticated for our services thrive. They never would have multiplied to such large numbers on their own if left in the wild. However, we have altogether enslaved those poor creatures, forcing most of them to live out their short lives in horrible conditions. Our alien visitor may well wonder how, if humans really believed in a God, they ever could expect to be forgiven for this. We are also quite capable to treat our fellow human beings equally miserable. So maybe it is not prejudice against non-human creatures that makes us treat other life with such contempt. Is it maybe just in our nature? We can only wonder what may go through our alien's mind when he reads about slavery, which in its most overt forms today is mostly gone. We may like to think that but then that is not true either. We may not call it that anymore, but slavery still exists in our modern times. Like in forms of bondage not to different from medieval servitude, mostly for women; sometimes it still exists as plain outright slavery, not disguised and presumably justified by ancient rituals and customs; or as one of the many lesser but nonetheless evil forms of bondage our societies seemingly still tolerate. Yes of course, we all want the alien to read about our great cultural and technological achievements. But does that really matter once our alien, while ruminating through our vast libraries filled with the records of all of our achievements, finally stumbles over the accounts of our many and atrocious wars, World Wars, and genocides? It does not. There have been so many human transgressions against any natural or divine law in the histories of our civilizations that our alien will have a hard time seeing us as civilized in the first place. Maybe he will concede us the status of fierce tribal cultures, always ready to kill, enslave, destroy and plunder when the occasion affords; always ready to go for the jugular of those who are not like us. However, it will not just be the mass murders, genocides, and wars standing out in our alien's report to her home world. It seems that even in peacetime human societies are not capable of taking care of their own in a compassionate way. Some of it our alien visitor may attribute to the inherent

competitive nature of humans. However, to comprehend it fully will be quite a struggle for our alien as in our better moments even we cannot find justification for it. Surely, any advanced alien civilization has long left greed behind and so it will be hard for our alien to understand our societal and personal priorities. Our alien friend may wonder how so many have lived and continue to live in abject poverty given that the planet has enough resources to provide for the basic needs of all humans. Our alien may wonder why so many humans seemingly believe they own a part of planet Earth. Humans surely must know that they are merely temporary custodians, and they own the planet no more than any dinosaur ever did. Why do they not behave accordingly? In some parts of our human world people have so many resources at their disposal while in other parts they barely can survive. How can that be? How can it be that even in what are seemingly the most prosperous countries on Earth large parts of their populations live in conditions that clearly fall under what humans themselves would call poverty? How can humans live with that? Our aliens mind must spin as all these observations are clearly nothing like the signature achievements of an advanced civilization. However, that may not even be the worst.

What in our alien's perception could likely be the straw that breaks the camel's back is the seeming inability of the human species to realize that its very survival as a species depends on the human species becoming the steward of life on Earth. Instead, humans seem to be hell bent on destroying the foundations for their own survival on Earth and for that matter, for much other life on Earth. If our alien had looked a bit into the evolution of life on Earth, she, he, or it may well conclude that the human species will most likely be responsible for the next mass extinction of life on Earth. More pointedly, the human species is already well on its way to level and impoverish life on Earth in a manner that in the past only natural catastrophes could achieve. Unfortunately, it is quite unlikely that an alien observer would come to see in the human species much more than a very sophisticated but parasitic species that lives of everything else on planet Earth and is likely to destroy itself and many other species in the process. Our alien's decision may well be to come back in a few hundred million years to see if another intelligent species will have evolved on Earth, one that may be capable of avoiding the mistakes of the human species. However, this could be a mistake in itself. One human characteristic is that we can learn from our mistakes. It will likely be difficult for our alien to see her or his way through to that conclusion, but maybe there will be a footnote in the report that the human species, after all, may still be able to redeem itself. Only then may there be a path for humans towards species civilization; for now, humans may be civil from time to time but truly civilized, our alien may conclude, they are not. So, what will it take for us humans to turn the tide and begin our journey towards becoming a civilized species? What is holding us back?

Before we go into this, we need to reflect a moment on a change in tone in this section. In a cursory discussion of our societies and what may ail them this is to be expected. These topics are more subjective and the paragraphs to come will reflect this. In a sense, this is also just a simple consequence of the predicament in which we find ourselves. On the one hand, we have all this progress so visible in our most modern societies. On the other hand, we still see forces at work not just jeopardizing our future in the here and now but the future of humankind itself. What is so very incomprehensible and reprehensible at the same time, is that behind much of what endangers our societies and potentially humanities future is nothing else but sheer bottomless greed in all its various ugly forms. There is unfortunately no better or nicer way to say this. It is against this background that what follows next may sound unduly harsh. However, it is with what we treasure most that we must be most forthright. In this case that is the best form of government that human civilizations have produced so fare, our democracies however imperfect they still are. So, what ails them?

Why is it that in today's day and age countless millions continue to be condemned to abject poverty? Why is it that too many rich nations do too little to help those who struggle to help themselves? Why do so many poor nations exploit their precious resources only to enrich their leaders at the expense of their suffering populations? Why are xenophobia and religious intolerance still rampant in many of our societies? We could go on with many more such questions, the list of grievances is much longer than the space we can allow for it here. It is those in power, we broadly refer to them here as the elites, we must look at for answers. They control the economies and politics of countries and nations. Failure of elites is an ancient disease that continuous to plague many countries across the globe, from rich to poor alike. History has endowed many of the richest countries, most of them democracies, with the advantage of strong political and economic institutions helping them to safeguard their political and economic systems; other countries have not been so lucky as their path through history has not left them with such safeguards to keep their elites in check. However, even nations with long democratic traditions can succumb to the influence of powerful elite interests, working to influence legislation unduly in many ways to their advantage.

What we can observe in places such as the United States of America as well as the European Union are likely weak cases of an affliction brought about by the so-called *Iron Law of Oligarchy*. This law, in the original German version called *Ehernes Gesetz der Oligarchie*, was first formulated by the German-Italian sociologist Robert Michels (1876 – 1936) in his 1911 publication titled *Zur Soziologie des Parteiwesens in der Modernen Demokratie – Untersuchungen ueber die oligarchischen Tendenzen des Gruppenlebens* (*Political Parties – A Sociological Study of the Oligarchical Tendencies of Modern Democracy*). Michels study focused on the evolution of the powerful German labor movement in the late nineteenth century. His core thesis was that inevitably, leaderships of organizations increasingly focus on their own personal interests, which they achieve through successfully influencing and controlling organizations to serve their own purposes. Politics always has to balance the interests of the various bodies of society. If one of them or a group of them manages to co-opt political leadership then they essentially have set themselves on the path outlined by Michels. In many ways, corporations have succeeded in accomplishing just that, in some countries more so than in others. Developing nations are often afflicted by much more severe cases of this disease with the *Iron Law of Oligarchy* ensuring that one despotic regime hell bent on extracting as many riches as possible for themselves and their corrupt elite kin is almost inevitably followed by another. Decades after the end of colonial rule, these elites condemn their nations to continued unspeakable suffering. Getting rid of such corrupt elites in one fell swoop in a French style revolution may have served these countries better than having again and again supposed saviors emerge from their elite ranks only to become the next crooks holding the reigns and extracting whatever wealth a country still may hold. Alas, some tried but to no avail as new elites quickly emerged from the carnage to emulate the ones, they just got rid of. Violence breeds violence and corruption breeds corruption. It is a seemingly endless viscous cycle, denying countless millions their basic human rights and denying humanity the benefit of their ideas, ingenuity, and compassion, as they are never allowed to flourish.

How do some of the milder symptoms of the *Iron Law of Oligarchy* afflict the United States and European nations? Well, one can for example make a good case that the political system in the United States today is less a democracy and more akin to an oligarchic democracy. The influence of the oligarchs, mostly corporate interests, outweighing the interests of the voting age population; not to mention those who do not have a voice yet but who will nonetheless suffer the consequences of our actions or inaction, our children. Why does the richest country on Earth have a health care system geared towards serving corporate and monetary needs, while leaving large parts of its population without access to the kind of health care that their neighbors to the north can enjoy in Canada? For an answer, we need not look further than to the *Iron Law of Oligarchy*. The situation is somewhat different in Europe where corporate interests have significantly less political power today, but things are changing there too. Over the roughly past two decades, under the mantle of modernization, British governments have for example much diminished what once was the pride of the nation, a well-functioning universal health care system. Did they do this to provide improved health care for the citizens they were supposed to represent? No, they did it to serve corporate interests. In a complete travesty, the demolition of a once great health care system makes a few considerably richer and many a lot poorer and with much reduced health care services. How can any politician helping to cause this train wreck claim to serve anything else but corporate interests?

One of the great quotes of the outgoing twentieth century came from the 40$^{\text{th}}$ president of the United States, Ronald Reagan (1911 – 2004). In his first inaugural address in January of 1981, he said *"In this present crisis, government is not the solution to our problem; government is the problem"*. Since then, Reagan's pronunciation has been the mantra of many, mostly conservative governments. It is a good guess that more likely than not almost anybody will be fed up with inefficient, unnecessary, and overreaching government. However, seeking the solution in privatizing what are and should remain essential government functions is dangerous. We already mentioned health care. Ensuring that all its citizens have access to and can afford using a functional health care system is surely no less important to a nation than defending the very same citizens against external threats. The public sector will only take over government responsibilities to the extent profits can be made. Businesses must make profits, and no one should blame them for structuring health care systems such that they profit as best they can. They are not at fault; the fault lies with the politicians that allow them to take over the profitable parts of the health care system or tweak them to the detriments of patients such that their profit margins meet Wall Street expectations. Common sense would imply that health care, no different from national defense, has to be fully tax funded to the extent that is covers the health care needs of its citizens. That does not mean that there is no role for private companies. A universal health care system can assure that all citizens have access to affordable health care; nobody should have to die or suffer financial ruin because of an illness. Those who want more, like single rooms in a hospital or unnecessary cosmetic surgeries can get additional private insurance that will cover those needs for a prize. Instead of complaining about inefficient public services, they should be fixed. A nation should be proud of its competent and efficient civil service and those employed by it should be equally proud of working for it. That will require some major changes in attitudes.

Maybe a look back into the past will help. Where did many of the major technological breakthroughs that so dramatically changed our lives over the past generations come from? Well, they came from tax dollar funded research work done by national research laboratories or universities and even sometimes companies with research contracts paid by tax dollars. For many decades now, taxpayers have been funding costly fundamental research and private companies have been reaping most of the monetary benefits. That happens when those companies gladly sell us all kinds of marvelous gadgets with little of the profits finding its way back to the taxpayers who funded the basic research enabling their technologies to begin with. Much of the twentieth-century modern technology development we are justly so proud of represents case examples of socializing losses and privatizing profits.

Frequently, politicians like to malign public organizations for their inefficiencies and incompetence. It is of course a vital government role to fund basic research, but the inefficiency and incompetence is here with the politicians who never seem to have found a way for the taxpayer to monetize the pay-off from such research and development; that was left to corporations who pocketed all the profits. Just imagine, if governments had been smart enough to ensure that a small percentage of the profits would flow back into the public coffers to fund more research and development or to pay for some other public services and if none can be found, maybe to lower taxes for its citizens. Why have our politicians not been able to achieve this? Because once they retire from political office, they all too often parachute into the soft executive chairs of private companies. Politicians, not all of them but excessively many, are part of this rigged scheme that uses the public coffers to enrich corporations for which they eventually will work; it is a revolving door that should have been closed long ago.

Will we ever get out of this mess? That depends. The situation in the United States is somewhat different from Europe. Hopefully, at some point in time, voters in the United States will be sufficiently fed up and finally elect a government that truly has their interests at heart. None of the current political parties seemingly has, the bulk of their politicians are just like managers in private companies, interested foremost in their own career and how to personally profit but less interested in the well-being of the civic body. We have been sliding ever deeper in this morass for quite some time and the ride is not over yet. How can this go on for so long? Few would associate the label welfare state with the United States. However, technically it essentially is one, albeit of a very different kind. There are substantial transfer payments in the United States, a lot of people depend on them, many more than most assume; some numbers suggest that close to fifty percent of the population receives some kind of benefit from one or more government programs. While that is a lot of money being spent it is seemingly being spent to keep the lid on what otherwise could quickly become an explosive situation, but it does not really change the outlook for this large part of the population.

This is a problem with many transfer systems. They tend to address the symptoms but not the root causes. Surely, these people need support. However, more importantly, they need the kind of support that helps them to get back on their feet and eventually support themselves. What madness is it that makes it seem right for a society to provide handouts just enough to ensure the malaise continues but not to do what is needed that all of its members can become productive citizens; if not in the current generation than surely in the next. Education is an essential government function as it must be highly desirable for any nation that all of its citizens can realize their potential to contribute to its prosperity. So one would think. However, it is astounding how unequal, even in a nation as rich as the United States, access to quality education is. Ensuring that those of the young generation growing up in poverty today will have acquired the skills that enable them to provide for themselves and their families once they are grown-ups, is fundamental to reducing rampant poverty.

Most European countries have a somewhat different form of capitalism, frequently referred to as social market economics. That does not mean that there are no poor people in Europe, there are, quite a few. Being poor in Europe is however different from being poor in the United States. Not that it is pleasant to be poor in Europe either, but far fewer poor fall through the cracks of the social systems there. Unfortunately, even there, poor homeless people still occasionally die in the streets of wealthy cities. Citizens of European countries have access to public health care that they frequently augment with private health care; most importantly, they do not face the threat of financial ruin because they unfortunately happen to have expensive health problems; similarly, public education is in good shape and compares favorably with much of the United States. Higher education in Europe is affordable with many excellent universities that students can attend without incurring massive debts that will burden them for many years to come because higher education is funded through taxes.

You may think that Europeans pay very high taxes to afford all of that and you are likely right, but it differs from country to country. However, you have to tally up what you are paying in total and what you are getting for it and there, things get interesting. Just take some of the richest states in the United States, California and New York for example, which are supposedly closer to the European model than many of the others. That may be so on a relative but not on an absolute scale. If one looks at what families in those states pay in terms of local taxes, mostly school tax and property tax, and adds to that state and federal taxes, then the relative numbers actually do not look so different from many European countries. However, once health care and education costs are added to family budgets and considerations are made for the services that are actually available to tax paying citizen there is quite a gap. The short of it is, residents of the richest states in the United States which are supposedly closer to a European model pay on a relative basis at least about the same but receive significantly less for it than most citizens in many European countries.

Why is that so? Well, it certainly reflects a different preference for how to spend taxpayer dollars or euros; and it has something to do with the inefficiencies that come with the legacy of the respective administrative systems in the United States and Europe. School systems in the United States are notoriously expensive but seemingly still underfunded. On top of that, they continue to enshrine a wealth and racial separation in education which should have been overcome long ago. For a country as large and diverse as the United States a nationwide unified school system is likely not a good solution either. Instead, states should take control of education within their boundaries and replace school funding through local taxes with a general income-based state tax tied to education, including also higher education. In all likelihood a unified statewide school system will not only ensure that wealth and race no longer can be factors in determining which school a student can attend. It may also be less costly.

While much of the discussion above has focused on corporate oligarchs, there are of course also other players with their own agendas, looking for ways to tweak the system such that they can control it to serve their interests. Worker unions in the United States have been crushed long ago in the interest of the corporate oligarchs and are mostly powerless today; at least when compared with their European counterparts. In Europe, worker unions still play a prominent role. They have been at least as successful as corporations in shaping politics according to their members and their own interests. Significantly, in many European countries partnerships between corporations and unions are now pretty much the rule and that is largely due to the fact that there is much more agreement between the various powerful interest groups and throughout the population on what they see as the defining elements of their societies. However, that also differs of course from country to country, and this consensus is for example different in France than it is in Germany.

Looking at democracies in Asia such as Japan and South Korea the picture gets a little more complicated. The sense of social cohesion in these societies is much more pronounced than in the United States or in Western Europe where it seems to be somewhat higher than in the United States but still markedly lower than in Japan or South Korea. At the same time, both of these countries seem significantly more stratified in their hierarchies and honor and respect seem to be much more pervasive virtues than in Western democracies. Stealing from the common good, because that is in simple terms what the *Iron Law of Oligarchy* is all about, is not just a punishable offense in these countries, like it is also in the West. If convicted of such an offense, the dishonor associated with it, though it seems, is culturally almost unbearable in these societies. Failing is human and human failure along the *Iron Law of Oligarchy* surly also happens in Japan and South Korea, but the cultural context provides a much higher hurdle and if committed it represents a much more grave offense than in the West.

The *Iron Law of Oligarchy* is at work in our democracies, but in subtle ways. Everything is done within the legal framework and policies are reshaped such that the respective oligarchs – in most instances corporate interests – can nudge the systems of government to work in their favor. Success in such efforts, which are often nothing else but blatant rent seeking, can have enormous payoffs; and so do monopolies or quasi-monopolies. As an example for rent-seeking behavior in the United States, one just has to look at the pharmaceutical industry being successful in getting politician to rescind the right for Medicare to negotiate drug prices with pharmaceutical companies. Medicare is a national health insurance program in the United States to which every citizen has access once reaching retirement age. As such it is the largest health insurer in the country and would be in a position to use this leverage for negotiating affordable drug prices with pharmaceutical companies charging exorbitant amounts; but Medicare is forbidden to do so.

Clearly, this is rent seeking by the pharmaceutical industry and rent granting by politicians. At the same time, these pharmaceutical companies sell their drugs for much lower prices in Western Europe and other countries because there they have to negotiate to be competitive. How frustrating for United States taxpayers whose tax dollars most likely paid for some of the fundamental research that may have pointed pharmaceutical companies to promising drug developments in the first place. The good thing is that in Western democracies we should be able to count on the pendulum swinging back as it has done before. Times of excesses in the United States such as the 1920's, were followed by the landmark social legislation of the New Deal. For all we can see the pendulum may already be on its way back, politicians just have not realized it yet. These cycles are likely to continue but as long as we hold strongly to our constitutions and democratic freedoms the resulting trajectory will be inclined towards were we need to go - towards increasingly more inclusive political and economic systems that strengthen social cohesion throughout nations and also between them.

Admittedly somewhat selective as to what ails our modern societies, the above narrative can only scratch the surface. However, it should make clear that much of what troubles us in our modern societies today are only the symptoms of a more profound failing, deeply seated in our human nature. Much of the discussion here is restricted to what we can broadly refer to as democratic countries even though they all differ somewhat from each other. The reason for that is simple. The government systems and societies in these countries, while far from perfect, are the best models of societies we have with functioning inclusive political and economic systems. Pointing out the many things that are going wrong in places that do not have the one or the other, or none of them, Russia and China certainly come to mind, is a futile exercise. In Russia, the *Iron Law of Oligarchy* has run its course, as it essentially has become an oligarchy. Russia should be an economic powerhouse, but its extractive political and economic systems have turned the country into an economic dwarf. China on the other hand has an economic system that in many aspects is becoming inclusive; however, it does not look like China will achieve the transition to a politically inclusive system anytime soon. One could be excused for having second thoughts about China in this regard, specifically when looking back at what happened in Russia after the demise of the Soviet Union. It may well be that the Chinese communist party retaining a firm grip on power as the country started its transition from a planning to a market economy may have helped China avoid the political and economic chaos that eventually resulted in Russia becoming an oligarchy. One can only hope that China at some point in time manages to transition from what in essence is still a dictatorship towards a civil society that respects the rights of its entire people. We would be amiss if we did not mention the largest democracy on our planet, India. India with its diverse ethnicity, its many religions as well as with countless different languages is the largest democracy experiment humans have ever tried. Indian democracy faces tremendous challenges but so far, it has managed to navigated them without floundering. If India can continue on its path towards an ever more inclusive democracy that it started only a few decades ago it will show that democracy does not only work for rich and mostly monolithic developed countries.

While the *Iron Law of Oligarchy* is in many ways seemingly pervasive, it certainly is not the sole explanation for what plagues our societies today. There are many other reasons that in different ways strengthen the ability of elites to rule countries with their own interests foremost in mind. It is largely the inability of the broader populace to ensure their interests, rather than the power of the elites, that in many countries allows elites to get away with subverting democracy. If people do not care or fall prey to disinformation, they are unlikely to fight for their interests. The 2016 campaign of the British right to get Britain's to vote for leaving the European Union is a good example for that. For most young people in Britain, leaving the European Union will clearly reduce their future opportunities, their interests alone should have ensured that the United Kingdom would remain in the European Union. However, they did not care about what seemed to be uninteresting politics and did not go to vote. Hence this issue was decided for them and the interests of the older and more conservative parts of the population won out; an ingenious misinformation campaign and outright lies likely also helped to convince many who otherwise would have still voted for staying in the European Union even though they were conservatively inclined. The lesson is simple. Democracy must be fully transparent and accountable, and it will only work if people fully participate and participate fully informed.

However, in many of our democracies, the reality is far different. Voter participation is low with people disillusioned as politicians and the supposedly representative systems continue to fail them; and disinformation seems to be the name of the game for most political campaigns, many of them funded by corporate interests. Does anyone think that corporate interests would be able to have the political influence they have today in the United States if everyone would care to cast their votes? Quite unlikely. Therefore, while there are many reasons to blame the failure of the elites, they have just been taking advantage of the situation that many poor people do not exercise their right to vote. In the United States the conservative right has well understood the danger of more poor people or colored people going to cast their vote as they have been trying every trick in the book to deny them their right to vote outright or by making sure that their votes would not count by re-arranging the boundaries of voting districts. Mind you, voter manipulation is certainly not a prerogative of the conservative right; the political left are quite good at it too in their own ways. Dangerously, politicians in all countries, on the right and the left, take increasingly populist stances abdicating their responsibility to pursue the greater good for all people by choosing to pander to our lower human instincts, framing our human condition as an *"us against them"* contest.

All of this is certainly not new. It is the very reason why competent and transparent institutions and the rule of law are so important. You may have to swallow an injustice for the moment but the institutions of a powerful democracy such as the United States allow everybody to legally challenge a perceived injustice or wrong. Political parties certainly have the means to do so. However, that is the crux, and it is another unfortunate example of the *Iron Law of Oligarchy*. Abraham Lincoln (1809 – 1865), the 16$^{\text{th}}$ president of the United States, in his second inaugural address, his famous Gettysburg Address, spoke of the *"government of the people, by the people, for the people"*. These words still resonate with many but the truth is a bit different, and today's reality is rather better reflected by a different choice of words such as *"government of the lawyers, by the lawyers, for the lawyers"*; the irony is that Lincoln was a lawyer, like many politicians since then. In this regard, a look at the 116$^{\text{th}}$ United States Congress elected in late 2018 is very informing.

Almost 37% of House members and some 53% of Senate members possess a law degree. However, that is already an improvement. In the nineteenth century, the representation of the legal profession in the United States Congress was even much higher than that, up to 80% at times. So, is it any wonder that the configuration of the legal system of the United States is such that at every turn, lawyers can set up a tollbooth to extract their pound of flesh? The United States has a legal system that works well for corporations or political parties or rich people. However, it completely fails the broader populace. If you are poor, you still have legal rights, but you cannot defend yourself if they are violated unless some charity helps. It is another one of those travesties where on paper you will find the greatest equality but in practice, it could not be more unequal. Actually, if you are a crook and have money you stand a good chance of getting off the hook; maybe a much better chance than a poor person seeking redress in the courts for a perceived or real injustice.

Looking just at a few examples of what ails our seemingly most advanced and civilized countries clearly tells us that we are still far from becoming a civilized species. It will be a long way before humans everywhere on our planet can live in societies with flourishing inclusive political and economic institutions. Far away may be the day when humans finally will understand that it is our obligation as a species to ensure that every human being can achieve her or his full potential; that the dignity of human life is sacrosanct, and violence is no way to solve our conflicts; that nobody owns planet Earth but that it is a gift to all of life, including us; and that our common interests as humans far transcend the narrow and selfish interests of tribes and nations. So, will we ever get there? It does not look to good for us, does it? Wrong, a look into the past will quickly convince us that we can be cautiously optimistic. Even though there is still a long way ahead of us, we already have made tremendous progress over the past couple hundred years or so. Many key statistical indicators do tell us that we are moving in the right direction. Good news usually does not make attention-grabbing headlines so you will not hear them on the evening news. That the world is getting a better place is just not something that garners a lot of attention. Fortunately, nowadays anyone can look up the respective facts on the internet. Until I did so myself a few years ago, I had no idea that for example the share of people living on less than two dollar per day had dropped in half over the past twenty years. Neither did I know that average life expectancy is now about seventy-two years globally, a little shorter for men and a little longer for women [74,75]. So yes, we can be hopeful that we will come to a point where we can put our differences aside and join on a path towards becoming a civilized species. We must acknowledge that we do live in an interdependent world and that only together can we succeed in building a human species civilization. A human species civilization can only become a reality if nations learn to put the interests of all of their people and humanities interests first, instead of bowing to the interests of their elites. This will require elites, corporations, organizations and individuals alike, so-to-speak human society at large, to denounce greed for what it is. We all are on this Earth for a short time only. We arrive with nothing, and we leave with nothing, unless we contribute a little to make this world a better place. Greed is not good, and it never has been good. It is about time that all of us realize this and act accordingly. If we can do that in time, we may be able to address the greatest challenge for all life on Earth - the threat of extinction.

The Future of Life

"It is difficult to make predictions, especially about the future". This quote or similar ones are attributed to a number of people, among them Albert Einstein and Samuel Langhorne Clemens, better known under his pen name as Mark Twain (1835 – 1910). When people think about the future, they usually have a specific context in mind guiding their thoughts. Naturally, for most of us, this context is of a personal nature and centers on our families. In this regard, the future of our civilizations or of our species usually do not matter. And if at all, then likely only to the extent of how peaceful it will be or if we need to fear that war eventually may destroy our future, just as it has done for so many people and still continues to do so today. Some are more concerned about their own and their countries economic future; others are concerned about the future impact that humans have on Earth and what that will do not only to the natural world around us but in the long run also to us. We all are simply nothing more but guests in a hotel called *"Planet Earth"*. Checking in when we are born and checking out when we die. In addition, for some of us, religions hold out the future of an afterlife; how realistic or concrete any such future may be is very much up to believers and beyond anything we should want to contemplate or comment on; we just should respect it.

Whatever our futures may hold for us on a personal level and whatever sense of permanence we have with regard to what surrounds us in the natural world or in our human-made environments, as far as human affairs are concerned, everything is transitory, and all is perishable. That insight does not make what we as humans strive for any less meaningful, actually quite to the contrary. Our human civilizations are impressive and may have the shine of a seeming eternity, but they are nothing like that. In the larger scheme of things, human evolution and everything humans have achieved so far is only white noise. If humanity would magically vanish from Earth tomorrow, in a few million years little would be left of what we may perceive today as almost eternal. Earth, our solar system, or the universe could not care less if there is something like humanity or not. Humanity only will have meaning if we as a species can endure to give it such a meaning. Therefore, future in this context is the very long-term outlook, a future when humanity would maybe at some point in time indeed matter beyond our Earth or our solar system. So, what can we know about this far-out future? As we have seen in the first chapter, science has given us an understanding of how stars are born and how they end their lives, how solar systems form, even how far back some 13,800 million years ago our universe originated in what we refer to as the *Big Bang*. So clearly, we know something about how such things play out on a cosmological time scale. But how about when it comes to more specific predictions that have relevance in our own neighborhood, our solar system, and for the future of life on Earth specifically? Are there any limitations that restrict how far out we can predict what our solar system will look like in the distant future; and for that matter, what Earth will look like?

A fundamental insight of twentieth-century physics, quantum mechanics to be specific, tell us that in the world of the very small we can never know the position and the velocity of any object accurately at the same time. This limitation as to how precisely we can know the values of such conjugate variables like the location and the momentum of a particle is known as Heisenberg's *Uncertainty Principle*, named after the German physicist

Werner Heisenberg.[13] There are other conjugate pairs of variables like energy and time, or particle number and phase information; for any of those pairs, if we try to measure the value of one variable in such a pair precisely, the information we can get about the value of the other variable in the pair becomes less accurate. Simply put, the product of position uncertainty and momentum uncertainty of an object, or the respective product for any other pair of conjugate variables, will always be larger than a quantity proportional to what physicists call Planck's constant. Because Planck's constant, named after the German physicist Max Planck, is so small – about 626.6 divided by the product of one trillion times one trillion times one trillion – the *Uncertainty Principle* does not matter for calculating trajectories of objects in the macroscopic world we live in. However, it most certainly matters for the world of the very small, such as for atoms and electrons, the stuff from which our world is made of. At the heart of quantum mechanics are probability distributions. The physics of quantum mechanics tells us how to calculate those but only once we make a measurement will probability be turned into certainty. Only through observations are probability distributions of potential outcomes reduced to certainties, like where an elementary particle will hit a detector screen. This raised the ire of Einstein who was not willing to accept quantum mechanics as the final word and strongly believed that it would eventually be replaced by a new physics that would restore the seeming certainty of classical predictions. However, no such theory has been discovered yet and quantum mechanics has become the most successful physics theory ever devised with its wildest predictions verified to the extent that modern technology allows for experimental verification. More importantly, not a single attempt to falsify it has ever succeeded. Testament to the success of quantum theory is that practically all of our most advanced technologies in some way or another rely on it. Quantum theory tells us that at the very small scale, which is the foundation of our reality as all bigger things are made out of smaller things, we ever only will be able to predict probabilities. The future is always open and can evolve in many different ways. This is just the nature of what we call future, and we have an intuitive understanding of this innate property of future without having to take recourse to quantum physics. Some of the more fantastic predictions by modern physics in the form of quantum mechanics maintain that all possible outcomes, which is all potential futures, will eventually be realized, just not in our universe. In some of the other hypothetical parallel universes, evolution may have played out quite differently. In any of them, other species, maybe even humans, may gaze in wonder upon the stars and marvel about life on their planets, just as we do. However that may be, it really is not relevant for our own future and as far as most of us are concerned, we likely all agree that we can only make sensible statements about the reality we live in. We may not be able to say exactly where each and every atom will be and how fast it will be moving at any point in time, but we still can describe and predict the behavior of large ensembles of atoms; for example, through the macroscopic properties of gases such as temperature, pressure, and volume. So, while some may lead you to believe that quantum mechanics does not allow us to make exact predictions that is only true if we try for example to predict how a single electron may travel through a slit and where it will register on a detector placed behind that slit. It is not true if we look at large numbers of electrons where we get the expected interference picture of electrons acting like waves when sent through small slits, waves of probability whose patterns we can well predict. Does that mean that in principle we can make accurate predictions on how our macroscopic material world will behave, now and well into the future? Yes, we can, but unfortunately only in a limited way; and that shortcoming has nothing to do with any restrictions imposed by quantum mechanics.

As it turns out, we already have very real problems in calculating how three macroscopic bodies will move relative to each other under the forces of mutual gravitational attraction. The realization of this problem goes back to Newton's time. While the two body problem, the classical example of it is a planet orbiting its star, has an analytic solution, the three body problem, such as the relative movements of Sun, Earth, and Moon, has not. Newton tried his hands on the three-body problem but quickly realized that the calculations were very complex; exceeding in his opinion what humans possibly could solve. Now there is no bigger challenge to mathematicians than a problem that is simple to pose but seemingly unsolvable. However, it would take until the end of the nineteenth century before eventually the French mathematician Henri Poincaré (1854 – 1912) discovered that the problem had indeed no precise solution; but for unexpected reasons. Back then, it was quite common for eminent scientists to participate in prize competitions intended to address what were judged as the most important scientific challenges at the time. Whether the solar system was stable or not, certainly was the kind of problem worthy of a prize competition, such as the one arranged in the name of the Norwegian King to be awarded in 1889 for an important discovery in the field of higher mathematical analysis. Henri Poincaré selected the three-body problem as the challenge of his choice and set out to find a solution and with that to succeed in the competition. If any among the contemporary mathematician stood a chance to solve this problem, certainly Poincaré would have been among the ones topping the list of potential winners. Poincaré thought he had found the solution he was looking for and was indeed awarded the prize for this effort. However, instead of solving the problem he had made it only more intractable. He quickly discovered this himself when to answer questions from the prize committee he needed to review some of the assumptions he had made in his calculations. What at first seemed like a major blunder was actually in hindsight one of the great discoveries in physics as Poincaré had incidentally stumbled into a field of mathematics that would evolve into what we know today as chaos theory. Poincaré had discovered that seemingly tiny changes in the initial conditions of his calculation, such as the relative positions and velocities of the three bodies, led to very large differences in the calculated trajectories for these three bodies at later times. There is no precise solution for the three-body problem, as the initial conditions can never be known accurately enough. Since the current orbital patterns of the planets in our solar system were established shortly after it had formed, they have remained stable. We may not be able to predict trajectories of small asteroids with precision for more than a few hundred years but we can do so for much longer for our solar system planets; as long as nothing undue happens. However, we know that there is a chance that Jupiter's gravitational pull on Mercury may change the latter's orbit in unpredictable ways which could produce orbital chaos among our inner solar system planets. It is fortuitous that this has not happened yet and hopefully it will stay this way with our planets maintaining their current orbits for a few more billion years. But if it happens, we would not be able to predict the outcome; such are the consequences of the three body problem.

The future of life is not just subject to the vagaries and unpredictability of the physical world, it is also closely tied to our very own future as a species. To inform ourselves about the latter we have to look at two different aspects. First, we need to understand what the past can tell us about the future. Second, we need to acknowledge that our species is the first animal species that is not only subject to the forces of evolution but has actually become a major force in shaping the evolution of life on Earth itself. In a way, life on Earth is a constant scramble where the chairs on the passenger deck of the ship *"The Life of Earth"* are continuously rearranged. From time to time, some species are washed overboard and new ones are being brought up from decks below; and sometimes the whole ship hits an iceberg, not unlike the Titanic did, with much of the life it carries lost, except for whoever makes it to the lifeboats in time. Eventually, the survivors will build a new ship and set sail again and the whole story seems to replay itself, just with mostly different passengers; some of them will make it through the next wreckage but many will perish, again setting the stage for new rounds to follow in recurring cycles of creation and destruction. Very few will be so lucky to make it through several such voyages of life; eventually even the hardiest characters run into an obstacle that stops them dead in their tracks before they can make it to the lifeboats or stops them long enough for all lifeboats to have already left for safety. As many will guess correctly, the above paraphrases the mass extinctions that life on Earth has experienced several times. Will there be more such mass extinction events in the future? Is the human species the cause of the next mass extinction that unbeknownst to many may already have started? Of course, this last question relates to our very own species characteristic, the ever-increasing ability of humans to shape their own future. Unlike other successful species before us, as for example the dinosaurs who ruled Earth for more than 150 million years, something we are still far from achieving, the human species is in a very different position. We are not just like other species characters in the drama of life that seems to be played in varying productions in theater Earth over and over again. We have become co-authors of this play; we write the roles that we then confer onto other species. It is us, who in most cases, often unknowingly, decide who gets cast out of the play and we may be soon in a position to create new characters to participate in the play. This has never happened on Earth before. Our Earth and life on it have been changing in the past in many ways and will continue to do so. However, the human species is the first species ever that can actively react to such changes. We truly are not just subjects of history as it evolves, we can increasingly write our own history and the history of life on Earth. Humans are the first ever species able to plan well into the future and have increasingly the means to realize plans that in the past we would have thought only supernatural powers could execute. That must have an impact on the future of life and on our future.

After this short excursion into some of the real or perceived limits of what we can predict about the future of the macroscopic world and how our very own species character may affect our future it is time to take a closer look. We start out with what we can know about the long-term future of Earth itself. From there we will look into some of the major disruptive events that life on Earth has experienced since the explosion of Cambrian life. After that, we will zoom into what we can know with respect to the near-term future on a geological timescale. Finally, we will reflect on how we shape our own future and by doing so the future of everything else on Earth.

The End of Habitable Earth

What can we know about the future of our planet Earth? How much longer can we expect conditions on Earth to be favorable for life? Contemplating our potential long-term future, we have to acknowledge that when it comes to planetary evolution or the evolution of our solar system governed by the physical laws of our universe, we as a human species just do not matter much at all. Today we can look back on more than 3,800 million years of life on Earth and a little more than 540 million years of higher lifeforms evolving on Earth. Almost certainly, this picture will continue to change as we learn more about the evolution of life on Earth, but its broad outline will most likely remain. However, what will not change is what we already know with certainty. Life on Earth is figuratively speaking beyond half-time and there will not be another 3,800 million years for life on Earth to continue. The time span that remains for higher life on Earth is much more likely somewhere in the range of 800 to 1,100 million years. That still looks like an eternity perceived from our brief existence as a species of some 5 to 7 million years, dating from the time the human and chimpanzee lines split. Looked at through the lens of an average human lifetime, the time remaining for life on Earth seems even more incomprehensible. So why bother? What matters for the evolution of life on Earth is not the comparatively short time span that it takes for a species such as us to evolve. What really matters is the evolution of life itself, long before it can sustain such complex and intelligent lifeforms such as us. That takes a considerably longer time, in our case some 540 million years. In this context, if life had to start all over on Earth for whatever reason, the remaining roughly 800 to 1,100 million years do not look like an eternity anymore.

How do we know that our planet will become inhabitable for species like us sometime 800 to 1,100 million years into Earth's future? To get an answer to this question we need to look to physics, or more precisely, astrophysics. Our Sun, the star in our solar system, is subject to the laws that govern all stars. As discussed in the first chapter, over roughly the past one hundred years or so we have learned a lot about the birth, life, and death of stars. Since our Sun was born about 4,600 million years ago, its luminosity, the measure for the total energy our Sun outputs per unit time, has steadily increased. Remember, the flip side of this luminosity increase is the weak early Sun problem we discussed in the second chapter, where science is still trying to understand how Earth could have had early oceans with such a weaker Sun. The rate at which our Sun's luminosity increases is equivalent to a $\sim$1% gain in energy output roughly every 110 million years. Because of that, our Sun's luminosity is today about 1.43 times higher as compared to when it started its life as a star by fusing hydrogen into helium in its core, some 4,600 million years ago. This quasi-linear increase of our Sun's luminosity will continue and result in a roughly 30% luminosity increase over the next three billion years. Unfortunately for us, if humanity should still exist then, already a 10% increase in luminosity over the next some 1,100 million years or so will alter environments on planet Earth dramatically. A 10% luminosity increase will change conditions on Earth such that life as we know it will no longer be able to exist. While a 10% increase in luminosity does not seem like much, it is sufficient for the habitable zone around the Sun to move significantly further outwards in our solar system, away from Earth.

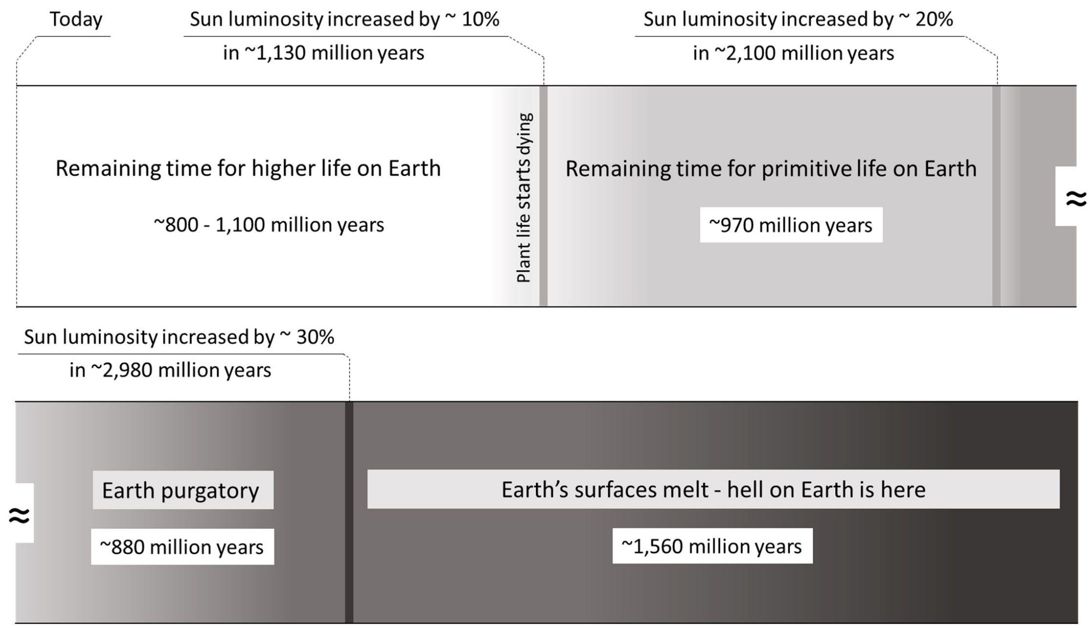


Figure 4.2: Expected increase in Sun luminosity over the remaining roughly 4,500 million years Earth may still exist. Conditions suitable for sustaining higher life can be expected to last maybe for another ∼1,000 million years.

Earth's orbit does not change and in a distant future Earth will no longer be in our solar system's habitable zone that can support higher life. Plant life is likely first to perish. With water becoming ever more scarce and humidity and temperature rising, planet Earth will cease to support anything else besides very primitive but sturdy lifeforms. Earth at that point will certainly not support anymore the kind of multicellular life our species represents. The lifeforms that arrived first on Earth and were able to survive in the much more inhospitable conditions of early Earth, these hardiest among bacteria and archaea will likely be the last ones to go.
In about three thousand million years, the surface temperature on Earth will have well exceeded the boiling point of water and by that time even prokaryote life will be gone. Eventually, with our Sun's luminosity further increasing, Earth surface will be in a molten stage in about four billion years. It seems as if Earth will return to resemble more and more its early phase, the Hadean, Earth first geological period, lasting from its formation until 4,000 million years ago. Whatever Earth will look like then, it may not be too different from some of the visions of hell that Christian theologians have been conjuring up in the not-too-distant past to frighten their flocks into God-fearing obedience. Where they were wrong is that hell is not a place in some unspecified location that awaits sinners after they die. Figuratively speaking, hell is what awaits life on Earth with the heat uncomfortably high for most lifeforms in about one billion years and reaching temperatures high enough to melt Earth's surface in about four billion years.
As discussed in the first chapter, our Sun's eventual destiny is to first evolve into a red giant star and to retire finally as a white dwarf star. Such is the fate awaiting stars that start their stellar life with a solar mass somewhere between approximately 0.3 to 8 times the mass of our Sun. In its last stages as a red giant star, our Sun will expanded so

enormously, that it will first consume Mercury, then swallow Venus, and finally it may also engulf Earth. Some predictions seem to indicate that there is a chance Earth may just stay outside the expanding Sun but that is a minute detail that will not make any difference for the fate of life on Earth. The fate of life on Earth will have been sealed long before our expanding Sun will obliterate the inner rocky planets Mercury and Venus or before it comes even remotely close to swallowing Earth. Regardless, the most likely scenario seems to be that our Sun's perimeter will continue to expand beyond Earth before our Sun will finally collapse to spend the rest of its star life as a white dwarf star. Nobody will be around to watch the final scenes in the cataclysm of our star as the Sun consumes the inner rocky planets. Unless, humans long before then have become a space-faring species and can observe the death of Earth from a safe distance. However, is it realistic to assume that humans will be able to survive as a species so long and eventually master space travel to find new homes for the life of Earth among the stars? If we leave out single-celled life such a bacteria and archaea and restrict ourselves to higher lifeforms, no species has ever managed to hang on for that long. The human species has about 1,000 million years left before Earth will become inhospitable. That is all the time we have at best to evolve from a planetary species bound to Earth to a species capable to find ourselves new homes among the stars.

Modern humans have been around for just a few hundred thousand years or so. For the sake of simplicity and because we are only interested here in rough estimates, we will take two hundred thousand years as the yardstick, knowing that modern humans have been around a bit longer. Persevering as a species for another 1,000 million years would mean on that scale to endure around 5,000 times as long. Or looking at a generational count assuming that on average humans tend to have procreated by their thirties this means that for us to be still around some 1,000 million years into the future we are looking at 33.3 million future generations to come after the roughly 6,600 generations that have gone by in the past two hundred thousand years.

Is it meaningless to think about anything so far in the future? If we continue in our current ways where we often cannot see much past our noses it indeed makes no sense at all. Only if we possessed something amounting to a species consciousness would this future be relevant for us. Unfortunately, our species possess nothing like that; at least not yet. If we did, we would react quite differently to the threat that climate change poses. Seemingly, we still have not understood that we are ruining the basis for our own existence on this planet as we continue with our greedy and shortsighted exploitation of each and every resource on Earth just as if Earth's resources were inexhaustible; but they are not. Muddling through may have been a viable strategy for our species a long time ago when there always was a new and untapped resource to replenish what we had already exhausted. Not so anymore. Nevertheless, we likely will not have some 1,000 million years to evolve a species consciousness. Mass extinction events have eradicated many species in the past and likely will do so in the future. Humankind could be wiped from planet Earth long before Earth will become inhabitable.

The Predictability of Mass Extinctions

In the previous chapter, during our brief journey through the evolution of life on Earth we saw that multiple times since the evolution of higher lifeforms started at the end of the Precambrian, life on Earth has been severely decimated in what are commonly referred to as mass extinction events. Decimated is actually the wrong word, rather the complimentary applies to mass extinction events, as more often than not they eradicated 70-90% of species at the time and not just every tenth, as this is what decimation really means.[14] In many ways, life on Earth has been shaped by these mass extinction events and arguably, without the last mass extinction some 66 million years ago wiping out the dinosaurs, mammals would not have had the opportunity to radiate as they did after the demise of what had been the dominating species for many millions of years. Consequently, it seems unlikely the human species would be enjoying its place in Earth's life today without this last mass extinction event some 66 million years ago wiping the slate clean. Major mass extinctions greatly reduce biodiversity or to put it more starkly, they clean the slate by consigning many species to extinction while providing opportunities for new ones to prosper once ecosystem have recovered. Of course, life is always precarious, and species can become extinct without a mass extinction event being responsible. We know that a large percentage of all species that ever existed on Earth are extinct. This fraction is estimated to be in the high ninety percent. The vast majority of all species that ever lived are extinct today, provided our species count is correct, as to which there is still some doubt. The more we look for new species the more we seem to find and just over the past few decades, we have found far more species of bacteria and archaea than we ever thought existed; and new species of multicellular life are still being discovered today. Just think of the lifeforms quite recently observed for the first time on the deep-sea floor around thermal vents that do not need the Sun to harvest the energy they rely on to sustain themselves. The high ninety percent estimate for the fraction of species extinct of all life that ever existed on our planet may have some uncertainty. Even though, given what we know about the average lifespans of species and that first life on Earth emerged at the latest some 3,800 million years ago, everything still points to the vast majority of all species ever existing on Earth to be extinct today.

The major extinction level events get their names from the geological periods during which they occurred, and we have encountered them already on our journey through geological time. It is not a coincidence that mass extinctions tend to occur as one geological period gives way to another. Boundaries between geological periods are associated with major changes in Earth geology and climate as this is what defines them; of course, any such momentous changes can take their toll on Earth's life. The evolution of higher lifeforms really started with the Cambrian explosion some 541 million years ago. Therefore, for good reason, we know about no mass extinction event before the Cambrian period. Could there have been events before the Cambrian that would have resulted in mass extinctions of higher life if it had existed in similar abundance before the Cambrian? Most certainly so. For example, any of the extreme glaciation phases in early Earth history would have put such higher lifeforms to the test. There have been a number of such events. In the geological periods preceding the Cambrian there was the Cryogenian glaciation, which may have resulted in an Earth almost completely covered in ice; it is still debated as to how much that amounted to a *Snowball Earth* versus a *Slushball Earth*. The Huronian glaciation almost 2,500 million years ago is the first severe glaciation event we know of and scientists believe it likely produced an earlier *Snowball Earth*. In addition to major glaciation events, there is the dance of the continents as they come together to form supercontinents only to separate again once that is achieved. All of this plays out over many millions of years and sometimes in can result in constellations, which if accompanied by some other unfortunate combination of factors can wreak havoc on Earth's life. This has happened several times since the Cambrian. While we only have the record for mass extinctions of higher lifeforms since such life evolved some 541 million years ago, we can almost be certain that mass extinction type events have been a regular occurrence throughout Earth geological history. Scientists identify five mass extinctions occurring over the past 541 million years.

The first mass extinction event we know of is referred to as the Ordovician extinction sometimes also called the Ordovician-Silurian extinction. It happened some 444 million years ago which is only about 100 million years after the Cambrian explosion of life. It is estimated to have wiped out 60-70% of all species at the time. While there is still speculation as to what exactly caused it, it seems to be agreed that the most likely reason was a short but intense glaciation period, the Ordovician ice age, sometimes also referred to as the Andean-Saharan glaciation. Water removed from oceans and turned into glacier ice resulted in sharply lower sea levels. We need to remember that by the end of the Ordovician only the first primitive plants had colonized land and life was still very much constrained to the sea so all complex multicellular lifeforms were marine organisms and a large number of them became extinct during the Ordovician extinction.

The next mass extinction event, commonly referred to as the Devonian extinction, occurred about 375 million years ago. It happened over a roughly fifteen-million-year time span between 375 and 360 million years ago and resulted in the extinction of up to 75% of all species. By the Devonian, animal life had followed plant life onto land but similar to the Ordovician extinction, the Devonian extinction seems to have mostly affected marine life. The Devonian extinction event was spread over a longer time period and a number of different factors may have contributed to it. No likely single root cause has been identified but is looks like changes in plant life during the Devonian period resulted in conditions that produced oxygen-depleted oceans or may have decreased the level of carbon dioxide in the atmosphere resulting in a cooling climate. The result was another ice age, the so-called Paleozoic ice age, sometimes also referred to as the Karoo glaciation or Karoo ice age.

The Permian extinction took place about 252 million years ago and is also referred to as the Permian-Triassic extinction event. It resulted in an enormous loss of biodiversity and for that reason it is frequently called *The Great Dying*. Marine life took the greatest hit with about 95% of species becoming extinct. This time, life on land was also hit hard with an estimated 70% of vertebrates being lost. *The Great Dying* is the only extinction event we know of that nearly wiped out all insect life. As to what caused it there is still speculation but it now looks like that there were two, maybe even three sequential extinction events. Some think potential catastrophic events more likely such as an asteroid impact, volcanic activity, or a sudden release of methane bound in methane hydrate on the sea floor. Others argue that processes that are more gradual could also have been responsible.

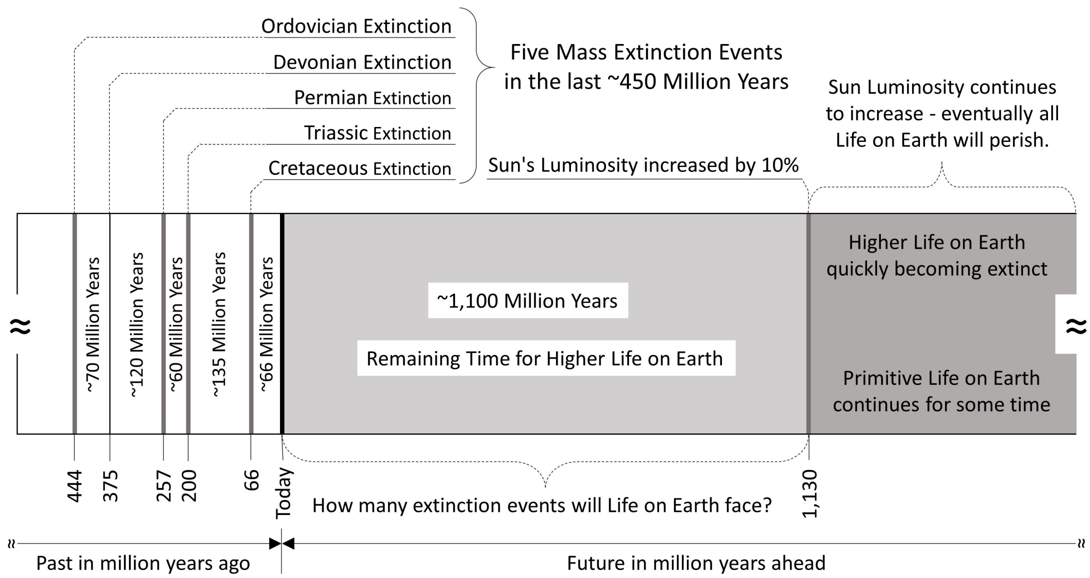


Figure 4.3: The five known mass extinctions in Earth past. We can count on more such events occurring during the remaining ∼1,100 million years for higher life on Earth.

Among those are changes in sea level accompanied by oxygen depletion of the oceans as well as much drier conditions on land. Pangea was almost completely formed by the end of the Permian and much of its interior would have been quite arid.

The next extinction event on our list is the Triassic extinction, sometimes also referred to as the Triassic–Jurassic extinction. It occurred about 200 million years ago, and it resulted in the loss of close to half of all species on land and seems to have reduced marine biodiversity by about one third. There is no agreed explanation as to the potential cause for the Triassic extinction but the usual suspects are being considered. This includes the release of huge amounts of carbon dioxide due to increased volcanic activity caused by the breakup of the supercontinent Pangea resulting in runaway global warming; an asteroid impact is also a candidate. Other possible scenarios scientists discuss are changing sea levels and increased acidity of the oceans due to large amounts of carbon dioxide being absorbed by the oceans. The latter off course could have been due to increased volcanic activity.

The last mass extinction on our list is the Cretaceous extinction also referred to as the Cretaceous–Paleogene or K-T extinction. It happened around 66 million years ago and wiped out about 75% of all species. With few exceptions, all large animals became extinct, including the non-avian dinosaurs. This extinction event marks the beginning of mammalian dominance. As bad as it was for all species which perished because of it, we have to be grateful for it. Without it, mammalian life - and with it eventually human life - would likely not have evolved the way it did. There is agreement that the most likely cause for this mass extinction was the massive asteroid impact creating the Chicxulub crater in the Gulf of Mexico.

Fig. 4.3 shows a simple graphic of the timeline including the five mass extinction events as well as the remaining time that Earth can be expected to support higher life on Earth. What should we read from this graphic? Specifically, as we know that before the existence of abundant higher life on Earth, likely several more events would have easily qualified as potential mass extinction type events. Most would probably say that we can quite likely count on more such mass extinction events happening over the roughly 1,100 million years remaining for higher life on Earth. However, at the same time, most also will likely not be concerned about this because the time scales that we are looking at here are really beyond what we can imagine. The oldest fossil records of the genus *Homo* date back just about 2.5 million years and modern humans arrived only about 200,000 to 300,000 years ago (see fig. 3.13). Our recorded history only goes back a few thousand years. How can we even begin to imagine what humanity will look like in a hundred million years or even a million years? Dinosaurs ruled on planet Earth for some 150 million years as the dominant species. For now, humans may be able to claim only a few 10,000 years in this role and that is likely a stretch.

The average time span between the five mass extinctions, from the first one about 444 million years ago to the last one about 66 million years ago, is roughly 90 million years. Based on a very limited statistic and an admittedly simplistic extrapolation that assumes the average interval will stay the same between such events, we could expect the next extinction event to occur in about 30 million years. The assumption of a continued ∼90 million year interval between mass extinction events is most certainly wrong but that is not relevant here. What matters is that such an event could happen in a few years, or it could happen in 200 million years, but the more time has elapsed since the last such event happened the more likely a new extinction event is to take place.

When looking at those five mass extinctions we need to keep in mind that even though we refer to them as events they much rather resembled extended processes. Usually, the word "event" implies something sudden that happens with immediacy and then is over. Nothing could be further from the truth. This is even the case for the asteroid impact where we definitively have a single event starting the mass extinction, which likely did have a catastrophic immediate impact on many species. However, many more species on different continents very likely survived the asteroid impact but then had to struggle to survive through what followed after, which practically must have been the equivalent of a nuclear winter. For several of the mass extinctions we know that they played out over millions of years, some of them unfolding over time spans much longer than the human species is around. This shows us that probably for many such occurrences there should be ample time to figure out survival strategies. In addition, our human species, with the technology we have available to us, should be in a much better position to cope with such events than other species in the past. While in some respects this is likely a good assumption, there are other factors to consider. Given how many there are of us today and how much we have come to rely for our very survival on our highly developed civilizations functioning, we could be in for a surprise.

With most mass extinctions unfolding over millions of years, when would we actually realize that we are in the midst of one? Some would argue we are already living through the next mass extinction and we humans are the major cause of it. Accordingly, there are some advocating for the naming of a new epoch following the Holocene: they call it the Anthropocene, or the human epoch. This new epoch is not yet an official subdivision of geological time but may well become one in the near future. There are various proposals for the start date of the Anthropocene. One seeks to tie it to the disappearance of Earth's megafauna which occurred in pulses starting some 50,000 years ago with a late prominent example being the demise of the woolly mammoths at the end of the last ice age, right at the beginning of the Holocene. Another option would be to associate the start date with the beginning of the *First Agricultural Revolution*. However, these Anthropocene start dates will collide with the definition of the Holocene. A better choice may be to correlate the start of the Anthropocene with some more concrete traces in any geological record future generations may find. To mind come the fall-out from the nuclear weapons tests that started in the last century, or the thin layer of highly compressed plastic that our civilizations will leave behind preserved in Earth geology, or maybe the trace record of accelerated global warming that we are producing now. Any one of these three would make for a good geological marker for the beginning of the Anthropocene and provide a proper bookmark to end the Holocene. As this would make the Holocene a very short geological epoch an alternative could be to convert the Holocene from a geological epoch to the first geological age in a new Anthropocene epoch.

While mass extinctions play out over very long times, the specific event starting them, such as the asteroid impact some sixty-six million years ago, can still have devastating immediate consequences. Fortunately, unlike the dinosaurs, we can understand the past and learn from it. Dinosaurs had no option other than to just continue procreating and accept whatever was coming their way. Their evolution took them down a different path than ours did. Lasting almost five hundred times longer than that of modern humans, their evolutionary path never led to the development of the kind of complex societies that characterize our species. With that, they lacked the vital ingredient that seems to have spawned human culture evolution. Consequently, they never were in a position to evolve into a species that could produce the knowledge to understand its place on Earth and the universe surrounding it.

Dinosaurs were never capable to realize the extent to which their own evolution as a species over their roughly 150-million-year rule on Earth may have contributed to the extinction of other species. They certainly had no way to calculate the trajectories of asteroids potentially threatening Earth and their survival as a species. It looks like Dinosaurs just got lucky enjoying their $\sim$150 million years of dominance on Earth. Of course, the basic blueprint of their species turned out to be a very successful one and continues in a sense to this day with their modern descendants, our bird species. However, dinosaurs did not evolve any of the capabilities that would have allowed them to take control of their destiny as a species. After some 150 million years of dinosaur rule, no dinosauroid evolution had taken place comparable to the humanoid evolution that produced us. Because of that, dinosaurs can be excused for not having figured out a way for their species to survive mass extinction events such as the one that wiped them out. Humans have no such excuse.

We increasingly understand how our species has affected life on Earth. Different from dinosaurs, humans can understand the potential consequences of mass extinction events whether caused by external events such as asteroid impacts or by gradual changes to our Earth environment, some of which are of our own doing. There is no excuse for the human species not to act on that knowledge. Or for that matter, not to do everything we possibly can to ensure the long-term survival of life on Earth. However, there is a problem. We know disturbingly little about what really caused any of the major mass extinction events except for the most recent one. How can we prepare for something for which we do not know the root cause? As the question already implies, the answer has to come in two parts; we just need to separate causes from consequences. Knowledge about what causes a mass extinction event to occur in the first place, even though unlikely to help us prevent it from unfolding, should give us a much better idea of how frequent the specific preconditions for any such root causes may actually be. A good analogy, albeit on a much smaller scale, would be what we already do with monitoring seismic activities in order to judge the likelihood for future earthquakes. In a similar way, we can detect changes in a number of parameters seen as the most likely indicators to precede mass extinction events. In some cases we are already doing this, like with monitoring the acidity of our oceans or tracking the amount of carbon dioxide in our atmosphere.

The second part in the above question deals with the consequences of mass extinction events. That for example could include developing solutions for how we could survive potentially several years of reduced sunlight because of particulates in the air blocking it from reaching Earth surface. Those particulates could be coming from extreme volcanic activity, or they could be produced by an asteroid impact. In any case, we would have to come up with solutions that ensured our and our life stocks food supply as well as anything we could do to protect the diversity of life on Earth. Therefore, it comes down to understanding the real likelihood of certain root causes for mass extinctions to occur and to be ready to deal with the specific major fallout from any such events. Coming back to the earthquake example: Do we not expect our leaders at the local, state, and national levels to have put in place solid plans and the necessary resources to be able to execute them when the next great earthquake strikes? We likely all agree on that, but how many of us believe that this is indeed the case? We all know the answer. If we approach our survival strategies and preparation for potential extinction level events with the same zeal, there will be no hope for us.

How we deal with the realities of climate change and its impact on the life of Earth is another instructive example. According to some of our eminent leaders, we are seemingly still stuck debating the root causes of climate change. Some politicians keep insisting that there is no evidence for climate change being caused by humans. Our world's experts on climate change, the overwhelming majority of scientists and scientific institutions, tell us that global warming and climate change are largely human made. Some of the same politicians are seemingly also incapable of understanding the difference between climate and weather. What else is this owed to but industrial and corporate monetary interests corrupting our democracies? It is the *Iron Law of Oligarchy* at work, a classic example of industrial oligarchs co-opting government and its regulatory bodies to serve their special interests. Fortunately, the majority of world leaders understand the critical issues at hand, and they have moved beyond debating the reality of climate change, if ever so slowly. They seem to comprehend what is at stake but that does not make it any easier to agree on how to address the root causes of climate change and how to fairly deal with its consequences. The process of doing so will be painful and will require compromises on all sides. This is the very first time that humanity must deal with what we could call a species level challenge. If we can do this successfully, we can be hopeful that we also will be able to address much bigger species level challenges that undoubtedly humans will be confronted with one day.

The most immediate threats to the survival of our species and to the continuation of life on Earth are coming from us. They are not related to external chance events but are a result of how our human societies have evolved and continue to evolve. The destructive potential of our thermonuclear weapons arsenals is still sufficient to wipe out most of humankind multiple times over. Any survivors of an all-encompassing thermonuclear war will face conditions that are likely to drive the human species to extinction. Since the two atomic bombs that obliterated the populations and cities of Hiroshima and Nagasaki, no other nuclear device has been used in warfare.[15] Although millions, mostly civilians, have been killed in conventional wars and genocides since then, mutually assured destruction has kept a fragile peace avoiding nuclear war. However, none-state actors have now joined in the slaughter of innocent people using conventional weapons and nobody knows what will happen if they ever should get their hands on nuclear weapons. How long can this fragile and imperfect peace last? The political rational of nations may continue to ensure no nuclear weapon will ever be used in war again, but no system is fail-safe. Probably unknown to most of us, in several instances humanity has only been a button push away from all out thermonuclear war. For preventing nuclear war, we have to thank unsung heroes who had the courage to let their better judgment rule the day. Two known examples are the refusal to act on erroneous information that suggested an enemy nuclear attack was underway in 1983 and not letting a seemingly desperate situation trigger the launch of a nuclear weapon in 1962.[16,17]

The risk of a system malfunction triggering an all-out nuclear war may be small but it is not zero. In the very long run, we can almost be certain that such a system malfunction will eventually happen. The very long run in this case translates to something much shorter than the geological time scales over which past extinction events have been taking place; not many millions of years, but still much longer than human lifespans, say on the order of ten thousand or one hundred thousand years. Can we avoid such an outcome by developing completely fail-safe systems? No, because completely fail-safe systems are not possible. The only option to reduce the risk of an accidental thermonuclear war to zero is to get rid of all thermonuclear weapons. Only a world without nuclear weapons will be a safer world. A world where individual nations horde and research such weapons is also a world where none-state actors can seek to procure them for a prize to achieve their vile purposes. Humanity must rid itself of such weapons and keep the respective knowledge under strict lock and seal.

The arguments just made with regard to thermonuclear weapons also hold true for other weapons of mass destruction with a potential outcome that is not controllable by humans anymore. Biological or chemical weapons, sometimes also called the poor man's nuclear weapons, come to mind. Just think about the risk of an extremely lethal virus getting out of one of those biological weapons labs and mutating to the extent that it becomes a global killer against which no antidote can be found in time. The human species is unlikely to be around as long as the dinosaurs lasted if it cannot transform itself into a species that is at peace with itself. If our species cannot achieve this we may as well not bother developing technologies that could help us survive externally triggered mass extinction events such as the ones in fig. 4.3 as we likely will have committed species suicide long before extinction events happening on a geological time scale may ever pose a threat to us.

An Ever-Changing World

Change is a common thread that connects our brief exploration of the solar system and the universe, our journey through Earth history, and the evolution of life on it. If there is one constant, it looks like that is change. The specific kind of change of course varies and often change is not immediate but plays out over sometimes very long time scales. Some of those changes seem irrelevant for us today because they encompass such vast time spans to make them seemingly incomprehensible. However, even if that is so, they still do matter very much. Without those changes happening in the past, their magnitudes and durations dwarfing anything that may be meaningful on our human scale, we would not be here today. There is the very long-term evolution of our solar system. We just have seen what we can expect for Earth future a few pages back, but does anybody really care what may happen over the next roughly 1,000 million years? Most of us would think not and would agree that for us as individuals it may be interesting to know such things; but do they really matter? The likely answer is going to be a resounding "No". After all, how can such things so distant in the future matter to us in the first place? Well, that very much depends on us, or more accurately on our very own perception of ourselves. If we see ourselves as only one of the many species that have come and gone on Earth, then the human species will at best become a highly interesting object of study to any other intelligent lifeforms that may or may not evolve after our species has vanished from Earth. In that case, the very long-term future will not matter to us at all. However, if we believe that as a human species we have a destiny, a mission to fulfill in the broader scheme of things, if we truly believe that as a species, we have a responsibility to be the steward of life on Earth, then yes, even those long time scales do matter.

When Noah built his arc, he certainly believed in such a mission; but his decision was a rather easy one as his God suggested the whole scheme and declining this mission was certainly not an option for Noah. It is an innate, deeply felt sense of responsibility for life on Earth that expresses itself, or more accurately that we articulate in our better moments, in stories such as the biblical story of Noah and his arc. Maybe evolution works in much smarter ways than we can imagine, and it may even include the evolution of a species that eventually manages to leave Earth and spreads Earth's life among the stars. Naturally, we have been bound to look at the evolution of life as a terrestrial phenomenon as for such a long time it seemed as if life was constrained to Earth. Scientists have been suspecting for quite some time now that this is not the complete story and we may well find traces of non-terrestrial life, however simple it may look like, somewhere in our solar system. When this eventually happens, it will only be the beginning of the much bigger realization that in all likelihood life is a truly universal phenomenon and not a local one confined to a small planet we call Earth, or an undistinguished solar system like ours. Our Sun is not the only star in our galaxy, which hosts 200 to 400 billion of them; and our galaxy is not the only one in our universe. In the early 1920's, Edwin Hubble was first to proof that Andromeda is a separate galaxy and not a part of our own Milky-Way Galaxy. Since then, we have learned to count galaxies in trillions or put differently in millions of million if that makes it any easier to grasp. We can assume with certainty that life is widespread throughout the universe. Therefore, what could be more natural than for the evolution of life to work on spreading the life of Earth out into our solar neighborhood? Looking at human evolution from this perspective we could well be just another evolutionary strategy that is part of the miraculous nature of our universe, how life spreads through our universe. Life started in the oceans, then moved onto land and from there into the air but still beholden to Earth gravity. Is it not natural to think of life as continuously evolving solutions to spread life itself ever farther, crossing one frontier after another as it does so? Following this logic, should we not expect life to leave Earth at some point in time, spreading out into our solar system and eventually among the stars? The only question being what form of life would achieve this. Can human evolution be a path towards such an attempt?

Provided we do not destroy ourselves and that our species is not extinguished by a chance mass extinction event because we failed to plan for such eventualities we may have a few hundred million years to develop the means to spread life among the stars. Just look at the technology development that we have seen over the past roughly two hundred years. Can anybody imagine what humans may be capable to achieve over the next say one hundred thousand years, or in a million years or in hundred million years; most of us would likely think not. Our visions are usually more limited. When yours truly was growing up, children and adults alike were still captivated by the stories of authors like Jules Verne, the French writer, frequently acknowledged as the father of science fiction literature.[18] Many of his visions of the future have indeed come true, we have gone to the Moon and back and we are finally becoming capable of exploring the still largely unknown deep seas. However, what is even more astounding are all the things that we use in today's world on a daily basis which would not even have been plausible fantasies back in Jules Verne's time.

Some may be tempted to believe that we have already explored much of what there is to explore for us humans and have discovered much of what science can help us to understand. We have been there before. Several times in the past, scientists fell into this very same trap, believing that everything there was to know already had been discovered and explained. This is why for example the young Max Planck was told in the 1870's that he should not study physics because there was nothing new to discover anymore and all there was left to do was to fill in a few blanks; at least that is how the anecdote goes. Fortunately, Max Planck did not heed the advice and became one of the fathers of quantum mechanics. Similarly, as we saw in the first chapter, until recently we thought we finally understood much of the universe only to discover even more recently that we cannot explain about 95% of its matter and energy content. We have managed to put a few pieces together that allow us to marvel about the wonders of the universe like no other living species on our planet can. However, that should not lead us to believe that there is not so much more to know and understand, more than we likely are able to imagine, ever.

Let us surmise that the evolution of life is not something that is constrained to Earth. The enormous vastness of the universe we live in is mind-boggling; filled with trillions of galaxies with each of them containing countless stars, some of them including more and some of them less than the roughly 200 to 400 billion stars in our own Milky Way Galaxy. As is often said, there are more stars in our universe than there are grains of sand on all of Earth's beaches. How could anybody in their right mind believe that the evolution of life is the prerogative of a small planet we call Earth, circling a star two thirds out from the center of a galaxy we call the Milky Way, one among a trillion of galaxies in our cosmos. Why would the evolution of life as we have come to understand it over the last 150 years not be a feature of the universe rather than something magically confined to our solar system? From what we know today, Giordano Bruno was quite right to think that there may be many other countless worlds out there in the vastness of our universe, full of life that does not answer to our religious authorities or any other power on Earth for that matter. What seems to us self-evident today was some three hundred years ago still punishable by death; as happened to Bruno when in 1600 he was burnt alive on the stake for this perceived heresy. Unfortunately, we are still not beyond killing people for nothing else than what they believe or think. What does that say about our species?

Several hundred million years ago, life may have made more attempts to move onto land than we now can discern from the fossil record. Today we look at Tiktaalik as the likely critical link that enabled our ancestors to move onto land; it may not have been Tiktaalik's species but one closely resembling it that was our first ancestor to leave the oceans. Of the other lineages that failed in their efforts to move onto land or whose lineages ended later without leaving a trace, we do not know. Could a similar kind of evolution be working to launch planetary bound species to a life among the stars? Could the evolution of a species like us just be one of countless trial runs in another type of evolution that works on a much larger scale? An evolution that is at work throughout our universe. We could be just one of evolution's trials to produce a species fit enough to survive as it spreads life among the stars. Maybe sometime in a distant future, someone will give this species successfully making the transition to life among the stars its very own name, so-to-say as the missing link connecting planetary life with space bound life; just as we did with Tiktaalik. Will that species be us? Surely, as the evolution of life must be a universal phenomenon, we can count on many other species out there in distant galaxies eventually making this transition, if they have not already done so.

We always have been curious if there is life out there beyond Earth. There is much speculation about the possibility of life in our own neighborhood, such as on Mars or on any of the many moons of the gas giants Jupiter and Saturn. Increasingly, that speculation now also extends to the ice worlds of the planetary bodies farther out such as Uranus or Neptune. As always, our visions have been constrained by what we can know and what we can imagine beyond what we know; and as what we know changes so does our imagination. A good example is our long-standing presumption that life would require sunlight to evolve and survive. We know today that this is not true, from life found at the very bottom of our Earth's oceans, surviving in what seem to be inhabitable conditions. Ecosystems of life can evolve as long as they can find an energy source that sustains them. Sunlight is one option to gain that energy, chemical energy is an alternative one.

We can be confident that eventually we will find signs or traces of life, however primitive, on some planetary bodies or moons in our solar system. That is, once we have developed the capabilities to explore for it in the places where it may be found. Some seven hundred million years ago, Earth was almost completely covered by ice and must have looked quite barren and lifeless to any fictional non-terrestrial explorer. However, even if life on Earth's surface would have been completely extinguished at the time, we know it was not, there would have been surely colonies of bacteria and archaea surviving on ocean floors around hydrothermal vents deriving their energy from sulfur. The instruments available to astrophysicists today allow them to scan planets or moons for evidence of life-supporting environments in ways that were inconceivable a few decades ago. What they are finding is that there are likely many more potentially life-supporting places out there than we would have thought possible not too long ago. The odds are that we can almost assume with certainty that somewhere out in the vastness of our cosmos the evolution of life is at work just as it is here on Earth.

We have to ask ourselves if we can muster the strength to evolve into a true space-faring species. If not, will we remain an Earth-bound species and either become eventually extinct or, in case we survive long enough, go down with our planet when our Sun will finally torch it? Counting from the time the major animal body plans first appeared, it took evolution give or take some 540 million years for intelligent life to emerge. Short of a major catastrophe such as a mass extinction, external or self-inflicted, we may have about the same time left to evolve from an intelligent terrestrial species to a space-faring species spreading Earth life among the stars. A tall order most certainly, but anything must seem possible in such a long time span. However, we should not wait too long before we get started. It would be a smart move to make sure that we establish human colonies in our own solar system as soon as we can. Think of it as an insurance policy. If our luck should run out and a major catastrophe should strike Earth, resulting in an extinction level event, having self-sustained colonies of Earth's life off our planet would be a prudent insurance policy. So yes, there are several hundred million years left for us to become a fully space-faring species but there is likely significant less time for us to take out such an insurance policy.

Ensuring the survival of our human species as Earth becomes uninhabitable cannot be separated from ensuring the survival of the diversified life on Earth that we are part of. This most certainly will be the ultimate challenge for our species. We do not know what our chances for success will be in spreading Earth's life beyond our planet but try we must as in the long-term there is no other alternative. In the short term but still speaking in geological time scales, we will have many other challenges to address. For the past 12,000 years or so humans have enjoyed quite favorable conditions on Earth. There were of course brief cold spells but overall, our civilizations have evolved in mostly stable climatic environments. However, this stability is not likely to last. Earth's climate is bound to change at some time much more dramatically than what we may have to confront with the human caused climate change in the very near term. The forces that produced the ice ages are still at work and so are the forces of plate tectonics which eventually will give us a very different global map (see fig. 4.4). However, in addition to those natural forces, life on Earth is now subject to a new force; let us call it the anthropogenic force.

It is not the first time that life itself has become a force capable to change the evolution of life on Earth. Life has reshaped our planet in so many ways since it first evolved. For example, bacteria did that long before us. Without them, there would not be an oxygen atmosphere to support larger lifeforms such as us. More to the point, without them, we could not survive. However, with humans we witness for the first time that an animal species has become as powerful as we have. Including the power to destroy all life on Earth and us with it.

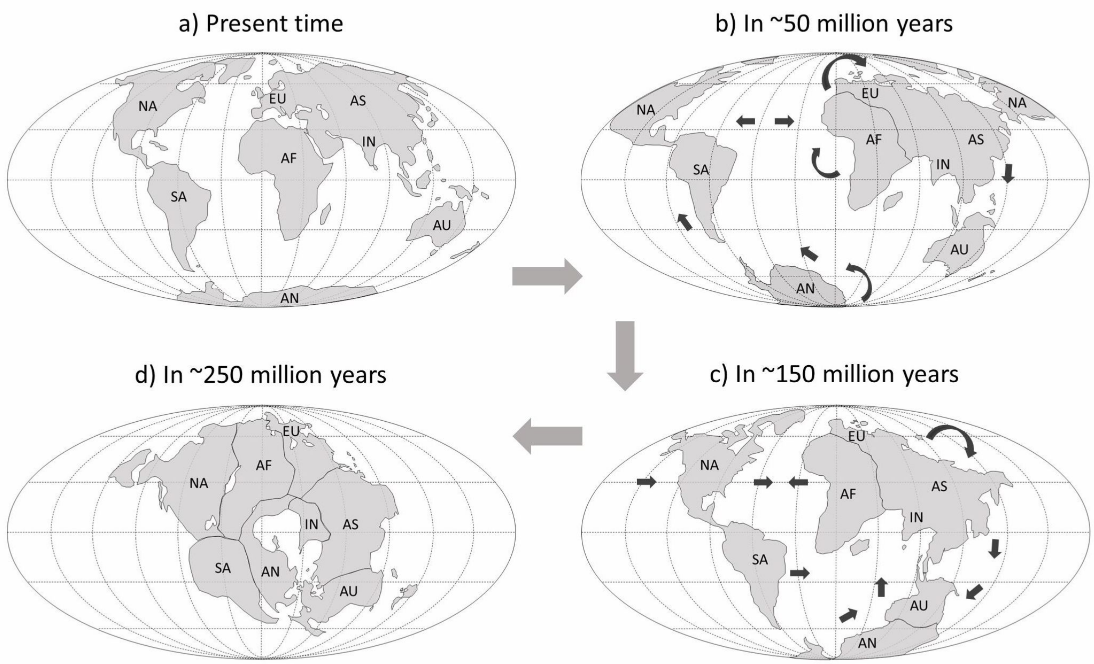


Figure 4.4: The current continental map (a) and expected changes due to plate tectonics illustrated for three different points it time: in 50 million years (b), 150 million years (c), and 250 million years (d) from present. The labels trace movements of continents / continent kernels: Antarctica (AN), Asia (AS), Australia (AU), Europe (EU), India (IN), North America (NA), and South America (SA). The maps sketch one of several potential scenarios of how the next supercontinent could form with the arrows in (b) and (c) indicating the major continental movements that could lead to a future supercontinent similar to what is shown in (d).

Dinosaurs may have ruled the world for some 150 million years, but they never had any power to destroy it or alter it for all other life. Unlike humans, they never developed the capability to drastically change the conditions for the evolution of future life on Earth. Our evolution and our civilizations have radically altered the conditions for most other life on Earth. A few hundred years ago, the human species impact on Earth's life was still quite limited. Granted, by that time we had already managed to kill off many a species, most of the very large animal were gone and their demise in many places eerily correlates with the arrival of humans. However, that was nothing compared to what was still to come. With industrialization, we became a truly global force.

In essence, we have become a terra-forming species, capable of altering conditions for life on Earth. The unfortunate fact is that we do so not purposefully for a greater good but driven by our insatiable hunger for economic growth that devours Earth's resources without any thoughts as to the eventual consequences. We brought global warming and climate change upon ourselves, driven by our own greed and disrespect for our home, planet Earth, and for the diversity of life to which it is home as well. We cannot change the past, but we can decide what our actions will be in the future. Climate change and global warming are facts that speak for themselves. Instead of litigating the past, we all must focus on how we will be addressing the challenges associated with climate change. Many of us are much concerned about the material impact, like how much money it will cost to build new dams, meet new building codes, or clean-up after increasingly frequent hurricanes and floods. But of much more concern is how we will deal with the larger societal changes, how we will for example balance the burdens of change across nations and populations, how we can ensure that political, economic, and national interests expressed by elite minorities do not outweigh the needs of the majority of the human species. The minority are of course the representatives of rich nations specifically those that actually stand to benefit from climate change more than they may be hurt. And the majority are those who are already living in difficult conditions today and who will have to struggle ever more to support themselves even at a subsistence level given the negative impact climate change will have on their ability to raise crops. In the face of climate change we must become more united and not be driven further apart.

The challenges that we must address to cope with climate change are only a precursor to the challenges humans will have to address long-term. Our planet Earth is an ever-changing world and always has been. As a species, we have only been around for a brief time on an evolutionary as well as on a geological time scale. Earth can be a very different place from what our species has experienced so far. At several points in time, we know it almost froze over completely, at other times it was much warmer than it is today. On a geological timescale, conditions for life have changed quite dramatically and life has adjusted; and we will have to adapt in the future as well. Fig. 4.4 shows how the continental map of our planet is likely to change over the next 250 million years. It illustrates the formation of the next supercontinent is one of several potential ways such a supercontinent could form and what it could look like.

Whatever the details of the formation of this next supercontinent will look like or if its eventual continental shape will be similar to the continental map in fig. 4.4d, one thing we know for sure, a new supercontinent will have formed by about 250 million years from now. Currently, North America and Europe are moving away from each other at a rate of about three centimeters per year as the Atlantic Ocean sea floor continues to spread along the Mid-Atlantic Ridge. At the same time, the Pacific Ocean will continue to shrink and the continental masses of the Americas, Africa, and Antarctica will shift further north. As this happens, Africa will rotate and so will Asia, which as it does will moves further south, as will Australia; the European landmass tied to Asia will follow Asia's rotation.

The rough directions of these continental movements are illustrated in fig. 4.4b and will likely play out over the next 50 to 100 million years. However, the expansion and contraction of the Atlantic and Pacific Oceans eventually will come to a halt and then revert to reunite in some 250 million years the landmasses of all continents in a new supercontinent, Pangea proxima, which we already encountered earlier. This reversal in continental movements is illustrated in fig. 4.4c with the Atlantic Ocean shrinking as Africa, Europe, and Asia continue their rotational movements with Asia's landmass moving ever closer to the equator. In this process, the long separated continents of Antarctica and Australia are expected to finally reunite with their joined landmasses moving towards the equator until eventually most of Earth landmass will be recombined in a single supercontinent, Pangea proxima, as shown in fig. 4.4d surrounded by one vast ocean echoing the Panthalassa Ocean of the Permian past (see fig. 3.11). Now, the specific results of this continental dance may look somewhat different from the one pictured in fig. 4.4. According to some scientists, there may actually be an earlier supercontinent formed. Therefore, fig. 4.4 should not be viewed as an exact prediction, which it is not, but as an approximate illustration as to how this new supercontinent may form over the next couple of hundred million years or so.

The main message of these scientific predictions is that Earth will be a very different place in some 250 million years. Just as they have been doing for more than 3,000 million years (see fig. 2.4), the unrelenting forces of plate tectonics will continue to move Earth continents around in a rather predictable fashion. Now why is it important to digest this? For the simple reason that we as a species with a rather short lifespan and a short experience on Earth seem to perceive our environments as quasi static. Our subconscious belief is that things will always be as they were even though we know quite well what havoc history can wreak. Not to speak of what nature could be able to do, next to which any human impact likely would be negligible; that is unless the human species plunges itself and everything else with it into the abyss of an all-out nuclear war. Now if the latter were to happen it would be an unfortunate event for humankind, but it will not be the end of the world, it just will be the end of the human species along with likely many other species. It will certainly set things back for life on Earth for quite a while, but eventually new life would emerge and likely flourish on a planet without humans. We think ourselves as way too important. The end of our species will not be the end of life on Earth. All it will do is clear the deck for evolution to start is magic all over again.

The ancient Romans used to date their years *"ab urbe condita"*, from the founding of their city of Rome in 753 BCE. Rome exists now for more than 2,700 years and many like to think of it as the eternal city. It is not. In a few million years, nothing will be left of it. And there most likely will be no more Berlin, Buenos Aires, Cairo, London, Moscow, Nairobi, New York, Paris, Peking, Rio de Janeiro, San Francisco, Seoul, Tokyo, or Vienna for that matter, or pick any other city you may like or live in; it likely will be gone. In the long term nothing that humans build on Earth will last, some things will be around for longer, some for a shorter time; but when looking out millions of years everything we know today will have vanished, everything will be different. It is we, who will have to adjust to a changing Earth. We as a species will have to learn to live with changes we cannot control, and the coming climate challenges will already give us a chance for doing so.

Looking back at Earth's history, it may actually be better to build maritime based societies and not terrestrial ones. Building large floating cities is currently still beyond our capabilities but once we can do that, it certainly would make it easier to adjust to Earth's longer-term changes. There are actually some encouraging signs as architects and engineers have already started to build floating homes. Similarly, aquaculture is still in its infancy but harvesting maritime fields may become much more important in the future. If indeed we have managed to set in motion a prolonged climate change that eventually will result in much higher sea levels, many of our coastal cities will be swallowed by the sea over the next several hundred years. We already know what the short-term impact of climate change is. However, that is nothing compared to the much larger changes that may be in store for us. Much higher sea levels would be nothing out of the ordinary for life on Earth to experience. Sea levels in the past have been hundreds of meters higher than today (see figs. 3.9 and 3.10). It is just that this time it may be our doing when this starts happening much sooner as it otherwise would have been the case. One of the things often missed in the climate debate is that we will not be able to prevent the climate change that is already dialed in. We need to start talking more about mediation of what we already know will happen with high certainty.

Even with our best efforts we cannot turn back the clock on climate change, all we can do is to stop making things worse for future generations. However, regardless of the effects of climate change, we can rest assured that Earth at some point in its future will again be a planet of much higher sea levels followed by much lower ones, a cycle that plays out over many millions of years. So eventually, we would have to deal with such changes anyway, that is if we were to be as long lasting a species as for example the trilobites or the dinosaurs were. As it is, we have to deal with the consequences of such changes earlier, and even if they are still small compared to what Earth eventually will have in store for us, it is a learning opportunity. We should not plan for a few hundred years; we must plan for thousands of years or otherwise we continuously will have to rebuild the infrastructure of our civilizations. On Earth that will mean making the oceans our home much more than that is the case today but in ways, which respect the life in those oceans. A second avenue is of course space, whatever we can relocate into space will not be subject to any earthquakes or floods, maybe we should build our data centers there as it also would provide for very efficient cooling; and of course, heating for anyone working up there.

Earth will continue to be our home for a long time to come. Over time, we will have to learn to live on Earth in harmony with the forces that shape Earth. We cannot succeed in working against them. Earth will always be our home base as long as it remains habitable. However, we need to start turning it into a real springboard for spreading the life of Earth into the vastness of space, settling new planets or moons. They may not be habitable now but maybe we can make them habitable for human colonies if we are successful in developing the technologies to do so. Space is sometimes referred to as the *Final Human Frontier*. Indeed, this is the case, but it is only a final frontier and not the only final frontier that we will have to explore as we continue to evolve as a species. We humans are to ourselves as much a riddle and unknown as the cosmos is still to our scientists. Exploring and understanding what it really means to be human, what it is that makes us do whatever it is we do or neglect to do, what makes us think about ourselves and how we think about others are no less important challenges than the final frontier of space. For us to evolve as human societies we need to understand ourselves at levels we cannot today. Without such understanding, I fear we will not be able to build the societies that can evolve into a human species civilization ready to face the challenges to come and to secure a future for Earth's life among the stars in the vastness of our cosmos.

The Future of Evolution

Is evolution still shaping the human species? Or has human civilization put the various mechanisms on hold through which the evolution of life previously worked its magic? Have we come to a point in our technological evolution where we can actually start to shape our very own biological evolution? Maybe not just our evolution, maybe even the evolution of life itself? Are we getting close to wielding what our distant ancestors would certainly have seen as divine powers? Asking any of these questions would have seemed preposterous only a few generations ago. We have come a long way since then. It is almost terrifying how dominant the human species has become on planet Earth over only a few thousand years and how powerful we seemingly are; but it is also deceiving. In a sense, humankind is still very much like a toddler, unfortunately handed a chemical or electrical kit to experiment with before she or he has even grown up to read danger signs. In such a situation one could only hope that the toddler will not be able to unwrap her or his presents any time soon. However, in the case of humankind, this is just what has happened.

We can expect the future of evolution to be quite different from its past. The evolution of the human species and its culture evolution have changed the boundary conditions for evolution on Earth itself. The human species is not only subject to the evolutionary forces shaping life on Earth, it is also by itself now one of the most powerful forces driving evolutionary outcomes. For most of other life as much as for itself. Why most and not all? The human factor will likely continue to shape the evolution of higher life on Earth as it has already done so for thousands of years, maybe even much more so than in the past. However, it is unlikely for the human species to have the same impact on the evolution of the most numerous lifeforms of all, bacteria and archaea. For the human species, one aspect of evolution has become much more important than for any other species, and that is cultural and societal evolution. It cannot be overestimated how much our own societies are shaping us as human beings not just in our individual development but also as a species. Just as life in nature adapts to fill specific ecological niches, we humans adapt to life in our societies while at the same time we also shape them. It is quite a unique situation. In a sense, we are bootstrapping ourselves in our evolution as the proverbial social animal. Uniquely for the human species, in addition to the "traditional" evolutionary forces shaping life on Earth in the past, we have become increasingly subject to powerful societal forces that we can expect to be no less consequential in shaping the evolution of the human species in the long-term. We will look first at "traditional" evolutionary forces.

There has been some debate if the emergence of civilization has so-to-say stopped human evolution in its tracks. For a time, some renown scientists even asserted that there was practically no evolution at all over the past 50,000 years and that we are still essentially identical to the same modern humans that moved into Eurasia around that time. Of course, we may easily identify with our more recent heroic ancestors as they speak to us for example through the epics of Homer, in the epic of Gilgamesh, or in the Mahabharata. But are we really still the same kind of people? In a few thousand years from today, will the human species be practically indistinguishable from us? How about one hundred thousand years, should we really be stuck in an evolutionary sense with who we are today because of civilizations? Quite to the contrary. There are indications that the human species instead of lumbering in the equivalent of an evolutionary stasis has actually evolved much more rapidly in recent times. Civilization seems to have accelerated evolution, not stopped it. If this continues, we can expect our species in the not-too-distant future to be very different from who we are today, maybe already in a few thousand years but certainly in a hundred thousand years. This should be good news because we will need to adapt to ever faster changing environments as our civilizations evolve. But let us get back to "traditional" evolution itself.

We all have heard that evolution is about the survival of the fittest. While evolution itself works on populations, selection or survival of the fittest is brutally individual. Predators cull the weakest animals in a herd and not the strongest. Most mutations have little to no effect on their hosts and the ones that do, fall into two groups, the few that confer an added strength and those that result in less fit animals. Again, it is the less fit animals that do not make it. Only the fittest will come out on top and will be able to have offspring; this is how evolution works as populations are constantly tuned through only their fittest members being able to add to its stock. As the argument goes, humans have sidestepped this process as our societies allow weaker members of society to survive and to have children. In the early decades of the last century, there was much concern about this, articulated in the language of the time in pronouncements sounding like this: *"Without the weak being weeded out naturally, the human gene pool would degrade"*. We all know how this ended. It started with seemingly innocuous eugenics movements, which turned into terrible tools that even in solidly democratic societies and in their purported interests violated those not deemed worthy to procreate, if not killing them outright. It culminated in the unspeakable horrors of genocide, mostly of Jews but also of many other ethnic and religious groups as well, committed in the deprived hope to purify a mythical Germanic race that never existed to begin with. Unfortunately, it did not end there. We can witness well into our times the savagery with which racial and ethnic biases can turn into mass murders and genocides. It is frightening how easy it is to set one group of people against another by exploiting racial and ethnic biases that have no justification whatsoever in science but unabashedly play on our innate fears of the supposedly "others". It is tribalism redux in its most destructive form. The Nazis were not just morally and ethically corrupted to the bone, they were factually wrong too. Aside from being murderous thugs, they were wrong with regard to race in general and they were wrong regarding their own superiority. Nazi racism had never a scientific base but was a cult and a primitive one at that.

Anyone who bothers to find out more about their very own genetic heritage is likely in for a surprise. The peoples we for example find in Europe today, and that was no less true when the Nazi's came to power, were not derived from some mystical pure genetic stock but were a melange of people that either peacefully or through warfare intermixed. With the help of DNA analysis science has been able to trace some of the major migrations. For example, the genetic makeup of the population mixture we can find today in Europe but also in Southeast Asia was to a good extent shaped by two major migrations in the last 10,000 years. The first one of these two migrations was a consequence of the development of agriculture, when peoples from the Near East first spread into what today are modern Turkey and Iran. These Anatolian and Iranian farmers, to name them after

the geographies from where their migrations started, then spread around 9,000 years ago into Europe and Southeast Asia respectively. There, they largely melted into or absorbed the hunter-gatherer populations who had moved into these areas much earlier. One such encounter seems to have led to the development of what would eventually become the Yamnaya culture which remained a pastoralist culture in Southern Russia, roughly between the rivers Bug and Dniester in the west and the Ural river in the east, a geography frequently referred to as the Pontic-Caspian Steppe. The story of the Yamnaya people is part of many high school curriculums covering Bronze Age cultures and most students will remember them as the builders of the burial sites called Kurgans. Nowadays, we also know them as the ones who likely were the carriers of the corded war culture as we see its front moving westwards as some of the Yamnaya migrated towards Europe in the west while others carried their culture to the southeast into India [76].
In this second migration, these Bronze Age Yamnaya horsemen with their horse drawn wagons swept into Europe and India roughly 5,000 years ago. They not only left their genetic imprint on the populations living there, but they are also seen as the likely carriers of the ancestral Indo-European language from which most of today's languages in Europe and in India derive. The only surviving none-Indo-European language in Europe from before the Yamnaya migration is the language of the Basque people.[19] We can almost certainly assume that there were many smaller migrations. We just have to look at the migrations that we know from the brief period of our recorded history such as the ones that ripped the internally weakened Roman Empire apart. This continued coming and going and intermixing of peoples through pre-history and history clearly tells us that any claims of racial purity are not compatible with what we know about the history of peoples over the past 10,000 years or so. It is positively ironic that usually many of the examples which some point to for evidence of racial purity are more likely to be examples of prolonged inbreeding and as such represent a genetically weakened population, not a stronger one. There are no inferior or superior races. However, there are still very significant differences in the access people have around the globe to education. Not just in nations that struggle to secure the financial means needed to provide for an inclusive educational system but also within wealthy countries such as the United States. Nurture is often far more discriminating than nature; or so it seems. However, acknowledging and understanding those disparities does often little to change deeply ingrained preconceptions. There is nothing new here. It is instructive to quote Herodotus, the father of history, in this context. In his *Histories*, he gives an account of how the Persians perceived the world beyond their borders:

> *"As regards nations, the Persians rank their own immediate neighbors as the fittest to be graced with their respect – always after themselves of course; then those who lie beyond their immediate neighbors, then those beyond them in turn, and so on, in ever-decreasing order of proximity. Accordingly, it is the peoples who live the farthest away from them who are the most despised. That this is so reflects the presumption that they are the greatest people on the face of earth, and that the quality of other people diminishes the farther one travels from Persia, until in the end, on the very margin of things, there is nothing but savagery."*

Please substitute your favored nation for Persia. No, seriously, we are not just like that anymore, are we not? It is amazing how little most of us still know about how people somewhere else on our planet live and what their cultures really look like and in how they are similar or different from our own. Different from the Persians in antiquity, our ignorance in this manner is not just an indication for how far away from us these other people live; often enough, we have not even a clue as to how much closer to home, some of our fellow citizens live. How different would the world look if each and every politician before being allowed to run for office would be required to spend at least one year of her or his life in the midst of the most unfortunate in our societies? We fear the things we do not know, and we cannot understand what we fear. Those who are not like us or about whom we know very little have to be inferior, right? Judging by our very own conduct as a human species as to how we treat those among us who are different and how we treat other species, this sentiment indeed seems to be pretty much the norm for human behavior. The ancient Persians in their time were not unique, just look at the Greeks of antiquity who looked down on everything non-Greek in similar ways. Are we different today?

Let us return to the question if evolution is still working on the human species. Or put differently, when did human evolution supposedly stop? It is a fact that the human species has a surprisingly low genetic diversity. If we look to our primate relatives or to other species, the genetic diversity of the human population worldwide is lower than what one should expect. If we only look at the genetic diversity of human populations outside of Africa, it gets even worse. The genetic diversity of African populations is quite a bit higher than what we find among the populations that derived from our African ancestors migrating into Eurasia. Much of the genetic landscape of human populations we find today is thought to be due to the small number of ancient humans comprising the most "recent out of Africa" migration between 70,000 to 50,000 years ago from which most modern human populations outside of Africa derive. The low genetic diversity of modern humans outside of Africa reflects a bottleneck effect.

As discussed earlier, genetic drift can quickly spread a mutation within a small population such that one hundred percent of the population carry the mutation within only a few generations. Scientists believe that this was the case when modern humans left Africa as every time small human populations set out to migrate towards a new territory genetic drift gets a chance to work. In Africa itself this must have also been the case for a long time before humans ever left the continent as populations spread all over Africa. Where population densities are low and genetic drift is high we also know that gene surfing [37] can take place leading to an amplification of gene variation along the frontier of a human population expansion. Because of this, the highest mutation frequencies are observed right at the frontiers of human migration waves as they expand in space and time. The tools of modern genetics allow us to trace back the trajectories of specific genes surfing such migratory waves and thereby allow us to retrace the migratory patterns of our human ancestors. It is the effects of both, genetic drift and gene surfing, working on small groups of humans migrating out of Africa which explain the lower genetic diversity of human populations outside of our continent of origin.

The human expansion first within Africa and then out of Africa has generated the genetic diversity we find in today's modern populations, a lower genetic diversity outside of

Africa and a higher genetic diversity within Africa. The small size of human populations migrating out of Africa and the resulting low genetic diversity may have shaped human evolution in more than one way. Without genetic influx from outside, a small human population cannot survive, particularly so when it starts out with low genetic variability. It is not clear where the critical boundary lies, defined by population size and genetic variability, below which the survival of a human population becomes impossible. If we assume for a moment that modern humans did not have an undue influence in the demise of the Neanderthals, maybe the Neanderthals themselves may provide a good illustration if we knew their population numbers and genetic diversity across which they interacted and presumably could mix. Maybe we will have that answer one day. Modern humans have not run into this problem. While genetic diversity outside of Africa is still lower than it is in Africa itself, we can expect genetic mutations to lead over time to a much higher genetic divergence among human populations than we can observe today. Certainly, that has been the case in the past as we today can look at a set of human populations that in some measures have clearly diverged from each other, even if minimally so.

For a long time, practically for some 30,000 to 40,000 years, this trend has been going on, only checked by migrations as discussed above. However, the distance over which most of us today can interact has very much increased as compared to only a few hundred years ago. We can and do intermix over much larger distances today. Not surprisingly, this will eventually lead to more homogeneous and well-mixed genetic pools within those larger sets of human populations that we can recognize today. Increasingly, we also see more intermixing between these sets of populations, much more so in some places than in others. This is all good news, as it should practically ensure that the genetic distance between any given pair of humans does not grow too large while at the same time, we all become a little more genetically diverse.

All of us inherited our genes from our parents and except for the so-called sex-chromosomes, all chromosomes carry a copy of each gene from a mother and a father. For females, this is also true for the sex-chromosomes but not so for males where the Y-chromosome inherited from the father and the X-chromosome inherited from the mother come with only one copy of the respective genes residing on the chromosomes. The two gene copies of the same gene inherited from a father and a mother can be and usually are different variants of a specific gene with a given functions. If they are different variants, they are called alleles. Which of the alleles is expressed in the phenotype depends on which if any of the alleles is dominant. Like the alleles for dark eyes and dark hair color which are dominant over the alleles that code for blue or light eye colors or for red or blond hair. Many alleles expressed in the phenotype we cannot readily see as we can with hair or eye color, but they certainly serve critical functions, like determining the blood types we carry, just to pick one example. Some alleles can convey a resistance to certain diseases if we carry only one copy of them, but if we inherit the same such copy from both parents they do not convey an advantage anymore but instead can potentially make us much sicker. Sickle cell anemia is a very good example of that. This disease is caused by a mutated version of the gene which makes hemoglobin, the red protein responsible for transporting oxygen in our blood. Blood cells carrying the mutant hemoglobin are not doughnut shaped but distorted into a sickle cell shape instead. A person inheriting such a mutant gene from each of her or his parents, ending up with two copies of the mutant gene, also called a homozygote, will suffer from sickle cell disease. A person that inherits a healthy gene for hemoglobin production from one parent and the sickle cell allele from the other parent, also called a heterozygote, will not develop sickle cell disease but will be protected from malaria. Nature runs a very fine balance here. A homozygote carrying two healthy genes when infected with malaria will develop the disease with all its negative consequences. A homozygote carrying two sickle cell alleles of the gene will suffer from sickle cell disease, which even with today's modern medicine makes for a difficult life but in pre-modern civilizations often would have equated to dying young, assuming one made it into adulthood at all. Finally, there are the heterozygotes, carrying one copy of the healthy gene and one copy of the sickle cell allele, which allows them to live a normal life albeit protected from malaria. Because of this benefit for heterozygote carriers, nature is willing to perpetuate the existence of a gene allele that in a homozygote would have been pretty much a delayed death sentence in pre-modern times. This is sometimes referred to as the heterozygote advantage, as it confers a higher fitness. Higher as compared to the homozygote with the two bad genes, here the sickle cell alleles, also called the recessive homozygote; or the homozygote with the two good genes, also called the dominant homozygote. Recessive and dominant refer to the gene allele being expressed in the heterozygote which is here the good gene being the dominant gene and the bad gene, the sickle cell allele, the recessive gene, which while not being fully expressed still conveys malaria protection.

What would happen if we found a cure for malaria? Would the sickle cell allele in the populations carrying it today be lost over time? That could well be but it would happen over an extended period of time as with modern medicine the sickle cell homozygotes, while still having a somewhat lower life expectancy, would be able to pass on their genes. When did this sickle cell mutation first arise? Was it millions, hundreds of thousands, or thousands of years ago? It is only now with modern genetics that we can pinpoint the time when this gene mutation first appeared. More than that, we can also determine if there is a single origin for this sickle cell allele or if there are several independent ones. The answer in this case is that the sickle cell allele appeared first about 7,300 years ago and it can be traced back to a single origin. The sickle cell mutation occurred only once, approximately 7,300 years ago, either in the Sahara or in west-central Africa [77]. 7,300 years is not a long time ago and it is most likely not a coincidence that the mutation appeared during a wet phase in the Holocene, a time when the Sahara was greening, also referred to as the African humid period. Mosquitos, the carriers of the single-celled malaria parasite, must have found many breeding grounds and a single mutation of the good hemoglobin making gene into the sickle allele would have conferred a significant advantage. It would have taken some time for the sickle cell allele to spread through a population before there was an increased chance of inheriting a sickle cell allele from both parents; and with that, the negative effect of the mutation, sickle cell disease, balancing the spread of the sickle cell allele. Interestingly, in other parts of the world different alleles than the one underlying the sickle cell mutation have evolved to counteract malaria. If there are evolutionary challenges, nature is seemingly able to adapt in more than one way and more quickly than we have given her credit.

Let us return to genetic drift for a moment to explore another aspect. As discussed earlier, small human populations are much more likely to quickly spread or lose a gene mutation. If an allele spreads across most of the population it becomes the dominant expression of a certain gene; the alternative is that the allele is completely lost. Genetic drift is random and can lead to significant genetic changes in a small population over only a few generations. While a single human can only carry two alleles of each gene at most, across a larger human population there will be many different alleles for any given gene. If these alleles confer no specific advantage or disadvantages in homozygotes or heterozygotes it will be up to pure chance how alleles propagate throughout a population and that in turn depends much on population size. Assume a specific allele occurs in a small population with a frequency of 50%, half of the population has it the other half does not. If population sizes are really small as they would have been in hunter-gatherer days, such an allele could have been, by pure chance, either be fixated, 100% of the population carries it, or completely lost, 0% of the population carries it; all within as few as ten to twenty generations. The larger the population is, the more generations it takes to lose or fixate such an allele and if the population is large enough, gene drift will only increase or decrease the probability of anyone carrying the gene mutations slightly, a few percent higher or lower than the 50% probability the group started out with. Genetic drift works efficiently in small populations to spread quasi-neutral mutations but can also equally quickly lose them. In large populations, a neutral mutation can hang around for quite some time without being fixated or lost.

Where it gets interesting is once an allele can confer an evolutionary benefit. Even if that benefit is small, the chances of it becoming fixated instead of being lost very much increase, in small populations as well as in large ones. There is another important aspect of large population sizes. The speed at which evolution works is largely driven by the frequency of beneficial mutations occurring. The mutation rate, the frequency at which gene mutations occur does depend on where within the genome a certain gene is located but overall, the mutation rate is low. Until recently, it looked like as if a large fraction of human DNA did not hold genes serving any known function at all; scientists thought that less than 10% of human genes served vital functions. While we know now that this is not the case, back then this indicated that many mutations would take place in areas of the genome that seemingly did not matter, and of the remaining ones, many would not result in functional changes either, even if they occurred in the relevant 10% of human DNA base pairs. Today we know that much more than 10% of our genes serve vital functions but still, beneficial mutations remain pure chance events and may happen only rarely, maybe too rarely to allow a species with a small population size to adapt successfully to changing environments.

This is where the controversial concept of adaptive mutation comes in. Different from chance mutations which are blind, the proponents of adaptive mutations believe that mutations are not random but happen in response to stress. While this may sound similar to the concept of epigentics which we discussed earlier, it is not. Adaptive mutations are thought to be real gene mutations occurring in reaction to stress and not functional inhibitions on the genome caused by molecules blocking critical sites on the DNA that disable the function of particular genes, which in the extreme case may also lead to epigentic markers being passed on to following generations. Such adaptive mutations may have been shown to be at work in bacteria but how they may play into human population dynamics and how they could have shaped our more recent evolution over the past few thousand years is still very controversial and much debated. Regardless, adaptive mutation is an intellectually intriguing concept even if the way adaptive mutation is thought to work is not correct. This is because a concept such as adaptive mutation could align with many observations if we just take a step back to consider it.

There is a perceptual difficulty here as we are so used to think of evolution only working over very long time scales, hundreds of thousands or even millions of years. Many will have trouble acknowledging that human domestication of plants and animals may be anything else but selective breeding. They are of course right, it is selective breeding. However, it clearly has resulted in dramatic changes that, just to take one example, have modified the wolf genome in many ways to produce all the countless dog-breeds that we know today; and this process is not at its end. The first dogs were already domesticated by our hunter-gatherer ancestors. Somewhat later humans started domesticating many more animals and plants as well, all of them undergoing their own human induced evolution. As pointed out earlier, the human species, in parallel to domesticating many plants and animals, may have been subject to its very own domestication without even realizing it, that is until now. If we just look at the past ~12,000 years, the Holocene epoch, is not what we see a pressure cooker environment for accelerating human evolution?

Arguably, never before has the human species been under greater pressure, much of it admittedly of its own doing, than since the beginning of the *Neolithic Revolution* which ushered in the transition from humans living as hunter-gatherers to becoming settled agriculturalists. In many ways, the stress factors we have subjected ourselves to during that period have been multiplying. But somehow, we have managed to keep up, so far at least. As pointed out earlier, the introduction of agriculture and the domestication of our farm animals as we know them today did come with its downside, biologically and socially. The diet of early peasant families was less healthy than that of hunter-gatherers. Closely living with livestock exposed humans to new diseases, and the societies they lived in also became less healthy, specifically for those parts of the population which ended up at the bottom of society in the process of social fractionation. If that is not stress, what is? This stress would not have been a temporary or short one, as it would have exerted its relentless pressure over many generations. However, there were also benefits down the road. Once cows were domesticated, several human populations evolved the ability to digest milk as adults. This happened independently and through separate mutations in Europe, the Middle East, South Asia, and West Africa [78]. Why should this be anything else but something like adaptive mutations taking advantage of a newly available protein source such as milk? Looking further back into the time window of 50,000 years, where supposedly we did not change in an evolutionary sense, there are other examples of recent mutations.

When the ancestors of human populations outside Africa left their continent of origin and moved into Europe and Asia they had to adapt; just as other human populations migrating out of Africa had done so before. A prime example of course being the ancestors of Neanderthals adjusting to local environments as they migrated into Eurasia. One of the genetic changes of the evolving Neanderthal population in colder climates was associated with the evolution of light-colored skin. Dark skin as a protection against skin cancer was less needed in cold climates and light skin could make better use of the reduced sunlight in northern latitudes to produce sufficient amounts of vitamin D. The light skin of Neanderthals may have been also the reason that they choose a dark color for body paintings instead of the red ochre used by their African ancestors. When our modern human ancestors started to move into Europe, they were the dark colored people and it was the Neanderthals, who were the fair colored people. As we know, in this encounter the dark colored people, our ancestors, persevered. For some time it was thought modern humans migrating into Europe may have acquired lighter skin color by interbreeding with Neanderthals, which we know they did. However, with the decoding of the Neanderthal genome we now know that this was not the case and that modern humans in Europe evolved their own mutation that gave them the advantage that lighter skin color has in northern latitudes. This seems to have been more of a recent development, as it looks like the gene modifications in most modern Europeans responsible for producing light skin all practically happened less than 10,000 years ago. These modifications are attributed to adaptations of farming populations from the Near East moving into Europe as well as the migration of the Yamnaya people into Europe. Seemingly, some 10,000 to 8,000 years ago Europe was still pretty much a patchwork of skin colors. Something similar happened among populations that had migrated towards East Asia instead of Europe, but their skin color changes are due to different mutations than the ones producing light skin color in modern humans in Europe. Therefore, we are looking at different examples of skin color changes, each one achieved by different mutations. It seems as if the genetic response of human populations a few thousand years after moving to latitudes where the new stress factors include a colder climate and less sunshine invariably is a rather predicable one.

While we inherit one copy of genes from each of our parents except for the genes on the Y-chromosome, we can also each carry multiple versions of any given gene. Of which we get also one copy each from our parents. One example for such a set of multiple gene copies is the gene that allows us to digest starch. The interesting fact is that it looks like the number of multiple genes of the starch digesting variety that can be found in a given population does correlate with how fond that population is of carbohydrates; the more carbohydrates in a population's diet the greater the likelihood that the number of starch digesting gene copies is higher. Our original starch digesting capability certainly goes back many millions of years, but this spread in the number count of starch digesting genes must be related to the introduction of agriculture and the increased starch intake associated with it.

Survival of the fittest in the strict Darwinian sense ensures that only the strongest have offspring. It has been the driving force behind the evolution of life on Earth as well as for human evolution. As humans progressed in their culture evolution other factors eventually became at least as important or maybe even more so when one looks at the evolution of human societies. In many ways, the Neanderthals were stronger than the modern humans moving into Europe but it is modern humans who remain the only human species on Earth today. We have seen in the previous chapter that one of the reasons that the Neanderthals are not around anymore is that their population size always was rather low and often at precariously low levels. Therefore, they were much more susceptible to becoming extinct than modern humans were; by natural forces or by competition from modern humans. The *Cognitive Revolution* of modern humans we discussed earlier did not really increase the strength of any individual. Actually, diverting time to produce art or engage in so-to-speak broadly cultural activities does not strengthen the one practicing the art in becoming a better hunter or warrior. Rather, it actually is likely to reduce hunting and fighting skills because of less practice. There must have been a compensating gain and the most likely candidate for that is a much stronger social cohesion and identity. The *Cognitive Revolution* must have strengthened the social bodies modern humans were living in and through that likely made them more competitive, not individually but as a society. The strength of a modern society on more than one level is not so much determined by its hard skills such as the fighting prowess of its soldiers, but by its soft skills, its technological and societal innovations. Einstein never would have made a good soldier but any society that would have had the privilege of hosting him would have benefited enormously. With the accelerated evolution of human civilizations, beginning with the industrial revolution, it was increasingly the brains and not the brawns that decided a nation's fortune. Since the industrial revolution, it were mostly the science, mathematics, and engineering parts of our brains that conveyed these advantages. However, with the resulting technology forces tearing at the very fabrics of our societies it became increasingly clear that human society itself, the social contracts within and among nations, would also have to change from the ground up. The soft skills of social science and social engineering, for the lack of a better term, have become as critical to the long-term survival of our civilizations as the other soft skills that produced all the technology gadgets surrounding us today.

The development of these two aspects of human soft skills, we can refer to them as human technology evolution and human social evolution, move unfortunately at different speeds. A good example for that is the science and technology of biology. Our own capability to manipulate the very code of life may soon be giving us the tools to engineer human biological evolution. However, in no respect are our societies prepared to deal with this in any responsible way. The most responsible approach for now is likely a highly regulated one that will be extremely restrictive as to what can and cannot be done with genetic engineering. This will be difficult because in essence, the cat is out of the bag. Given human nature we can almost assume with certainty that money will be spent on "improving" children; why not select their hair or eye color, why not make sure that they will be resistant to many diseases that still plague mankind today; why not make her or him smarter and stronger at the same time. However, we are learning that understanding the human genome and how it works is much more complex than initially thought. In some cases, so-called switch genes can be identified which if activated or deactivated will turn on or turn off certain biological functions. But more often than not, our biological functions are regulated by many genes working together in intricate ways. This is where the genes come in that are part of the 90% of human DNA that scientists until recently considered junk DNA, seemingly not serving any vital function. Not so anymore. As it turns out, genes in the supposed junk part of our DNA do serve functions after all, as it looks like mostly in controlling the activity of the 10% of genes deemed vital before. Because of that, understanding how genes actually achieve what they do has become much more complex. After all, a Frankenstein world is not just around the corner yet.

We must use the time science needs to figure out how nature achieves its magic and before technology evolves that can reproduce it, to give ourselves rules that safeguard the use of genetic engineering. Without such safeguards, human genetic engineering will virtually open Pandora's box, unless our societies find a way to prevent the evolution of super humans.[20] There will be many challenges where genetic engineering can help us and in some cases already does, hereditary diseases are but one example. We all wish for this to become part of our medical tool kit allowing genetics to help people that can get no such help today. However, most of us likely also agree that genetic engineering can never become a tool to generate more or less worthy humans resulting in what would be nothing else than a genetic version of a caste system.

In the first decade of this century, scientists succeed in decoding our human DNA. At the time, this was an enormous and very expensive scientific undertaking. Progress in technology over just a little more than one decade then made it possible for individuals to screen their complete DNA at an affordable cost. Mapping human DNA patterns against various human traits such as biological, medical, psychological, or social ones, is now becoming routine. As indicated above, how genes work together to achieve what they do is much more complicated and involves many more genes than initially thought. But this is where technology comes to the rescue. Combining a technique called genome-wide association (GWA) studies and the computational power to handle big data allows scientists to search the human genome for small variations, for many millions of such small variations by now. These many small variations and their specific combinations are then compared against specific human traits. The weightings of millions of such genome variations can for example be fitted to test for correlation with specific human diseases. The result of analyzing a very large number of human genomes in this way will then be a distribution of genome variations, the kind of bell-shaped curve all of us have likely seen at some point in time. On one end of this distribution, there is a high likelihood for someone being afflicted by a certain disease in his or her lifetime while on the other end the probability for such an event is low. In many ways, this will revolutionize the practice of medicine.

Imagine such a curve built from very many data points, so the statistics is solid, was available for example to screen for susceptibility to Alzheimer's disease. If you then had your genome tested for the very variations that had been used to build the curve in the first place, then your doctor could give you the likelihood for you actually having the disease at some point in your life. The important fact is that this test can be done decades ahead, well before the potential on-set of Alzheimer's disease because your genome at the time of the test at a young age is the same that it will be when you turn say eighty or ninety. Instead of diagnosing if someone has the disease or not in their old age, we will be able to predict if someone could get it or not with a certain probability when they are much younger. Some may want to know this while others may not. Insurance companies would certainly like to get their hands on such information, but their customers would likely not want to hand them this information. We use here the example of a disease to illustrate something that has a much broader impact. Just as one can predict the potential susceptibility of people with certain genome variations to contract this or that disease, the same technology allows to predict many other things about every one of us, and most of it, could be done right at birth. This includes for example simple things such as your height and weight as an adult, innocuous information for most of us. But what about intelligence or potential for academic achievement? As it turns out, our genomes are also very good predictors for all of those traits.

Psychology was among the first to discover GWA studies and it looks like they may become a much better tool to diagnose disorders than the traditional tools psychology has used until now. You see where this is going. How far should we take this? Should we have mandatory screening, maybe as early as at birth, to determine the likelihood that for example someone will commit a serious crime at some point in his or her life? If so, what then should we do with this information? We may well be on our way to a brave new world here. The reality is, and there is no running away from this, almost everything that scientists are learning from GWA studies confirms that nature seems to be much more in control of who we will be than nurture. The scientific view on heredity is changing rapidly but the broader public is little aware of it. Much good can come from GWA studies but we also can count on some parts of society trying to use such data and access to our human genomes for purposes that are more sinister.

The best it seems we can do is to have an open debate about how human genome data can and cannot be used. Unfortunately, too little of that is happening. Although there are efforts to educate the broader public, the complexity of this topic does not lend itself easily to the kind of simplistic discussions many of our media and public discourses nowadays prefer to engage in. Another question is how different societies will handle this. Democracies allow for an open debate, but other regimes may choose not to do that and use such genetic technologies for their very own purposes. Just imagine the kind of abuse such tools could serve in dictatorial regimes or veritable police states of which there are still many. In some countries, Aldous Huxley's *Brave New World* could become a reality very quickly. How can we achieve an international consensus that regulates what human genome information can be used for and how it can be obtained in the first place, and of course, who owns it?

The dramatic changes to our understanding of the human genome in just over the past three decades are quite likely something that the early pioneers of genetics may have dreamed off. But equally likely, they would have seriously doubted that we could get there so quickly. As many of us have experienced in one way or another, the revolution in technology over the last several decades has presented us with ever-increasing capabilities; all of which are becoming ever-more accessible to an ever-increasing number of us. Because of that, we can expect modern-day biology and genetics to have far-reaching consequences for our societies. If we had the same genetic knowledge we have today already one generation ago we could have done little with it because we did not have the technology. That has changed and today there is a wide range of practical things we could possibly do with this knowledge, thanks to our ever increasing technological capabilities. As a biological species we are coming close to the juncture where we will be able to engineer our own evolution. That certainly raises a number of important question that our societies will have to answer. Just because we can do something does not mean it would we a good thing to do in the first place.

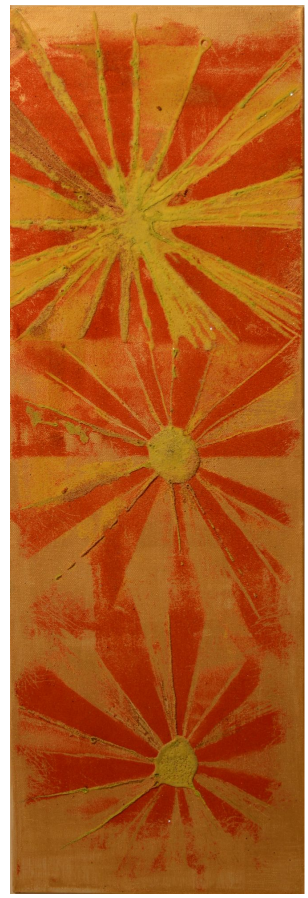
In the same way this applies to any human engineered biological evolution, of us or of other life, it also applies to the evolution of our civilizations. Given the impact our advanced industrialized societies now have on all of Earth's life as well as the future for any life on Earth, we have arrived at a critical juncture there too. In the larger context of all life on this planet, we as a species will have to make difficult choices as to how we would like our future relationship with nature to evolve. The questions all of this raises and the kind of discussions these topics require go well beyond the scope of this book. However, understanding how we got to this point, which is the story of this book, will be critical to approach such discussions; and hopefully help us make our decisions in any such matters with the humility they require. We are not the masters of this planet and we are not masters of our evolution. We are much more like the proverbial sorcerer's apprentice, always too convinced that we know it all and only realizing how wrong we were once we find ourselves in an inextricable mess of our own making.

What remains as we look out into the far future is only one more question that we need to ask. What can we say about the potential future of human evolution in the very long-term? For that, it may be instructive to look back in time to when fish became amphibians moving onto land. For animal life to leave the sea and to live on land, it had to adapt to life on land including the evolution of respiration systems such as lungs or tracheae and legs. Because land and sea are intimately connected this evolution played out over many million years in a gradual process until at some time, maybe after many failed attempts, one of our ancestors, a transitional animal such as Tiktaalik was equipped to be at home in the shallow seas as well as on land. Once amphibians had evolved, reptilians and mammals would follow. Now, crossing the frontier from sea to land is very different from crossing the frontier from Earth to space. Most of us would likely agree that it is the technological evolution of humankind which enables us to leave Earth and not our biological evolution. After all, there is no way to adjust to life in space biologically unless one has technology to support life in outer space in the first place. We already imitated the evolution of life in a way when we used our technological evolution to leave the ground and conquer the skies above us. Life took to the air several hundred million years ago and our species achieved this in the early twentieth century.

Clearly, conditions for life in outer space are different from those on Earth and humans for now can stay in outer space only for a limited time. Much of that has to do with the lack of gravity to which our bodies inadvertently adjust in outer space, limiting the time we can stay in zero gravity environments. So we will have to learn to build spaceships with artificial gravity. That may help us to travel more freely within our solar system and maybe to build colonies on some planets or moons. However, to start colonies on a moon or planet we again would have to adjust to different gravity. Gravity on Mars for example, because of its smaller mass, is only a little less than 40% of Earth's gravity. Therefore, someone weighing in at say seventy kilograms on an Earth scale would only weigh a little less than twenty-eight kilogram on a Mars scale. If humans would stay for many generations on Mars, we could expect them to become physically and physiologically notably different from humans on Earth. Our first inclination may be to think they will become taller but that may not be the case if supporting a smaller body in a lower gravity environment is energetically more favorable. Similar if humans would colonize a planet much more massive than Earth. There, humans would of course weigh more than on Earth and over generations we could expect them to become physically smaller as gravity will leave them little choice; unless their bones become much more massive and their hearts a lot more powerful; just two examples for potential adaptations. Of course, all of that also depends on the atmosphere on any such planet where this evolution would take place. As we have seen earlier, a more oxygen-rich atmosphere in Earth's past has allowed animals to grow quite a bit larger.

The short of it is, unless humans recreate Earth-like environments wherever they move, they will have to change biologically. It does not look like we have any choice in this regard. The umbilical cord that binds humans to all other life on Earth may leave us no other option than to ensure that wherever our destiny in space takes us, we take the biosphere supporting us on Earth with us. In essence, for all we know, humanity can only survive in the context of life on Earth and not by itself. From where we stand today, even with all the tremendous technological advances over the past few generations, biological evolution towards becoming a space-faring species sounds like science-fiction. However, given that we can count on human technology to evolve, we may eventually be in a position where we can choose to make those changes through genetic engineering instead of leaving such work to biological evolution, which would require much longer times to effect comparable changes. In essence, the path that our biological evolution will take in the very distant future seems likely to be governed by how advanced our technologies will be. This far out in the future we also have to ask what human lifespans will be. We are already developing the technologies to replace many of our human body parts so at some point it looks like human longevity will only be limited by the biological clocks seemingly built into every one of our cells. We have not become space faring societies quite yet and it will still take a long time until humans have settled in large numbers on other planets or moons and take to space like we take to the air nowadays. By the time that will be the case, we can assume that humans already will be genetically quite different from who we are today. We can only hope that human moral and ethical evolution will have kept track with human technological and biological evolution.

Epilogue



Rising Sun

Humans always have been travelers in time as well as in space. Our own personal journeys may seem rather brief ones, but they are not. If fortune smiles on us at birth, we may live to a ripe old age. Those not so fortunate among us may live shorter lives but their journeys may actually take them further. Length of life does not translate into more or less full and meaningful lives. We all start out as participants in what is Earth's greatest lottery, the lottery of birth. Depending on where this will put us on Earth, we will have very different experiences. However, those differences are small compared to what we all have in common. We are born blank slates, empty vessel with no personal past and just like dry sponges, we will soak up human culture and humanities past in whatever form or fashion it flows to us. We biologically inherit and culturally learn our human condition. Our bodies evolve in ways that our biological evolution has worked out over time spans unimaginable to us, shaping us into the humans we are today. At the same time, over the few years between birth and adolescence, our minds recapture much of the vastness of human culture evolution. Each of us is truly unique as every newborn human develops into a person that never existed before and as such will never exist again. However, we all connect to our shared human past and every one of us contributes a tiny little bit to our species common future. In Greek mythology, the Titan Prometheus had a brother, the Titan Epimetheus. Most of us are familiar with Prometheus, credited with creating humans from clay and bringing them not only the first fire, stolen from the Olympian Gods as it were, but also the gifts of civilization and art. The name Prometheus, literally translating into *fore-thinker*, or more liberally into foresight, identifying the one who looks forward, pretty much sums up what he stands for: human progress. His brother Epimetheus is likely a less well-known figure to most of us but for that no less important. In a sense, Prometheus and Epimetheus are representations of the same kind of duality that Yin and Yang represent in ancient Chinese philosophy. The name Epimetheus literally translated means *after-thinker*, more freely interpreted as hindsight, or the one who looks backwards in time. However, a better analogy would be to see Prometheus and Epimetheus as the two faces of one head facing into different directions, with Prometheus looking into the future and Epimetheus looking into the past. Humanities future and past as it is. In this sense, they represent the continued struggle between the past and the future as humankind evolves through time and space. Humans would be lost without their past, but progress is only possible with leaving some of that past behind. This conflict between our past and our future, the old and the new, lies at the heart of our human condition. It is what always has defined humans. As we strive to create new and better futures for all of us, we remain tied to where we come from. It is this fact, which makes it so important for us to understand what that actually is where we come from. Because only if we do, will we be able to navigate a future that is not imperiled by the past, and only then will we have a past that can wisely inform us about the sensibility of our future decisions and choices. Prometheus can only look towards the future because Epimetheus connects us to the past. Prometheus creating new futures ensures that Epimetheus past will never become stale. The two, Prometheus and Epimetheus, represent equally important aspects of humanity. They are the Yin and Yang of humankind, the duality that needs to be in harmony to be a productive force for humanities evolution and not a destructive one. Of course, other cultures than the Greek and Chinese have also articulated such dualism underlying humankind's history. It is just convenient to pick those two as one resonates in the East while the other one does so in the West.[1] This duality, or should we say the dialog between the past and the future, very much stands out when we look at the evolution of our understanding of the human condition. What we frequently refer to as established orthodoxy, the enshrined and approved knowledge and wisdom of the past, will always be in conflict with the new. Fortunately, over time we have found ways and means to have this discourse between the old and the new in a manner that does not exact blood anymore as it still did when a few hundred years ago, following the *Copernican Revolution*, Earth lost its place at the center of the universe. Today we settle such disputes in a much more civilized way, but the convictions held on either side are certainly no less strong or determined. The recent discussions with regard to climate change certainly come to mind. However, with climate change it is not religious belief pitched against scientific truth, it is commercial interests, or greed for short, pitched not only against scientific truth but even more so against the interests of future generations to come.

Figuratively speaking, each one of us carries all of humanities past while we work to create new futures. Living in the presence, sharing our common human past, and being able to envision and plan our future is fundamental to our human condition, for us as a species as well as individuals. We do not know about any other species that can engage in a discussion about the past, present, or the future, on a personal or on a societal level. The perceptions of time we have and how we can place real and imagined events into the stream connecting past, present, and future seems to be uniquely human. More than that, we can create our own pasts, of our cosmos, our planet, and of life on Earth by discovering it through the language of science. Of course, we also can create fictional pasts, presents, and futures in various ways. Actually, we have become quite good at that and for many it has become increasingly difficult to distinguish between fiction and fact. However, there is only one reality, at least in our universe, as we know it; and this is the reality of how our universe evolved, how our solar system was born, how our planet Earth formed, and how Earth and all life on it evolved together including our own species over the past roughly 3,800 million years. This reality is the firm ground we all must stand on as we look out into the future. Without our perception of reality solidly grounded in what we know about our universe, our planet, and life evolving on it, our arguments will be shaky, and we only will deceive ourselves as well as others. We also must acknowledge that this knowledge itself evolves over time as we continue to understand ever better how the world around us shapes our very human condition. Unfortunately, for some among us, deception and ignorance seem to be name of the game. How else could one still pretend to believe in the creation of all life just a few thousand years ago or more broadly, in any of these creationist worldviews based on ignorance?

When we are born, we start out as blank slates but in a couple of decades, we more or less represent the values and beliefs inherited from countless generations past. To that, we have added over the last several hundred years a much-increased objective knowledge of the world around us. We all live today in a world where it has become impossible to know pretty much everything as it may still have been true a thousand years ago. That is, if one had access to the accumulated knowledge of humankind. However, even a few thousand years ago, accumulating anything like complete encyclopedic knowledge would have been a tall order. When we look at what our children learn through their roughly twelve years of schooling, each one of them in principle has so much more factual knowledge about our world than any of the ancient sages could have ever had. Does that make our kids smarter than any of these ancient sages? As most of us would venture, quite likely not. The better question would be a different one. What could have any of these ancient sages have achieved if they would have had access to the same factual knowledge that typically high school students have today? Just imagine an Archimedes working on new inventions using computer engineering. Well, a genius in her or his time will likely always be a genius, regardless of where fate puts them in time. Newton, not known to be a humble man, made the rather humble statement that he only could see farther than those before him because he stood on the shoulders of giants. What he was saying is that we discover the truth by building on the discoveries made by others before us. Now, it is certainly important for a scientist to understand the discoveries made by those preceding her or him. However, it is no less important for each one of us to understand what those who came before us have found, and how and when they did so. It is in this respect that it really matters to have at least a sense of how vast our universe is, what its general characteristics are and how we only discovered much of that over the past four hundred years or so. In a similar vein, the same is true for our own world, how Earth and life evolved on it closely interdependent, eventually including our very own species.

It is the presence or the lack of a general understanding of these things that set the framework for our human condition. In an ideal world, it all would be a matter of choice between ignorance and knowledge. However, that is not the case in our seemingly modern world. Much of humankind still has no such choice and too many children grow up in traditions and beliefs that limit what they can learn and know. The duality of Prometheus and Epimetheus, of Yin and Yang, shapes our human condition as we move through time. At any given point in time, Earth has been and continuous to be home to diverse human cultures. This cultural richness of humanity is a gift to all of us but at the same time, it also is a source of conflicts. How we manage the latter and how we ensure that our human cultural richness continues to thrive while eventually all of humankind moves into modernity poses a tremendous challenge. Resolving such conflicts has never been easy. The different speeds at which human societies and cultures have changed since the beginning of the industrial revolution has made this even more of a challenge. Sometimes these conflicts are not so much between societies like for example between the economically lesser and the more prosperous parts of our world but within societies, rich and poor ones alike, highly developed or not. Economic interests may be stirring these conflicts in part, at least sometimes, but more often than not the issues at stake are of a much more fundamental nature.

On a journey through our understanding of the universe, Earth, and life such as the one in this book, much has to be condensed and simplified. But that should not leave anyone in doubt about the magnificence of it all. Many such narratives use the language of science because it allows us to understand our world. This language, however, is suspect to some among us as it seemingly replaces the divine humans used to see in nature throughout much of their history with profanity. Far from it. Anyone who truly pursues a deeper understanding of our cosmos, our world, and life as we know it through science will be in for a spiritual as much as a scientific journey. Sadly, in our everyday world, this connection has largely been lost. Science delivers utility and the world of engineering, mathematics, and natural sciences could not be further apart from the human world of spirituality.

Spirituality is as much a fundamental part of our human condition as is our curiosity that has produced the sciences which provided the foundation of our modern world. This is not a conflict between any specific religion and our modern world even though sometimes it may be tempting to see it this way. This is a conflict at the heart of human nature itself. In a way, much of our world has disconnected from the spirituality that for our ancestors was part of their lives. Throughout human history, and we can safely assume the same for much of human pre-history, humans had a spiritual as well as a material life. It was the equivalent of shamans and later of one organized religion or another that served their spiritual needs. That actually may be where the problem lies. Over many centuries, organized religions almost completely co-opted human spirituality. Not to say that this is a bad thing in itself. However, these religions over time served more and more the needs of those in power and religious life became increasingly constrained, resembling more of a divine bureaucracy than a divine spirit. Of course, that led also to revolts and spawned for example much of the protestant movements throughout Christianity. The religious and philosophical currents in the Eastern World followed a different path. Notably, for most of the time, there was more open-mindedness towards a multitude of religious and spiritual experiences. In a sense, Eastern World spirituality was always suspicious of one or another religion claiming to possess the indisputable and final truth.

Circling back to our modern world, we see religions seemingly at odds with much of science. What once was the indisputable truth is no more. The Bible cannot help those who were born into Christianity in discovering the universe as it is or in understanding the evolution of Earth and life. Nor can any other religious text. It may take Muslim societies a little longer to realize that than it has taken for example Christianity. For Muslims, the Koran is not a collection of texts written by humans but according to belief has been dictated by God right into the hand of their prophet. Does anybody really think that such beliefs will still be around in their current forms a few thousand or a million years from today? Religions, just as everything else, change over time, just much more slowly so. Everything in our societies is constantly changing and humans are constantly adapting. Inevitably, our societies move towards becoming ever more technology based. While difficult, imagining the future of human civilizations from a technological perspective is still something we can meaningfully engage in. We all may have our own visions of life in the cities of a distant future or how far our colonies on Moon or Mars will have evolved in a few hundred years. A much harder and more difficult challenge is to envision our spiritual future. As pointed out above, religion while still providing consolation for many has largely failed our modern societies, but so has philosophy. Arguably, the lives of our ancestor, whether they were more religiously or philosophically inclined or not, were from a spiritual perspective on average quite a bit richer than ours are today. With traditional sources of spirituality mostly failing us, many seek out new sects or spiritual experiences unless they already have succumbed completely to our materialistic societies. However, most of these are even less capable to meet their spiritual needs.

One could wonder if art has not to a good extent taken much of the space that in the past was the preserve of our spirituality. In addition to bringing humans civilization, of which worship of the Gods must have been part of, Prometheus also brought humans art. Art provides a unique way for humans to express themselves and it always had and frequently continues to have a deeply spiritual and religious connotation. Not just in Christianity but also in other religions, even if restricted in what kind of art they allow, like it is the case with calligraphy in Mosques. Anyone who has listened to some of the church music written by humanities greatest composers must feel the deep spirituality this music carries. There can be no doubt that art in a sense can and does stand in for spirituality and religion. However, there also can be no doubt that there is no replacement for true philosophical, religious, or spiritual thought exploring and reconciling the contradictions that are the very essence of what conflicts our human condition.

Given the grave inequities which characterize human life on Earth, there are those who believe that satisfying material needs will solve all problems. But ultimately, this approach will fall short. Humanity has impressively demonstrated its capability to use the tools of science to understand our world and that it can come up with engineering solutions for most of the needs of our modern societies. There clearly is some hubris associated with that but overall, the evidence is still sufficient to count on humans to be able to evolve technologically way beyond where we stand today. Nothing like that is the case for our cultural and spiritual development. We cannot solve what are deep societal problems with an engineering solution approach, and there is certainly no way to engineer human spiritual needs out of humans. We place great care into how we educate our children in the practical skills they will need to find interesting jobs, but we invest far too little in their spiritual development. In today's world and even more so in the future, it will not suffice to educate children in the religion of their ancestors and take them to prayer every Friday, Saturday or Sunday, depending in which religious communities they happen to grow up. A long time ago, in generations long gone by, communities revealed the meaning of life as they understood it gradually to their younger members as they grew up. Those tightly knit communities are mostly gone now and families in our modern world cannot replace them in this respect. Nowadays, nearly everyone is pretty much left to discover the meaning of life on her or his own. What a challenge and what a journey to be taking by yourself. Even more so as finding answers to such questions that every human being asks at some point has become ever more difficult.

In the distant past, people also grew up in a context that gave meaning to their lives in a much more natural way. The meaning of life became a much harder question to answer with the onset of the industrial revolution. What meaning is there in spending all of your adult life, which back then started way earlier with what we would call today child labor, doing nothing else but seemingly meaningless repetitive work for endless hours, day for day without relent. Of course, there was family and other ones to be cared for and there was the consolation of religion. There still was hope for something better to come in a different life. In some places, similar naked exploitation still prevails such as in textile factories producing fashionable garments for people who may have no inkling that they are wearing a product that is only so inexpensive because the workers who made it, mostly women, are poorly paid and have no rights. While things have improved and in most countries such conditions are long gone, new and bigger challenges are plaguing our economies and societies.

What should young people do today with their lives when in practice their societies tell them that they are not needed anymore? Humanity is at the threshold of entering a new age, the age where we may run out of work. How will people be able to support themselves without work? How can we replace the sense of dignity and meaning that work gives to so many even today? What dignity is there in living off handouts? It may not happen in our generation but in the not-too-distant future, we can envision a society where robots will do much of the work leaving much less for humans to do. The difficult task will not be adjusting the distribution of material wealth in our societies. There is enough food and material wealth to go around with plenty left over once everybody has their basic needs satisfied.[2] However, it will be much more difficult to assure the spiritual well-being of a largely non-working society. For most, work is a source of motivation, gratification, and dignity. Without work, where do people get this from? Societies that just redistribute wealth and otherwise do not care about the personal development of their every member will not function.

But we need not look that far into the future. Even today, it is already clear that, as important as technological evolution may be for our societies, it all will come to nothing if we cannot find a way to offer a meaningful life to all. That will first require a deeper look into what we mean by a meaningful life. If some are tempted to associate a meaningful life with how most of the people live now in the developed nations that may not work out. Spiritually, people in the rich countries of the developed world may be poorer than their fellow-humans are in the more economically disadvantaged communities of the developing world. A meaningful life likely has little to do with being materially well-off. Material safety yes, but riches, are they really necessary for a meaningful life? We will have to re-think the cultures of money and greed that pervade the rich countries today to different degrees. Do we really wish this on the currently less fortunate in our world? One of the great travesties is that in principle there is more than enough food and water for no one on this planet to go hungry or thirsty today; but still, many continue to die of male nourishment or hunger and access to clean drinking water remains a luxury for many. Why is this not something that we can change? How can one part of the world sit down with a full plate and a glass of wine in hand in front of their evening news and watch the misery of another part of this world while enjoying their dinner. What is wrong with us? Even more so, how can this often happen in one and the same country?

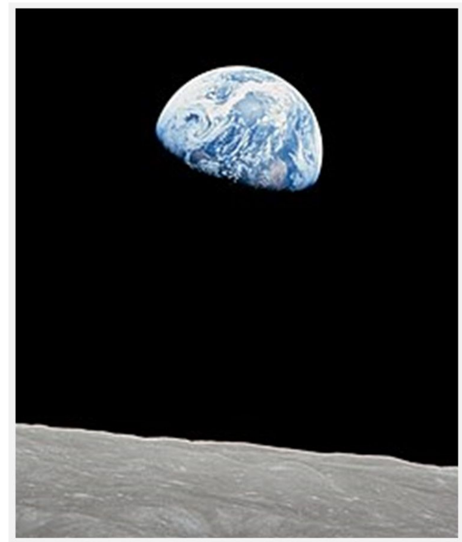
Well, if you are looking for answers yours truly does not have them any more than you do. However, it should be crystal-clear that our human species is unlikely to endure and master the challenges of colonizing space as long as we have not solved our challenges here on Earth. If we are truly to become evolution's way to spread life beyond Earth, we need to change very much if we really want to succeed.

The pendulum of history has kept swinging back and forth since humans built their first civilizations. Sometimes we moved forward a few steps only to retrace half of it later. However, the overall trajectory has been in the right direction, at least for humans; for other life on Earth it is likely a quite different story. Today, more people live in relatively free societies than ever before, and we have continued to move closer together. We only have to let the past century with all its atrocities and horrors pass before our mind's eye to understand how far we have come. However, none of these achievements towards a more integrated and united world are unchallenged.

The use of language once associated with barbarity or racism seems to become acceptable again. All too easily are we ruled by fear and it is our fears that populists and their ilk leverage to nudge us back into a world which we believed we had already consigned to the past. We will have to see how long this lasts but eventually, the pendulum will turn the other direction again. Long-term, there is no alternative to a much more integrated and open world with much more equitable living conditions and opportunities. We should have no illusion about how long it will take humanity to get there but eventually it will. It is at that point that we can and must devote our energy to spread Earth's life among the stars.

When people ask why we are here, there are few better answers than for spreading life throughout the universe. It is through life as humanity represents it that the universe knows itself. If it were not for us, who would wonder about the miracles of the cosmos, who would try to understand it? Our universe is a universe of life, and it is highly unlikely that we are the only intelligent species evolving in this universe. It is up to us if we will be a species able to spread Earth's life among the stars or not. If we fail, it will all have been a grand evolutionary experiment, nothing more. We know that our solar system and our Earth have a finite lifetime. We do not know how long the universe will exist but from all we can understand, it will be there for countless billions of years to come, ever so slowly changing from our human perspective. We will likely never stop wondering why there is a universe in the first place, but it remains doubtful if we ever will find an answer to this question, but we can hope.

Whatever the answer could be we may come up with, it may not be any more satisfying as the little we currently know and it likely will only continue to raise new questions. The situation we find ourselves in is not unlike with those Russian dolls where there is always a new smaller doll hidden in the one we open; with the slight difference that there may be a countless number of dolls to open. The big question – "Why is there anything at all?" - will not go away. We may become ever more technologically sophisticated as a species. As we do so, we must hope that our spiritual and cultural evolution will keep track with our technological evolution. They have to because the two are at the core of what defines our human condition. If they diverge too much in their speeds of development, it may not bode well for our species.



Earthrise

Will we ever be able to answer the question of what life is and how it did originate? I do believe so. However, as interesting as this question is and as fascinating the pursuit for answers may be, this is not what is most important about life on Earth. What really matters is the future of life on Earth and beyond Earth. Only if there is a future for life on Earth and beyond the time when Earth becomes inhabitable, only then will it matter to have answers to those questions. Evolution has endowed us with capabilities that other animals do not have, and it has handed us the responsibility of becoming the stewards of life on Earth and beyond it. This is what we must understand and accept. If we do not or if we should fail, life on Earth will have just been a one off, not meant to go beyond Earth, maybe just one such experiment of evolution among very many in the universe. It is up to us to ensure that the life of Earth spreads throughout the universe as far as we can take it, hopefully seeding new life as we go. For now, we are doing pretty much the opposite. We are destroying the basis for life on Earth, and we even may manage to go extinct as a species before our time. If so, we can almost expect with certainty that new life will emerge on Earth eventually, life that in time could produce another intelligent species such as us. However, too much time may have passed before such a new caretaker of Earth's life could emerge to ensure its endurance beyond Earth's expiration date. This very responsibility should give meaning to all of our lives as all of us have a role in it to play. Ensuring an Earth environment that remains hospitable for life as long as possible while preparing a future for life after Earth is the true mission evolution has given to our species. If we succeed, then we will not only have technologically evolved to a level we cannot even imagine today, we also will have become a spiritually new species with our human moral and ethical center of gravity in a very different and much better place than it is today.
It is now a little more than fifty years since the crew of Apollo-8 took the famous *Earthrise* picture as their spacecraft reemerged from behind the Moon to encounter a rising Earth, just as from our Earth we see a rising Moon. For good reason, this picture has become one of the most famous images taken in the last century. As all the astronauts who experienced this *Earthrise* firsthand were quick to realize, this image more than anything else should make it plain clear to any human what a unique, beautiful, and small place Earth is. A small blue marble as seen from the Moon and nothing more than a speck of dust in an enormous universe, if at all. Humanity with all life on our planet is hurtling through space on this small spaceship called Earth. How can anyone who has seen this picture not realize that we are all in this together, all of life on Earth including our species? We will either all survive and thrive together on spaceship Earth or we will be erased from the records of the universe just as we had never existed. The choice is ours.

Notes

Prologue

1. It may be a worthwhile science endeavor to understand how much of this greed is just part of our species characteristics vs. being an acquired characteristic. We do pride ourselves in what we believe differentiates us most from other animals, but it may not be so much our big brains or reasoned thought that are our major species characteristics. It could just as well be our greed and other darker character traits that on an individual and societal level differentiate us from other animals.

Gaining Perspective

1. Not many other animals – if they could voice this - would likely agree with us that they fared equally well. Please look for Yuval Noah Harari's *Brief History of Mankind* in the bibliography section if you are interested in an eloquent indictment of human transgressions, including our crimes against other animal species.
2. There are several phases of Stonehenge roughly being dated between 3,000 to 2,000 BCE. We should note that there are also a number of other megalithic structures that align with the sunrise of the winter solstice.
3. The term pre-Socratic philosopher refers collectively to a group of philosophers active in the ancient Greek world in the sixth and fifth centuries BCE, before and up to the time of Socrates .
4. The ruins of ancient Milet are close to the modern village of Balat in the Aydın Province of Turkey.
5. Our source for this is the Greek scholar Diogenes Laertius, the third century CE biographer of the Greek philosophers, about whom we know little more than his only surviving work, the *Lives and Opinions of Eminent Philosophers*.
6. The spheres had to be of a transparent nature so one could see the stars shine through them; crystalline spheres were the natural choice.
7. The literal translation of the Ancient Greek word *atom* is uncuttable.
8. The word polyhedra is a combination of the Ancient Greek *poly* for many and *hedra* for base or seat. Many "seats" is the more proper translation as all regular polyhedra have in common that their sides are all made made up of either triangles, squares, or pentagons. They are essentially all dices which when thrown on a flat surface always will come to sit on one of their identical faces. The criteria defining a regular polyhedron are: 1) the polyhedron is convex; 2) every face is a regular polygon; 3) all the faces are congruent, i.e. identical; and 4) every vertex is surrounded by the same number of faces.
9. The ether, this fifth essential substance needed to explain the universe in addition to the four others – fire, air, water, and earth - often was referred to as the the quint-essential substance, "quinque" meaning five in Latin. Therefore, when you refer to something as quintessential, you are actually making in a manner of speaking an astronomical statement.
10. Eratosthenes realized that the curvature of the Earth results in different lengths of noon-shadows in Northern vs. Southern places which he measured at the Egyptian cities of Aswan and Alexandria. He also came up with the first practical way to determine the number of primes between any two specified numbers, the so-called *Sieve of Eratosthenes*.
11. When we talk about the geocentric model of the universe, we may automatically assume that this means that people back then thought Earth was as sitting at the very center of a sphere around which everything revolved. Not quite so. Hipparchus found that for his model to describe the apparent movement of the Sun, he had to move Earth slightly away from the center of the circle that the Sun's orbit seemed to

follow. The Sun's circular orbit, the so-called deferent, was centered on the so-called eccentric which was removed from Earth's position. Later Ptolemy would introduce other points controlling the movements of planets, removed from the eccentric the same distance as Earth but in the opposite direction from the eccentric, the so-called equants controlling the speeds of planet movements. Therefore, in a strict sense, the geocentric model had Earth only approximately at its center. We should note here that the earliest use we know of concepts such as eccentric, deferent, and epicycles by Apollonius of Perga (c. 262 – c. 190 BCE) precedes Hipparchus by one generation.

12. Others include changes in the shape of Earth's path around the Sun between more or less elliptic, and changes in the tilt of Earth's rotational axis with respect to the plane of this ellipse.
13. More correctly, an epicycle describes the movement of a fixed point on the circumference of a rolling small circle whose center moves around the circumference of a larger circle. As pointed out in chapter note 11, Ptolemy was not the first to think of epicycles to describe the motions of planets. The concepts of eccentric, deferent, and epicycles traces back to Hipparchus and Apollonius, and may predate them but we do not know that. However, Ptolemy brought them together in a systematic model of our universe that would continue to be in use until the early Renaissance.
14. In Greek mythology, Andromeda was the daughter of Cepheus and Cassiopeia, king and queen of the African kingdom of Aethiopia, then referring to the lands south of ancient Egypt. Andromeda, meaning the ruler of men, was punished for her mother's hubris who thought her more beautiful than the sea nymphs, an offense to Poseidon. However, she was rescued by Perseus.
15. They also tried to invade the eastern part of the empire, but they were less successful there. It took another ∼1,000 years and the added onslaught of Goths, Arabs, Persians, Slavs, and Turks to finally bring down the remaining vestiges of the Eastern Roman Empire in 1453.
16. Not really in science but certainly in philosophy and theology as evidenced by Augustine of Hippo (354 – 430), whose influence on Christian religious belief, thought and doctrine can hardly be overstated.
17. The term science would not have made sense to scholars back then in the way we understand it today, but it is the right analogy to use here. Let's just keep in mind, there were no scientists back then but certainly intellectually curious people who sought true knowledge and that is what science is really built on.
18. His real name was Peter Bienewitz but as many in his time he used a Latin version of his name. "Biene" is the German word for bee and the Latin word for bee is "apes".
19. The first use of the name America dates to a 1507 world wall map attributed to Martin Waldseemüller (c. 1470 – 1520). Peter Apianus is credited with producing the first printable map that used the name America.
20. Uranus was discovered in 1781, Neptune in 1846, and Pluto in 1930. In 2006, Pluto has been demoted from a planet to the status of a dwarf planet.
21. Topology, the mathematical discipline that is concerned with such things, describes spheres as having a single surface and no edge, whereas a flat cylinder, i.e., a flat Earth, has three surfaces, the mantle and the bottom and top lids, plus two edges separating the mantle from the lids. Therefore, an ant could crawl on a sphere surface forever without encountering an edge but not so on a flat disk.
22. This is the Latin version of his name that most of us are familiar with, but the birth name of this German-Polish astronomer was Niklas Koppernigk or Mikołaj Kopernik in its German or Polish form, respectively.
23. This change happened as a consequence of a conflict in which the Polish Crown and the Prussian Union, an alliance of Prussian cities and nobles, had joined forces to fight the Order of the Teutonic Knights who posed a constant threat to both.
24. A canon at the time was an ecclesiastical position in the Catholic Church just below that of a bishop that involved all kinds of administrative responsibilities with regards to running the local church economies as the church not only owned much land but also ran mills, bakeries, or breweries. It was not uncommon for someone like Copernicus coming from a well-connected family to be awarded such a position already in younger years to provide him an assured income.
25. In Greek mythology Helios is the personification of the Sun, descended from his Titan-father Hyperion and his Titan-mother Theia.
26. The Latin word for Earth is Tellus.

27. It is said that he held his newly published works firmly in his hands lying on his deathbed, knowing that the Catholic Church would not be able to make him retract anything of his work. True or not, it is a good story.
28. Ephemerides, a plural word, derives from the Latin word "ephemeris" which the Romans adopted from the same Greek word which means "diary".
29. His real name was Filippo Bruno.
30. The Inquisition was a Roman Catholic Church institution founded in the mid-thirteenth century to combat heresy and was finally abolished in the early nineteenth century. We probably will never fully know how many people over the more than five-hundred years of its existence the Inquisition killed, tortured, or how many lives it destroyed even though the victims physically survived their ordeals.
31. The representatives of the Reformation perceived Giordano Bruno as a similar danger to religious authority. They could be equally intolerant and vindictive as their Roman Catholic counterparts, including burning their critics on the stake, when they deemed their authority threatened.
32. It must be remembered that the power of rulers was seen as divinely derived and rulers were only answerable to God, and of course to his representatives on Earth, i.e., the Pope and the Catholic Church.
33. That Giordano Bruno's view of an infinite universe and many other worlds may have been the real threat to exclusive Catholic Church authority is supported by the fact that only two years before Bruno published his two major works, the 1582 calendar reform commissioned by Pope Gregory XIII (1502 – 1585) used the work of Copernicus. For worldly affairs, we still use today what is known as the Gregorian calendar not only in the West but practically worldwide. However, Christian religious calendars differ since then, as the Eastern Orthodox Church continues to hold on to the Julian calendar to this day.
34. The four Jupiter moons Io, Europa, Ganymede, and Callisto are also known as the four Galilean Moons and are the four largest of the 79 moons of Jupiter that we know of today.
35. Sidereus Nuncius also included the drawings of the Moon surface.
36. The Medici rose to power in fourteenth-century Florence first as bankers and by the fifteenth century had become the most powerful family, ruling Florence in all but in name for much of the time. Medici power reached its zenith in the sixteenth century when they established themselves as hereditary Dukes and then Grand Dukes of Florence and Tuscany, a title they would keep until the early eighteenth century when they not only were virtually bankrupt but more importantly, failed to produce a male heir. The Medici would produce four Popes, three in the sixteenth century and another one in the first decade of the seventeenth century.
37. The warning was given to Galileo by Cardinal Bellarmine (1542 – 1621), one of the most respected Catholic theologians of his time and in essence the judge in the dispute between Galileo and his church opponents.
38. The sixth-century philosopher Simplicius of Cilicia (c. 490 – c. 560), one of the last pagan philosophers, was well known for his commentaries on Aristotle, in fact, he seems to have done nothing else.
39. It would take the Roman Catholic Church more than 350 years and a thirteen-year investigation before finally in 1992 acknowledging its error in condemning Galileo Galilei.
40. Leiden University, founded 1575 in the city of Leiden in the Dutch province of South Holland, is the oldest university in The Netherlands.
41. The most famous one was the supernova from 1572, now referred to as SN 1572, a supernova of type Ia in the constellation Cassiopeia. At the same time this supernova was also observed and discussed in other places, China being one of them.
42. It is with the *Rudolphine Tables* that Tycho Brahe's original star catalog, which he had compiled in 1598 just before he was exiled from Denmark, finally was published.
43. For example, early on Kepler used the five Platonic Solids to build a model of the solar system.
44. Conical sections are curves produced by slicing through a cone. They include circles (eccentricity $= 0$) and ellipses ($0 <$ eccentricity < 1), as well as parabolas (eccentricity $= 1$) and hyperbolas (eccentricity > 1).
45. Of the planets discovered in later times, only Pluto has a greater eccentricity than Mars but because Pluto lost its planetary status in 2006, Mars again is planet number one in our solar system in terms of highest eccentricity.

46. Newton's corpuscular theory of light only came into its own in the twentieth century when Albert Einstein wrote his seminal paper on the photo effect in 1905 and then in the 1920's when the particle / wave duality of matter and mass-less particles such as the photon became enshrined in quantum mechanics as we know it today.
47. It was almost unfathomable at the time that such a mathematical innovation could be produced simultaneously by Newton and Leibniz but today we know, this is what happened. Unfortunately, each of the two geniuses, who both could lay claims to many other firsts, thought the other stole his work and bitterly disputed priority until the end of their lives.
48. The modern value for Earth mass is $5.97 \cdot 10^{24}$ kg, a number with twenty-four figures before the decimal point or expressed differently, 5.97 trillion trillion kg.
49. Quoted from a letter that Newton wrote 1692/93 to the classical scholar Richard Bentley (1662 – 1742), a future Master of Trinity College.
50. Georgium Sidus is Latin for George's Star, so Herschel's name implies a star or comet rather than a planet. Saturn in Greek-Roman mythology was the father of Jupiter, and Uranus the father of Saturn. So, there is some logic to the planetary naming sequence as one moves outward from Jupiter to Saturn to Uranus.
51. This sounds like the description of an eclipse and in essence, a transition is indeed a very partial eclipse of the Sun.
52. The exact value is: 1 AU = 149,597,870,700 meters.
53. Thule, referring to an island north of Britain, is the farthest north location mentioned in ancient Greek and Roman texts. In this sense, Thule marked for the ancient Greeks and Romans the end of the world in the North and hence the proper translation for Ultima Thule is something like *"Beyond the known world"*.
54. He measured the elapsed time between successive eclipses of the Jupiter moon Io when Earth was moving away from Jupiter as compared to when Earth was moving towards Jupiter. From the differences in the measured intervals, he could estimate the speed of light.
55. The States General of the Netherlands declined his patent invention as similar claims for invention had also been made by others. It seemingly was one of those cases where the time was ripe for an invention. Nevertheless, the Dutch government realized the importance of this invention for the help it could provide in its ongoing war for independence from the Kingdom of Spain and rewarded Lipperhey accordingly for his design.
56. Nebulae is the plural of the Latin word "nebula" which means fog but is often translated as cloud which is not quite accurate as cloud in Latin is "nubes". Nebula in astronomy today refers to different kinds of interstellar clouds.
57. Orion is the famed giant huntsman of Greek mythology, placed into the heavens by Zeus himself, according to other sources by his daughter, the goddess Artemis. Orion was out hunting with Artemis when he was killed by a scorpion, sent out by Gaia, the Greek mother Earth herself, to prevent Orion from killing every living beast on Earth as he had promised to do. Coincidentally, along with Orion, Zeus or Artemis placed the scorpion into the heavens as well, thereby naming two star constellations.
58. Today we know that the Orion Nebula is a star nursery where thousands of stars are born. In honor of Huygens discovery, the bright interior of the Orion Nebula is referred to as the *Huygenian Region*.
59. There are different explanations for the naming of the Pleiades cluster, two of them more often cited than others. The first one derives it from the ancient Greek word *plein* meaning "to sail", an indication of their importance to navigation since antiquity. The second one traces their name to the seven daughters of the Titan Atlas, referred to as the Pleiades after their mother, the Oceanid Pleione, a Nymph. As to how the seven sisters were placed into the heavens there are also different accounts. One is that they killed themselves out of grief after hearing about the death of the Hyades, their half-sisters from Atlas and another Nymph; another one has them rescued by Zeus after being pursued for seven years by the giant huntsman Orion. In both accounts, Zeus places them into the heavens where, as the star constellation Pleiades, they now are next to the Hyades cluster, also placed there by the gods, while still being pursued by the giant huntsman Orion placed into his own constellation.
60. Today there are many different and more extensive star catalogs, some general looking at everything that can be found in the night sky and some specializing on specific types of objects. However, references to celestial objects by their Messier or NGC numbers are still frequently used.

61. The name Milky Way is derived from the Greek word *galaxias*, which refers to the “milky one” because it appeared to the ancient Greek as a “milky” band in the sky. According to Greek mythology, milk was spilled by the goddess Hera when she realized she had been made to breast feed infant Heracles, one of her unfaithful husband’s Zeus many children, in this case from a dalliance with Alcmene, Heracles human mother.
62. Thomas Wright did not actually say how far radially out our solar system lies from the Milky Way Galaxy center but that has no impact on his argument as it just needs to be removed from and not close to the center.
63. In his Milky Way map William Herschel assumed that our solar system was located near the center of the galaxy.
64. Our Milky Way Galaxy is a so-called barred spiral galaxy which are spiral galaxies characterized by a bar-shaped structure of stars aligned with the longer axis of the spiral galaxy. The disks of spiral galaxies are not circular and therefore usually characterized by two size parameters in the plane of the disk bracketing the extensions of the spiral arms. For the Milky Way Galaxy those bounds bracket the disk size of our galaxy between about 150,000 and 200,000 thousand light-years. The central bulge of our galaxy containing the majority of stars is more symmetric with a central diameter in the disk of around 40,000 light-years and a bulge extending out of the disk in both directions with a diameter of roughly 30,000 light-years.
65. The Hooker telescope was named for the American businessman John Daggett Hooker (1838 – 1911) who helped fund it.
66. In Greek mythology Cepheus was the name of a mythological king of Aethiopia featuring in the Perseus legend as the husband of Cassiopeia and father of Princess Andromeda. The name Aethiopia is itself derived from the Ancient Greek *aethiops*, referring to a dark-skinned person. Historians such as Herodotus used the name Aethiopia for the people living south of ancient Egypt.
67. We now know that the Magellanic Clouds, the Large Magellanic Cloud and the Small Magellanic Cloud, are two dwarf galaxies about 160,000 and 200,000 light-years away from us, respectively.
68. At the time of Edwin Hubble, it was not understood why Cepheid stars behaved as they do. However, that did not matter as long as there was experimental verification of it.
69. The effect was discovered by the Austrian physicist Christian Andreas Doppler (1803 – 1853) in his studies of acoustic waves.
70. Joseph von Fraunhofer (1787 – 1826) was a Bavarian physicist and master optical lens maker. He was the first to discover and study the dark absorption lines in the spectrum of the Sun that are now referred to as Fraunhofer lines in his honor.
71. It was the Scottish botanist Robert Brown (1773 – 1858) who in 1827 while looking through a microscope first observed particles, in this case pollen, moving through the water. What caused the random movement of those particles remained a mystery until Albert Einstein’s 1905 explanation showed it to be proof of the existence of atoms and molecules.
72. Their postulation became known as the Lorentz or Lorentz–FitzGerald contraction. It was made after the Michelson-Morely experiment of 1887 had found no evidence for the ether. The purpose of the proposal was to save the concept of a stationary ether in which light would propagate by suggesting that the length of a moving object was shortened in the direction of its movement.
73. The muon is about 207 times heavier than the electron. The other large brother of the electron is the tauon which is about 3,477 times as heavy as the electron. Muons and tauons carry the same negative electrical charge as the electron. Similar to the electron which has as its positively charged counterpart, the positron, there also are positively charged muons and tauons.
74. Here is the math: 2.2 millionth of one second multiplied by 300 million meters per second, the speed of light, equals 2.2 times 300 meters or 660 meters.
75. You would not want to be in that elevator. For a typical elevator with a distance between the sidewalls of 3 meter the acceleration required to move 1 millimeter upwards before the light beam hits the wall opposite to the hole is orders of magnitude higher than the gravitational acceleration we experience at Earth’s surface, which is 9.82 m/s^2, often referred to as $1\,g$. As a quick back of the envelope calculation will show, this acceleration would have to be of the order of two-million-million g. Anybody in that elevator would be complete squashed. However, that is the beauty of a Gedankenexperiment – nobody

does have to be in the elevator. Given the equivalence of acceleration and gravity that also means that it will take enormous masses, such as stars, to bend light in a measurable way.

76. The relationship between frequency f and wavelength λ of light is simply $c = \lambda \cdot f$ with c being the velocity of light.
77. Cosmos is the Ancient Greek word for describing an ordered world in contrast to Chaos. According to ancient Greek mythology at the beginning there was Chaos which either referred to the primeval emptiness of the universe before things came into being or the abyss of Tartarus, the Greek underworld. Our today's meaning of the word Chaos goes back to the Roman poet Ovid (43 BCE – 17 CE) whose full name was Publius Ovidius Naso. He identified Chaos with the original disordered and formless mass out of which order, that is the Cosmos, was created.
78. The term parallax derives from the Ancient Greek word *parallaxis* and translates as "alternation". It describes a difference in the apparent position of an object when viewed from two different positions. One can simply visualize the concept by stretching out one arm with the thumb pointed up and watching how its apparent position changes against the background if we look at it alternating with the left and the right eye closed. The parallax angle between the respective lines of sight measures such changes of perspective.
79. The star he chose, 61 Cygni in the constellation of Cygnus, which is Latin for swan, is actually a double star but that was not known to Bessel at the time. His measurement estimate of about 10.3 light-years was close to the actual 11.4 light-years.
80. In fig. 1.6 the distance Earth to star is the hypotenuse of the right-angle triangle formed by Earth, Sun, and Star. Simple trigonometry gives: $tan(1") = 1\,AU/1\,pc$. Each of the 360 degrees of a circle contains 60 arc-minutes and each arc-minute has 60 arc-seconds so there are 3600 arc-seconds in a circle. Converting from degrees to radians we get $1\ arcsec = (1/3600)\cdot(\pi/180)\ radian$. Finally using $tan(x) \approx x$ which holds for such small angles we can solve for parsec and get: $1\,parsec = 1\,AU \cdot 3600 \cdot 180/\pi$ or $1\,parsec = 648{,}000\,AU/\pi$. As one light-year is equivalent to 63,240 AU we get the relation $1\,parsec \approx 3.26$ light-years.
81. The apparent magnitude m_1 of a star with Flux F_1 is given by $m_1 = F_1 = L_1/(4\cdot\pi\cdot D^2)$. When Herschel devised his apparent magnitude ruler, he established it by defining the relationship of the magnitudes m_1 and m_2 of two stars such that $m_1 - m_2 = -2.5\cdot log_{10}(F_1/F_2)$. If star *1* is 100 times brighter than star *2* then $m_1 - m_2 = -5$. The absolute magnitude M of a star with apparent magnitude m is nothing else than the brightness this star would have if it were observed at a fixed distance D_M versus its real distance from the Sun D_m. Using the flux definition, the difference between this absolute magnitude and the apparent magnitude can be written as: $M - m = -2.5\cdot log_{10}(L_M/L_m)$ - $5\cdot log_{10}(D_m/D_M)$. Specifying distances in parsecs and setting $D_M = 10\ pc$ while realizing that the intrinsic luminosity of the star has not changed, $L_M = L_m$, we get: $M = m - 2.5\cdot log_{10}(1) - 5\cdot log_{10}(D_m) + 5\cdot log_{10}(10)$. Cleaning this up, the absolute magnitude of a star is then given by $M = m - 5\cdot log_{10}(D) + 5$, where m is its apparent magnitude and D is the true distance of the star from the Sun measured in parsec.
82. A bolometer measures the power of electromagnetic radiation via heating of a material that absorbs all wavelength like a black body and has a temperature dependent electrical resistance.
83. A black-body radiator/emitter is a calibration device that allows to precisely compare temperatures of bodies emitting electromagnetic radiation. It derives its name from the fact that a perfect black body is the ideal absorber and emitter of electromagnetic radiation. In practice, a black body radiator can be realized in its simplest form by heating up a large metal block that is hollow on the inside and has a tiny opening. The electromagnetic radiation emitted through the tiny hole is a close approximation to the radiation that would be produced by a perfect black body at the temperature of the interior of the metal block.
84. Most stars are black body radiators or very close to being such. Therefore, the luminosity L of a star with radius R and temperature T is given by $L = 4\cdot\pi\cdot R^2\cdot\sigma\cdot T^4$.
85. The relationship is $M_V = 4.83 - 2.5\cdot log_{10}(L_S/L_\odot)$ where L_S and $L_\odot$ are the stars and the Sun's luminosity, respectively. It is a special version of comparing the absolute visual magnitudes of two stars as discussed in chapter note 81, with one of the stars being the Sun.
86. Using the expression for a star's luminosity L as stated in chapter note 84, the relative luminosity of a star with respect to the Sun's luminosity $L_\odot$ is $L/L_\odot = (R/R_\odot)^2\cdot(T/T_\odot)^4$. To deduce what the curve looks like which for example describes all stars which have a radius 10 times the Sun's radius $R_\odot$ we

simply replace the star radius R with $10{\cdot}R_\odot$ and get $L/L_\odot = 100{\cdot}(T/T_\odot)^4$. In the Herztsprung-Russel diagram the logarithm of $L/L_\odot$ is plotted over the logarithm of the temperature so we get: $log_{10}(L/L_\odot) = 4{\cdot}log_{10}(T/T_\odot) + 2$, which is a straight line in a double logarithmic plot. Using $T_\odot = 5{,}780$ K and inserting the high and low temperature values on the temperature axis in fig. 1.7 we get: $L/L_\odot = 2.5{\cdot}10^4$ at the high temperature side (23,000 K), and $L/L_\odot = 1.43$ at the low temperature side (2,000 K) of the graph, a straight line starting high on the left and ending lower on the right side of the graph. The same exercise can be done for stars with multiples or fraction of 10 of $R_\odot$ producing lines that are shifted parallel up or down from the one with $R = 10{\cdot}R_\odot$ used as an example here.

87. For Earth to have a stable orbit around the Sun, Newton's gravitational force must balance the centrifugal force Earth is subject to as it is slung around the Sun in the Sun's gravitational field: $F_{Gravity} = F_{Centrifugal}$ or $G{\cdot}M_\odot{\cdot}m/r^2 = m{\cdot}v^2/r$. Solving this for the Sun's mass $M_\odot$ results in: $M_\odot = v^2{\cdot}r/G$. We only need to know the velocity of Earth around the Sun, the radius of its orbit and G, Newton's gravitational constant. G can be measured, similar for the radius r and the velocity of Earth around the Sun is $v = 2{\cdot}\pi{\cdot}r/T$ where the time T measures Earth orbital period around the Sun. The Sun's mass is then given by $M_\odot = (4{\cdot}\pi^2/G){\cdot}(r^3/T^2)$. All that needs to be known to calculate the mass of a star is a companion planet or star of which the orbital period and the orbital radius can be measured. This is actually nothing else than a restatement of Kepler's third law where the ellipse has become a circle. Given the large distances involved in measuring the orbital distances in binary star systems, this is usually a very good approximation.
88. Four years later, Henry Norris Russell found through his own work that Cecilia Helena Payne-Gaposchkin was right, and he gave her credit in his 1929 publication.
89. The neutron was discovered by the British physicist James Chadwick (1891 – 1974). Four physicists, the Russian-American George Gamow (1904 – 1968), the Dutch-Austrian-German Fritz Houtermans (1903 – 1966), and the British Robert Atkinson (1898 – 1982) used the quantum mechanical tunnel effect to show that the likelihood of fusion reactions predicted by quantum mechanics was much higher than what classical physics would predict. The American physicist Carl D. Anderson (1905 – 1991) discovered the positron in 1932 which was predicted by British physicist Paul A.M. Dirac (1902 – 1984) in 1928. The existence of the neutrino was predicted by Austrian physicist Wolfgang Pauli (1900 – 1958) in 1931 as a necessity to balance energy and momentum in nuclear reactions. The experimental verification of the neutrino had to wait until the 1956 measurements by American physicists Clyde L. Cowan (1919 – 1974) and Frederick Reines (1918 – 1998).
90. Today's mass-luminosity relationship used for stars with masses from around the solar mass $M_\odot$ to up to 20 times $M_\odot$ is $L \propto M^{3.5}$. Dividing the total energy available to the star which is proportional to a star's mass by the rate it is burned up gives the lifetime estimate $\tau \propto E/L = M/L$ for the star. Replacing L with the mass-luminosity relationship we get for the lifetime $\tau \propto 1/M^{2.5}$, which only depends on a star's mass. Stars heavier than the Sun live shorter and stars lighter than the Sun live longer. Using this back of the envelope estimate, stars with masses of five or ten times $M_\odot$ would have lifetimes shorter by factors of 56 and 316 than the Sun, respectively.
91. In Greek mythology, the Hyades, half-sisters of the Pleiades, are the five daughters of Titan Atlas with the Oceanid Aethra, a Nymph. As with so much in Greek mythology there are different accounts as to how they received their place in the night sky as the five stars in the head of the constellation Taurus, or Tauros to use the Greek version. In one of them it is a reward for nursing the infant Greek wine god Dionysos in another version they were stricken so much by grief over the death of their brother Hyas that Zeus took pity on them and placed them into the heavens.
92. Famed because of four lines of surviving archaic Greek poetry frequently attributed to the poet Sappho (c. 630 – c. 570 BCE) from the Island of Lesbos. antiquity considered her one of the greatest poets and the surviving lines of her so-called midnight poem are familiar to many: *The moon and the Pleiades have set, it is midnight, and the time is passing, but I sleep alone.*
93. The force counterbalancing gravity in this case is what physicists call electron degeneracy. Two electrons cannot occupy the same space, the closer they are being pushed together the stronger they will resist; this is also known as Pauli's exclusion principle and applies to all fermions, electrons being one example. The gravity of a white dwarf star is not sufficient to overcome this electron degeneracy pressure.
94. The remaining neutrons are in part the neutrons from the iron and nickel atomic nuclei and in part

the protons of these nuclei that were converted into neutrons. The principle of this neutron degeneracy pressure withstanding further gravitational collapse is similar to the electron degeneracy pressure holding up white dwarf stars, but their physics underpinning is quite different.

95. The lower limit comes of course from the iron-nickel core mass the parent star must have had to trigger the supernova that produced the neutron star in the first place.
96. The title of John Michell's paper read in 1783 is: *On the Means of discovering the distance, magnitude, &c. of the Fixed Stars, in consequence of the Diminution of the Velocity of their Light, in case such a Diminution should be found to take place in any of them, and such other Data should be procured from Observations, as would be farther necessary for that Purpose.*
97. The term *Big Bang* was coined by the British physicist Fred Hoyle. Hoyle was a brilliant physicist, but he believed in the equivalent of an eternal universe and never accepted that our universe had an origin. When he first referred to the theory of origin for our universe his colleagues were developing as the *Big Bang* it was not meant as a compliment but intended to ridicule the idea.
98. Gamow added a second author to the paper who had actually done no work on this paper at all. It was one of Gamow's many practical jokes that arranging the last names of the sequence of authors as Alpher, Bethe, and Gamow would invoke the first three letters of the Greek alphabet: alpha, beta, and gamma. While Hans Bethe did not contribute to this specific paper, he was one of the first to take Arthur Eddington's idea of nuclear fusion in stars seriously and became one of the main contributors to the theory of stellar nucleosynthesis.
99. Lord Kelvin was known as William Thomson before he was knighted. The Kelvin temperature scale is named after him.
100. How is a 46,500 million light-year radial extension of the observable universe not a contradiction with light only being able to travel 13,800 million light-years since the Big Bang? Consider the simple analogy of a balloon, which we blow up just sufficiently to mark two dots on its surface. Then we start to draw a sine wave from one point towards the other as we continue to blow it up. As the balloon's surface expands, we get closer to the other point but the distance between them keeps growing. In addition, we can observe the distance between peaks and valleys in the sine curve we already have drawn increasing. Just as we cannot separate the expanding balloon surface from the line which we are drawing on it we cannot separate the properties of light from space. We cannot think of light as propagating in a space that is somehow separate and expanding. Space is an underlying property of light and the stretching of light-waves is the measure of the expanding space. This is the reason why we can observe objects that first sent out light 13,800 million years ago, even though they are today some 46,500 million light-years away from us.
101. Werner Heisenberg formulated the *Uncertainty Principle* named after him in 1927: $\Delta x \cdot \Delta p \geq h/4 \cdot \pi$, where Δx is the uncertainty of the location x of a particle and Δp is the uncertainty of its momentum p which is given by the product of its mass m and velocity v as $p = m \cdot v$. If the location is known exactly then Δx becomes close to zero and as a consequence Δp, the uncertainty of the particles momentum and velocity, can become arbitrarily large. A similar relation exists for the variables energy and time: $\Delta E \cdot \Delta t \geq h/4 \cdot \pi$. The smaller the time interval Δt is we look at, the larger the uncertainty in energy ΔE can be. Loosely speaking, this energy then is available for matter / antimatter particle pair creation without violating the conservation of energy by ever so briefly borrowing it from the vacuum; that is as long as the matter / antimatter pair recombines to annihilate within that time interval, returning the energy to the vacuum. Because of their fleeting existence physicists call such matter / antimatter pairs virtual particles.
102. Erwin Schrödinger's 1925 wave equation was the first mathematical description encapsulating the wave-particle duality fundamental to quantum mechanics.

Our Home - Planet Earth

1. Ionia is an ancient region on the central part of the Mediterranean coast of modern-day Turkey including the Ionian Island chains in front of it which are part of present-day Greece. It derives its name from a Greek tribe called the Ionians who settled on these shores and on many islands sometime in the sixth

century BCE. Later on, the name extended beyond the original Ionian Greeks and became common for Greeks in much of the Eastern Mediterranean sharing a similar culture.

2. Greek citizens were not citizens of countries or nations, these concepts did not exist back then, but were citizens of their native cities. When they went to war, a frequent result was that part of the losing side ended up as slaves of the winning side. Becoming a mining slave as for example in the silver mines of Athens was certainly one of the worst outcomes in that case.
3. The Sui Dynasty (581 – 618) produced only two emperors and succeeded to unify China for the second time after it first had been unified under the Qin Dynasty (221 – 207 BCE). Introducing the civil service examinations was part of the Sui Dynasty government reforms to restore and centralize imperial authority.
4. The marine fossil location was removed hundreds of kilometers from the nearest ocean shores in the Taihang Mountains which stretch through China's Hebei, Shanxi, and Henan provinces.
5. Georgius Agricola is the Latin version of his German name Georg Bauer, "Bauer" being the German word for farmer and "Agricola" the Latin translation of that.
6. Nicolas Steno is the Latin version of his real name, Niels Steensen.
7. Fossilized shark teeth were known since ancient times, but their origin remained a complete enigma. Pliny the Elder believed they fell from the sky during lunar eclipses. Later, they were thought to be the tongues of serpents that Saint Paul had turned to stone while visiting the islands of Malta. Thereby they came to be called *glossopetrae*, a Latin version combining the Greek words for tongue *glossa* and stone *petros*.
8. The word *isogonic* translates from Greek as "having equal angels". Isogonic lines are lines where the angle between the true north and the magnetic north, the magnetic declination, remains constant.
9. Neptune was the name ancient Romans gave to Poseidon, the Greek god of the sea and earthquakes. He was the brother of Jupiter, the Roman equivalent of the Greek god Zeus, and of Pluto, the Roman equivalent of the Greek god Hades who ruled the Greek underworld.
10. Vulcan was the Roman god of fire and as such the equivalent of the Greek god Hephaistos.
11. The first volume of Lyell's *Principles of Geology* that Darwin took with him on the five-year HMS Beagle around-the-world exploration was a present from the Beagle's captain, given to Darwin on their departure from Plymouth in December 1831. Before that, Darwin was not familiar with Lyell's work. However, once he started reading it and more importantly, when it helped him understand the geology of places the Beagle visited, he became a convinced disciple of Lyell.
12. The word seismology derives from the Ancient Greek word for earthquake *seismós* and the ending *logía* for "the study of". It literally means the study of earthquakes.
13. Paleontology is the branch of science studying all life, as it existed in the distant past. The name combines three anglicized derivatives from the Ancient Greek words for old, being, and logical discourse, paleo-onto-logy.
14. The word tectonics stems from the Latin version of the Ancient Greek *tektonikos* with the meaning of "pertaining to building", which for geology translates into the building processes resulting in the structure of Earth's crust.
15. There is not complete consensus among geologists on how to count and label lithographic plates. Some count not seven but eight major plates because they split the Indo-Australian plate into an Indian plate and an Australian plate. There are quite a few more minor plates than listed and shown here with the count varying between twenty-one to twenty-two plates. In addition, the naming of some of the minor plates is not completely consistent either and reflects some regional and national preferences. Some sources for example do show an Anatolian plate while others do not.
16. Gondwana was named in reference to geological formations in the Gondwana region of central northern India which are similar to formations of the same age on Southern Hemisphere continents.
17. Pangea is Greek for "All Earth".
18. Isostasy is derived from the Greek words for equal, *iso*, and standstill, *stasis*, and is the name given to Archimedes principle as applied to explain the uplift of Earth's crust in plate tectonics. Archimedes principle states that a body completely or partially submerged in a fluid is pushed up by a buoyant force - the uplift - equal to the weight of the fluid displaced by the body.
19. The name pyroxene is derived from the Greek words for fire *pyr* and stranger *xenos*.

20. Peridotite is named after the gemstone peridot which describes a gem-quality olivine mineral. The etymology of the word peridot itself seems not to be clear.
21. One of those places is the Tablelands region of Gros Morne National Park in Newfoundland, Canada.
22. The mineral olivine derives its name from its typically olive-green color. A gem-stone quality olivine is referred to as peridot.
23. Like all sciences, geoscience carries with it some historic nomenclature, and this is where the D" layer label comes from.
24. If a material undergoes a phase transition the chemicals in the material do not change but the way in which they bind to each other does. Examples of phase transitions we are all familiar with are the transitions from ice to water and from water to steam.
25. Wadsleyite is named after the Australian chemist and crystallographer Arthur David Wadsley (1918 – 1969).
26. Ringwoodite is named after the Australian Earth scientist Ted Ringwood (1930 – 1993).
27. 1 GPa, one giga-Pascal, is close to 10,000 times the atmospheric pressure we experience on Earth's surface.
28. The perovskite mineral structure is named after the Russian mineralogist Lev Perovski (1792 – 1856).
29. Bridgmanite is named after the American physicist Percy W. Bridgman (1882 – 1961), considered to be the father of high-pressure experiments.
30. The wüstite mineral structure was named after the German metallurgist Fritz Wüst (1860 – 1938).
31. The periodic table of elements we use today goes back to the Russian chemist Dmitri Ivanovich Mendeleev (1834 – 1907). He not only figured out how to properly order the elements according to their chemical properties but importantly had the foresight to leave empty spaces for elements not yet discovered but which he believed must exist based on the gaps in his table. The gaps were due to the much smaller number of chemical elements known in his time. These gaps now have all been filled and in addition humankind has also created a small number of short-lived synthetic elements that Mendeleev could not have dreamed off, but which perfectly fit in his scheme.
32. From our high-school teachers we all learned about Pythagoras of Samos as the discoverer of $\boldsymbol{a}^2+\boldsymbol{b}^2=\boldsymbol{c}^2$, the relationship between the three sides of a rectangular triangle with $\boldsymbol{c}$ being the longest side, the so-called Pythagorean theorem. However, more importantly, he was an eminent philosopher and the founder of Pythagoreanism, one of antiquities most important philosophical schools.
33. The late Bronze Age civilization of Mycenaean Greece roughly dates from 1600 to 1100 BCE and is named after one of its most prominent archeological sites, Mycenae with its famous Lion Gate, in the north-eastern Peloponnese of modern-day Greece.
34. It is still an unanswered question today if there ever really existed a poet genius called Homer or if the most famous epic poems of antiquity, the *Illiad* and the *Odyssey*, are the works of many authors over a much longer period of time that somehow became associated with the name Homer. We may never know.
35. The Canadian Shield is a massive U-shaped rock formation that encircles Hudson Bay in Northern Canada.
36. Stratum is the Latin word for something spread or laid down and strata is the plural form of it. Stratigraphy is the branch of geology that studies rock layers, the strata, and processes of stratification, how strata are being laid.
37. In ancient Greek mythology Kratos is the divine personification of strength and if used as a verb it means strength.
38. The name Rodinia derives from the Russian word "rodina" meaning motherland or birthplace or the related verb for giving birth or to beget, "rodit".
39. The name comes from *chondros*, the Ancient Greek word for grain.
40. Carbonaceous chondrites were originally assumed to be rich in carbonaceous material because of their typical gray-to-black appearance but it turned out later that their appearance owes more to morphology, a fine-grained structure, than to elemental composition. While "carbonaceous" is a misnomer the name stuck. It may actually gain some validity again as several of them have been found to contain amino acids, the foundation for carbon-based lifeforms as we know them.
41. In Greek mythology Theia was the mother of Selene, the goddess of the Moon.

42. The Moon completes a rotation around its own axis every 27.3 days which is the same time it takes the Moon to rotate around Earth. Because of this condition, scientists call it synchronous rotation, it looks to observers on Earth as if the Moon does not rotate at all.
43. That is unless we reject a linear extrapolation or comparable simplistic back-calculations. But it is better to acknowledge that nature usually is not that simple, and we still know too little about the dynamics of the early Earth-Moon system.
44. Rocks are made of several different minerals and do not have one single melting point such as a homogeneous material like a bar of iron. Rather, they melt in stages until eventually the temperature is high enough to melt all of the material in the rock. At which temperature this occurs depends of course also on the surrounding pressure. The deeper a material sinks into Earth the higher the pressure will be and the higher the temperature needs to be to melt parts of it or eventually melt it completely.
45. As with all things in classification schemes, there is usually never complete agreement on what to count as a single craton or for example two cratons. Hence, it depends on who does the counting, and this is why there are roughly 35 cratons.
46. The name Vaalbara is composed of the last four letters of the two cratons it is supposed to have been composed of, the Kaapvaal Craton, today located in eastern South Africa, and the Pilbara Craton, today located in northwestern Australia.
47. "Ur" in German refers to the oldest instance or the fundamental essence of something.
48. The name Kenorland land derives from the Kenoran orogeny, a mountain building process that took place on what is now the Canadian Shield about 2,500 million years ago. The rocks formed at the time are among the oldest on the North American continent, prominent in places such as the Lake Superior area of Canada; this is why the supercontinent Kenorland is also referred to as the Superia supercontinent.
49. Please search on NASA's website, https://www.nasa.gov/, for what differentiates weather from climate.
50. Foraminifera is a combination of the Latin words for holes "forami" and bearer "nifer".
51. The process is a bit more complex as first the heavier oxygen is preferentially condensed out of the air to form ice sheets in higher northern or southern latitudes after which precipitation from the remaining O-16 enriched air continues to grow the ice. Moist air rising from the oceans is O-16 enriched thereby replenishing the already O-16 rich polar air.
52. Albedo is the Latin word for whiteness. It measures the fraction of solar radiation Earth receives that is reflected back towards space and its value ranges from 0 to 1. A value of one means everything is reflected back, like from a perfect white surface, and a value of zero means everything is absorbed, like from a perfect black surface. Albedo is also used to describe the energy balances of other planets with regard to the radiation they receive from and reflect back into space.
53. The Scoville heat unit to measure food spiciness traces its origins to the American pharmacist Wilbur L. Scoville (1865 – 1942) who was first to devise a scale of pungency.
54. The names of the atmosphere layers are derived from Ancient Greek words combining *tropos* for turn, change or response, *stratos* for layer or level, *meso* for middle, *thermo* for heat, and *exo* for outside, external or beyond with the word *sphere* for globe.
55. The word meteor derives from the Ancient Greek word *meteoros* literally meaning "high in the air".
56. The word thermohaline derives from the combination of the Ancient Greek words for temperature, *thermos*, and salt, *halas*, with haline having the meaning of salty from which our word saline of course derives.
57. Along coastlines a reverse process can also occur when wind blows in a direction that causes surface waters to build up along the coastline with the surface water eventually sinking toward the bottom; this reverse process is called down-welling.
58. There are of course much more detailed climate classification systems. The most widely used one goes back to a system originally proposed by the same Wladimir Köppen we already encountered when discussing the ice ages.
59. For anyone really interested to see for themselves how far some climate models were already in the late 1960's, the landmark paper to read is "Thermal Equilibrium of the Atmosphere with a Given Distribution of Relative Humidity" authored by Syukuro Manabe and Richard T. Wetherald in 1967

who back then were with the United States Geophysical Fluid Dynamics Laboratory of the National Oceanic and Atmospheric Administration in Washington, D.C.

Life On Earth

1. The Latin word "anima" can mean breath, spirit, life or soul and is the Latin translation of the Ancient Greek word *psyche* whose original meaning identified the life-giving breath.
2. We have little biographical information on Alcmaeon of Croton, but he may have been born around 501 BCE.
3. Magna Graetia is the name given to the Greek colonies that were established by various Greek city-states in Sicily and in southern Italy. Croton, or Crotone as it is named today, was a colony founded by the Achaeans. Achaea can refer to a region in the Northwest Peloponnese but is also used by Homer to refer to Greeks in general as *"the Acheans"*.
4. Akragas has become the modern city of Agrigento which is also the capital of the Italian province of Agrigento in Sicily.
5. Democritus Abdera in Thrace is today's Avidra, a town in in the Greek province of Thraki on the peninsula opposite to the island of Thassos.
6. The Greek world seemingly spread everywhere as Greek city-states and later during the Hellenistic period Greek states founded colonies wherever their ships could sail, trade was lucrative, or their armies were victorious. Because of that, many places along the shores of the Mediterranean and Black Seas as well as all over Asia Minor and as far as into the Hindu Kush can trace their origins to Greek founders including such cities as Alexandria, Beirut, Istanbul, Marseilles, Naples, Odessa, Qandahar, Syracuse, Tripoli, or Varna; just to name a few of the larger cities today outside modern Greece.
7. The island Kos is part of the Dodecanese Island chain a few miles off the Turkish coast.
8. Aristotle used different names for the groups, but it is better to use illustrative examples for how he arranged life along the ladder as, e.g., today we know that the first two groups belong to the same family as whales and dolphins are indeed mammals.
9. Taxonomy is concerned with the science of defining and naming groups of biological organisms on the basis of shared characteristics.
10. Lyceum is the Latin translation of the name of the philosophical school Aristotle founded in Athens. The French translation of the name, Lycée, was introduced during the time of Napoléon Bonaparte (1769 – 1821) for the main secondary education establishments in France, the French equivalents of High Schools in the United States and of Gymnasiums, another borrowing from Greek culture, in German speaking countries. Eventually, many francophone countries also adopted the Lycée.
11. *Agricola* is the Latin word for farmer and agriculture, a term only coined in medieval times from the Latin words for land or field, "agri", and cultivation, "cultura". Agriculture and farmers mattered very much to Romans and the Roman state. It was the citizen armies fielded by their farmers which protected and built Roman power in the first place.
12. Dioscorides and Galenus are both Latinized names, their original Greek names were Dioskouridēs and Galenos. Anazarbus, then a city in the Roman province of Cilicia is now modern-day Anavarza in the Adana Province of Turkey. Pergamum, or Pergamon in Greek, was a pre-eminent city in ancient Greece, and then a city of the Roman province of Asia, today the city of Bergama located in the Izmir province of Turkey; its archeological site lies about half-an-hour drive from the Aegean coastline.
13. India's Vedic Age is the period following the Indus Valley Civilization. Vedas are the ancient religious texts of India. The Sanskrit words *"ayur"* and *"veda"* translate into "life" and "knowledge", hence, Ayruveda denotes the *Knowledge of Life*.
14. We only have to remember that even the great Isaac Newton would still indulge in alchemy.
15. The Latin name Paracelsus translates into equal to or greater than Celsus. Aulus Cornelius Celsus (c. 25 BCE – c. 50 CE) was a Roman encyclopaedist living in the first century and was known for his writings on medicine, the only part of his larger work that survived. This work, *De Medicina*, was first published in 1478, only a few years before Paracelsus was born. Maybe Paracelsus most famous statement was: *"Only the dose makes the poison"*

16. It is still being debated who the actual inventor of the microscope was and there are several candidates, but it seems to be agreed that it was invented in the lens-making centers of The Netherlands, which have also given us the telescope.
17. One specific discrimination Maria Sibylla Merian faced as a woman was that the painter's guild of her adopted home city of Amsterdam did not allow women to paint in oil so her amazing color illustrations were mostly done using watercolors.
18. Linnaeus was the first to understand that whales are actually mammals. As for the class of verme, it basically included all invertebrate species that were not invertebrates with exoskeletons, segmented bodies, and paired jointed appendages. Examples for the latter being spiders, scorpions etc. which are grouped under then name arthropods; the latter derived from the combination of the Ancient Greek words for joint and foot.
19. The full title is: *Le Règne animal distribué d'après son organisation, pour servir de base à l'histoire naturelle des animaux et d'introduction à l'anatomie comparée* (*The Animal Kingdom Arranged after its Organization; Forming a Natural History of Animals, and an Introduction to Comparative Anatomy*).
20. It would have likely been much more difficult for Darwin's views on evolution to prevail as quickly as they eventually did without the dedicated support of his closest supporters, the core of which included: the British naturalist Thomas Henry Huxley (1825 – 1895), sometimes also referred to as Darwin's bulldog; the British botanist Joseph Dalton Hooker (1817 – 1911), the longtime director of the Royal Botanical Gardens, Kew; and of course the British geologist Charles Lyell, as mentioned in the text. In addition to that, Darwin had also strong early supporters in mainland Europe, specifically so in Germany and interestingly also in the United States.
21. Chromosome is a combination of the Ancient Greek words for color, *chroma*, and body, *soma*. They received their name because they were first observed as color-stained filaments under the microscope. Chromosomes are threadlike structures of nucleic acids and protein found in the nucleus of living cells and carry genetic information in the form of genes. Mitosis is derived from the Ancient Greek word *mitos* for thread in reference to the thread-like structures seen during mitosis. Meiosis comes from the Ancient Greek word *meiosis*, which means lessening and refers to meiosis reducing the chromosome numbers by half in daughter cells compared to the parent cell.
22. The names given to the three germ-layers stem from the sequence of their appearance. In Ancient Greek *ecto* refers to outer, *meso* to middle, and *endo* to internal or inner while *dermis* translates to skin or tissue.
23. The word ontogeny derives from the Greek words *on* for being and *geneia*, essentially referring to how something is made.
24. The word phylogeny derives from the Greek words *phylon* meaning race or stock and *geneia*, referring to how something is made. The phylon is also a taxonomic category below the level of kingdom (see also tab. 3.1).
25. The name protista is a Latin version of the German word *Protisten* which Haeckel derived from the Ancient Greek word *protistos*, having the meaning of "the very first" or "primordial".
26. The name eukarya refers to cells having a nucleus as indicated by the combination of the Ancient Greek words *eu* for good, well or right and *karyon* meaning nut or kernel.
27. The word cytology derives from the Ancient Greek words for cell, *cyto*, and discourse, *logía*. The prefix "cyto" is also used to reference a number of cell properties as for example in cytoplasm or in cytoskeleton.
28. Robert Brown made also contributions outside of biology, notably the discovery of what became to be named *Brownian Motion* and they arguably made him more famous than his description and naming of the cell nucleus.
29. The name virus is of Latin origin where it denotes a venomous substance, or a poison for short.
30. Bateson derived the word genetics from the Ancient Greek word *genno* which translates into "to give birth".
31. The reason why X-rays are called this way is because the letter X designated for physicists a then unknown radiation. Since Röntgen's discovery they are also known as Röntgen rays. Like for visible light-waves, there is a whole spectrum of X-rays from hard X-rays with wavelength of 0.2 to 0.1 nanometers to soft-X-rays with wavelength of a few to a few ten nanometers. For comparison, the visible green light

our eyes are most sensitive to has a wavelength of around 550 nanometers; one nanometer equals a meter divided by one-thousand-million, or one billion.

32. The Rosetta Stone found in 1799 was key to understanding the ancient Egyptian language as on the stone was written a text in two languages, Egyptian and Greek, using three different scripts, namely hieroglyphic, demotic, and Greek. With two of them already known, the hieroglyphic script could be deciphered.
33. Fred Griffith made one of the greatest discoveries in modern biology, but it looks like he did not see it this way back then as he himself did not pursue his original discovery but devoted himself to research on other bacterial topics.
34. The two life cycles of bacteriophages are referred to as the lytic cycle and the lysogenic cycle. The lytic cycle causes the disintegration of the host cell by rupture of the cell wall or membrane in a process called lysis which is ancient Greek for "to unbind"; the lysogenic cycle leaves the host bacterium intact while the viral genome is spread through bacterial reproduction, that is until the cycle turns lytic and the host cell dies too.
35. The word epigenetics combines the Ancient Greek prefix *epi* for "over or outside of" with the word genetics (see chapter note 30) to describe changes that effect gene activity and expression without altering the DNA sequence.
36. It does say something about us that we designate ourselves as the wise ones; and it begs the question what the next stage in human evolution will be called.
37. Scientists refer to the combined mass of water found on, under, and above the surface of a planet as the hydrosphere.
38. Rocks are composed of various kinds of minerals, which can be significantly older than the rock itself as for example discussed earlier for the 4,400-million-year-old zircons found in Australia.
39. The word stromatolite is a combination of the Ancient Greek words for layer *stromatos* and stone *lithos* an appropriate description for the layered stone built from fossilized microbial mats.
40. The letter was addressed to the botanist Joseph Dalton Hooker, one of Darwin's closest friends.
41. In an auto-catalytic reaction one of the reaction products is also a reactant and therefore a catalyst in the same or a coupled reaction.
42. Chert is a hard, fine-grained sedimentary rock composed of silica compressed under high pressure.
43. The name derives from the Ancient Greek words *phyto* for plant and *plankton* meaning to wander or drift. Phytoplankton, sometimes also called microalgae, are microscopic photosynthetic organisms. While many of them are plants, many are also not and include species more closely related to bacteria and protists than plants. Examples of phytoplankton are cyanobacteria as well as diatoms, dinoflagellates, radiolaria, or ciliata.
44. As for prokaryotes, which did not have a nucleus with DNA, they were likely relying on RNA to carry their genetic information. In our bodies RNA carries the instruction from our DNA for synthesizing proteins.
45. The name mitochondrion derives from the combination of the Ancient Greek words for thread, *mitos*, and granule, *chondrion*. They break down food molecules to produce for their host cell the chemical adenosine triphosphate (ATP) which is the energy carrying molecule common to all life.
46. The label acritarch combines the Ancient Greek words for confused, *akritos*, and origin, *arche*, thereby acknowledging the seemingly confused origin of these tiny fossils.
47. Protozoon, the singular form, is a combination of the Ancient Greek words *proto* for primitive or first and *zoion* for animal.
48. Pseudopodia is another Ancient Greek word combination from the word *pseudo* for false and *podion* for little foot, from which we also derive the Latin word podium that has entered the English language.
49. The word eustatic was coined by the Austrian geologist Eduard Suess to refer to worldwide sea level changes and combines the Ancient Greek words *eu* for well or good and *stasis* for "standing still".
50. As is the practice in paleontology, discoverers of new species get to name them. In this instance, the name combines an indicator for the fossil location, Charnwood Forest, and the last name of the discoverer, the schoolboy Roger Mason, resulting in *Charnia masoni*; only later did it become clear that the schoolgirl Tina Negus had discovered the fossil already a year earlier than Mason.
51. The name conodont is derived from a combination of the Ancient Greek words for cone, *konos*, and tooth, *odont*.

52. In ancient Greek mythology, the Titans Iapetus and Atlas are father and son. What the naming implies is that the Iapetus Ocean was bounded by some of the same continental masses bordering today's Atlantic Ocean which was named after Atlas.
53. The word orogeny is derived from a combination of the Ancient Greek words for mountain, *oros*, and creation or origin, *genesis*. It describes the process by which Earth's crust is folded and deformed by lateral compression, resulting in the formation of new mountain ranges.
54. Many graptolite fossils resemble hieroglyphs written on rock which is reflected in their name, deriving from the combination of the Ancient Greek words for written, *graptos*, and rock, *lithos*.
55. The continental mass known as Euramerica or Laurussia is not to be confused with the supercontinent Laurasia which came later. Because of the distinctive sandstone deposits with oxidized iron mineral inclusions, the landmass referred to as Euramerica or Laurussia is sometimes also called the Old Red Continent or the Old Red Sandstone Continent.
56. Stomata is the name given to the pores found in plant leaves or stems through which the gas exchange required for photosynthesis occurs. The name is derived from the Ancient Greek word for mouth, *stoma*. Functionally equivalent pores in organs are also referred to by the same name.
57. We humans also have one large trachea through which air is transported to and from our lungs. Arthropods have many smaller tracheae bringing oxygen inside the arthropod exoskeleton from where oxygen diffuses through the tissues of the body.
58. The name lignin is derived from the Latin word for wood, "lignum".
59. The name coelacanth is a Latin version of a combination of the Ancient Greek words for hollow, *koilos*, and spine, *akantha*, referring to the hollow spines in the tail fin found in the first fossils of these fish; interestingly, first described by the very same Louis Agassiz we met when discussing the discovery of the ice ages.
60. Tiktaalik means large, freshwater fish in Inuktitut, the language of the people of the Nunavut Territory in Northwestern Canada.
61. The word amnion derives from Ancient Greek referring to a little lamb, in the sense of giving birth to one.
62. The word placoderm derives from the Ancient Greek words for plate, *plaka*, and skin, *derma*.
63. The name Panthalassia is derived from the Ancient Greek words for all, *pan* and sea, *thalassia*.
64. The origin of the German word Zechstein can indicate a very hard material as "tough stone" translates to the German "zäher Stein"; or it can be a reference to the German word "Zeche" which refers to an organization set-up to run a mining operation.
65. The name dinosaur is derived from the Ancient Greek words for terrible, *deinos*, and lizard, *sauros*.
66. The name pterosaur is derived from the Ancient Greek words for winged, *pteros*, and lizard, *sauros*.
67. Echoing this transition, almost two hundred million years later, the mammalian ancestors of whales and dolphins would make a similar journey back to the sea.
68. The names for ichthyosaurs and plesiosaurs are derived from the Ancient Greek words *plesios* for "near to", *ichtys* for fish, and *sauros* for lizard.
69. The name sauropod is derived from the Ancient Greek words for lizard, *sauros*, and foot, *podion*.
70. The name Archaeopteryx is derived from the Ancient Greek words for ancient or old, *archaios*, and feather or wing, *pteryx*.
71. The name cephalopod is derived from the Ancient Greek words for head, *kephaolos*, and foot, *podion*.
72. The word placenta comes from a Latin word for a cake which has its origin in the Ancient Greek word *plakoenta* for flat or a slab-like object, referring to the appearance of its human version; the name marsupial is derived from a Latin version of the Ancient Greek word *marsippos* for pouch; and the name monotreme comes from the Ancient Greek words *monos* for single and *trema* for hole, referring to where the egg is pushed out of the body.
73. The name Mosasaur is derived from the Latin name "Mosa" for the French river Meuse and the Ancient Greek word for lizard, *sauros*.
74. The Cretaceous is named for the chalk deposits that are characteristic for it. The German word for chalk is "Kreide" and this is where the K in the K-T extinction event comes from; the T did not have to be changed as the old period scheme introduced in 1760 by the Italian geologist Giovanni Arduino used a Latin derived word, the Tertiary, to name the third period in this now obsolete scheme.

75. The Messinian is a geological subdivision of the Miocene epoch which lasted for around two million years, from ∼7.3 million years ago to the beginning of the Pliocene, 5.3 million years ago.
76. The name Dryopithecine is derived from the combination of the Ancient Greek word for tree, *drys*, and the Latin form of the Ancient Greek word for ape, *pithēcos*.
77. The five-to-seven-million-year time frame is based on genetic analysis and not on fossil records.
78. The genus designation *Sahelanthropus* reflects the geography where the fossils of this human, *anthropos* in Ancient Greek, was found, the Sahel desert, while the species designation *tchadensis* indicates the specific country where it was discovered, modern-day Chad.
79. In the local language of the Toubou people of Chad where the skull had been found Toumaï means "hope of life".
80. The genus name *Orrorin* has the meaning of "original man" in the Tugen language spoken in parts of Kenya while the species designation *tugenenis* reflects the specific geography of the fossil site, the Tugen Hills in Kenya.
81. The genus and species names reflect on the region where the *Ardeptihecus* fossils were found. In the East African Afar language *Ardi* translates into ground or floor which combined with the Latin form of the Ancient Greek *pithēcos* for ape results in the name meaning ground-ape; *kadabba* in Afar identifies the father of a family; and *ramidus* comes from the Afar word *ramid* meaning root.
82. The genus name Kenyanthropus reflects of course the geography where the fossil of this human species was found, Kenya.
83. The genus name *Australopithecus* meaning southern ape derives from the Latin words for southern "australis" and the Latin form of the Ancient Greek word for ape, *pithēcos*, reflecting the geography where the first skeleton of a member of this genus was found, southern Africa.
84. The genus name Paranthropus reflects the sentiment that this linage is not seen as a direct ancestor to modern humans as it is a combination of the Ancient Greek words beside, *para*, and human, *anthropos*.
85. It was the American-born British mining engineer Charles Watson Boise (1884 – 1964) who financed the 1959 expedition to the Olduvai Gorge in Tanzania that found the *Paranthropus boisei* fossils, led by the British-Kenyan paleoanthropologist couple Louis Seymour Bazett (1903 - 1972) and Mary Leakey (1913 – 1996).
86. The wiser than wise human...

The Human Endeavor

1. The British biologist, banker, and politician John Lubbock (1834 – 1913) in his *Pre-historic Times*, published in 1865, first used the term pre-history. Lubbock was a younger friend of Charles Darwin and Charles Lyell, and he introduced the term pre-history to refer to the time span bridging the geological past as described in Lyell's publications and our modern times.
2. John Lubbock was also first to divide the Stone Age into the Neolithic and Paleolithic in his 1865 *Pre-historic Times*; the Mesolithic was inserted a year later as an intermediate category by the British archeologist Hodder Michael Westropp (1820 – 1885). The Ancient Greek word for stone is *lithos*, and *paleo*, *meso*, and *neo* are the respective Ancient Greek words for old, middle and new. So Paleolithic, Mesolithic, and Neolithic refer to the old, middle and new stone ages respectively.
3. The name Oldowan is derived from the Olduvai Gorge, a major fossil site in modern-day Tanzania. The name Acheulean is derived from a locality named Saint-Acheul near Amiens in northern France. The names Levallois and Mousterian are of course also of French origin, the first one referring to a fossil site that is now a suburb of Paris, Levallois-Perret, and the latter derived from the location of the fossil site Le Moustier in southwestern France.
4. It is called the *First Agricultural Revolution* to distinguish it from the *Second Agricultural Revolution*, which started in England with the adoption of industrial style farming along with the industrial revolution and spread to the rest of the world, ushering in the transition towards monoculture, the agricultural practice of growing single crops on a large scale. Arguably, growing genetically modified food crops could be the *Third Agricultural Revolution*. The full consequences of the latter are not known yet, such as its

impact on biodiversity and its percolation throughout the food chain, and of course human consumption of genetically engineered foods.

5. This is a translation of Heisenberg's original quote from 1958 which reads in its original: *Wir müssen uns daran erinnern, dass das, was wir beobachten, nicht die Natur selbst ist, sondern Natur, die unserer Art der Fragestellung ausgesetzt ist.*
6. The word ethology is derived from the Ancient Greek word *ethos* meaning character in this context and the ending *logía* meaning "the study of".
7. For those seriously interested in the archeology of this site there is a dedicated website that documents the excavations at Çatalhöyük: `www.catalhoyuk.com/`
8. Most of the time it was a "Him" and not a "Her". But there may have been more "Her" than we know today as it was not uncommon to write women out of history.
9. Given the pace in genetic research and more recently in genetic engineering that may not be true anymore because at some point in the future humans may well have developed the capability to reverse engineer genomes.
10. This is paraphrasing one of the pioneers of Victorian human studies, Francis Galton, Charles Darwin's half-cousin, who said: *"As a matter of history, our Anglo-Saxon civilization is only skin-deep"*.
11. There are certainly other and earlier examples but the one used here is likely the most well documented in our human history and one may say the most consequential one in its societal impacts.
12. Even though we clearly seem to enjoy seeing fellow-humans being hurt and sometimes hurt badly in all kinds of so-called martial art sports. In some way, the cage fights one can tune in today on television bring our societies much closer to Roman style entertainment for the masses than we have been for a long time.
13. For the *Uncertainty Principle* please see chapter note 101 for the chapter ***Gaining Perspective***.
14. Decimation was a tool used in the legions of the Roman Empire to enforce discipline among its troops, be it in the face of imminent defeat or to make it clear that losing a battle was just not an option. The punishment was dealt out randomly, but the end result was inevitably the same with one out of every ten soldiers having to be killed by his comrades.
15. The total death toll will never be fully known, and sources disagree on exact numbers but about 200,000 and 100,000 may have perished immediately or later because of the atomic bombs dropped on Hiroshima and Nagasaki, respectively. We are surely not surprised when we see what we call evil regimes having no problems with indiscriminately killing civilians, but it looks like war draws out the worst from all of us regardless to what moral virtues or ethics our governments officially profess otherwise.
16. Stanislav Y. Petrov (1939 – 2017), lieutenant colonel of the Soviet Air Defence Forces, judged that five apparently detected incoming American missiles where not real but a system malfunction and reported it as such to his superiors.
17. Vasili A. Arkhipov (1926 – 1998), deputy commander of a nuclear armed Soviet submarine during the Cuban missile crisis, withheld his consent as the only one of three officers who had to agree unanimously to authorize a nuclear launch.
18. That may be so with regard to many of the technical capabilities Jules Verne envisioned. However, if we define science fiction more broadly, then certainly the social utopias and their authors discussed earlier in the chapter do precede Jules Verne's visions of the future.
19. The other three non-Indo-European languages spoken in Europe, Estonian, Finnish, and Hungarian all belong to the Finno-Ugric language group and arrived in Europe with much later migrations from Eastern peoples into Europe. The Basque language truly stands alone as it is not related to any other language we know.
20. In Greek mythology Pandora was the first woman and her name translates into "the one who bears all gifts". When Prometheus stole fire from the Olympian Gods, Zeus created her to punish humankind which Prometheus had created from clay, seemingly restricting himself to making only men but not women. Pandora was sent to the brother of Prometheus, Epimetheus, to become his wife, carrying with her the proverbial box or more likely a jar, which the Gods told her never to open. Pandora could not resist to open the jar and all evils and plagues of humankind the Olympian Gods had put in there escaped; except for hope, which the Gods had put in there too. Hope remained in the jar when the will of Zeus had Pandora close it before it could escape too. This is a troubling story as it essentially pictures women as the source of all evil, insinuating that men would not have even hope on their side to

withstand evil. It looks like it was not only Near Eastern religions that pictured women as the source of all evil; the sons of Abraham could look back on a long tradition of paternalistic societies ascribing all evil to women.

Epilogue

1. This is just another example of human limitations in this case of the author growing up in a Western cultural context, practically imbibed from birth with Europe's and our World's classical cultural heritage. As a young boy yours truly was spell-bound by learning about Greek mythology and could not get enough of it. For that matter, one of the most treasured Christmas gifts I ever received was a translation of the stories of Greek mythology. However, nobody ever gave me a book about the mythologies of non-European people; except from what I could find in the public library bus. Therefore, it was only as a young adult that I really started to appreciate that there was so much else. I am still learning and for now not sure-footed enough in any other cultural tradition to include them in this book's narrative. My fault alone.
2. There is more than enough food to feed every hungry mouth in the world today, but many still go hungry and too many still die from male nourishment or hunger. So even when we have enough food to go around there is something else that is keeping us from sharing it in a way that nobody has to starve or die from hunger. Why is that so? That has nothing to do with our ability to grow enough food or with the availability of technology capable of distributing it. But it has all to do with a human condition that seems to prevent us in sharing with strangers the same way we would share with family. That holds true within and between countries or nations, rich or poor. It is primarily a spiritual challenge that we must overcome here to do the right thing and not an economic or technical one.

Media Credits

Prologue

Paintings:

Falling I	Painting by Brigitte M. Wurm, courtesy of Sevencatsgalerie, © Sevencatsgalerie LLC.

Gaining Perspective

Paintings:

Red Dots Leaving	Painting by Brigitte M. Wurm, courtesy of Sevencatsgalerie, © Sevencatsgalerie LLC.

Figures:

Fig. 1.1b	Image extracted from a photograph on Wikimedia Commons of the drawing in Peter Apian's *Cosmographia*.
Fig. 1.3a	Image extracted from a photograph on Wikimedia Commons of the drawing in *De revolutionibus orbium coelestium*, Biblioteka Jagiellońska Kraków, Poland.
Fig. 1.3b	Thomas Digges' diagram of the universe, courtesy of Wellcome Collection, Attribution 4.0 International (CC BY 4.0).

Portraits:

Nicolaus Copernicus	Portrait extracted from a photograph on Wikimedia Commons of a circa 1580 portrait painting by an unknown artist hanging in the town hall of Toruń, Poland.
Giordano Bruno	Portrait extracted from a photograph on Wikimedia Commons of a derivative modern illustration of Giordano Bruno after the 1578 published *"Livre du recteur"* (*The Rector's Book*), University of Geneva, Switzerland.

Galileo Galilei	Portrait extracted from a photograph on Wikimedia Commons of the 1636 portrait of Galilei by Justus Sustermans, Le Gallerie Delgi Uffizi, Florence, Italy.
Tycho Brahe	Portrait extracted from a photograph on Wikimedia Commons of an engraving by an unknown artists circa 1600.
Johannes Kepler	Portrait extracted from a photograph on Wikimedia Commons of a 1610 painting by an unknown artist, Observatory of the Benedictine Abbey of Kremsmünster, Austria.
Isaac Newton	Portrait extracted from a photograph on Wikimedia Commons of a 1689 painting by Godfrey Kneller, Private Collection.
Frederick William Herschel	Portrait extracted from a photograph on Wikimedia Commons of a 1785 portrait painting of Herschel by Lemuel Francis Abbott, National Portrait Gallery London, England.
Thomas Wright	Portrait taken from a photograph of a 1737 mezzotint - a print made from an engraved copper plate - by Thomas Frye, courtesy of Wellcome Collection, Attribution 4.0 International (CC BY 4.0).
Charles Messier	Portrait extracted from a photograph on Wikimedia Commons of a 1771 painting by Nicolas Ansiaume, Private Collection.
John Lewis Emil Dryer	Portrait extracted from a on Wikimedia Commons of a pre-1926 portrait picture taken by an unknown photographer.
Immanuel Kant	Portrait extracted from a photograph on Wikimedia Commons of a circa 1795 copper engraving by Johann Heinrich Lips, Staatsbibliothek zu Berlin, Germany.
Pierre-Simon Laplace	Portrait extracted from a photograph on Wikimedia Commons of an engraving published in *Gallery of Portraits*, London 1833.
Heber Doust Curtis	1914 portrait photograph of the astronomer courtesy of Allegheny Observatory Records, 1850-1967, UA.5.1, University Archives, Archives & Special Collections, University of Pittsburgh Library System.
Edwin P. Hubble	Portrait extracted from a photograph on Wikimedia Commons of a portrait taken by the photographer Johan Hagemeyer in 1931.

Henrietta Swan Leavitt	Portrait extracted from a image on Wikimedia Commons of a photo taken by Margaret Harwood, courtesy of American Institute of Physics Emilio Segré Visual Archives, Physics Today Collection, Shapley Collection.
Albert Einstein	A portrait of Einstein when he became professor in Zürich, ETH-Bibliothek Zürich, Bildarchiv / Fotograf: Langhans, Jan F. / Portr_05936 / .
Ejnar Hertzsprung	Photo courtesy of Yale University showing the astronomer as 1937 Bruce Medalist.
Henry Norris Russel	Portrait extracted from a image on Wikimedia Commons of a photo published in *The World's Work*, New York City 1921.
Arthur Eddington	Portrait extracted from a photograph on Wikimedia Commons from the George Grantham Bain Collection, Library of Congress Prints and Photographs Division Washington, D.C.
Cecilia H. Payne-Gaposchkin	Portrait extracted from a photograph on Flickr: The Commons, Acc. 90-105 - Science Service, Records, 1920s-1970s, Smithsonian Institution Archives.
Susan Jocelyn Bell Burnell	Portrait extracted from a 1967 photograph of Jocelyn Bell on Wikimedia Commons; Attribution-ShareAlike 2.0 Generic (CC BY-SA 2.0).
Karl Schwarzschild	Portrait extracted from a photograph on Wikimedia Commons of Schwarzschild in his garden.
Subrahmanyan Chandrasekhar	Portrait extracted from a photograph on Wikipedia showing Chandrasekhar in 1934 as Fellow of Trinity College, Cambridge, NASA Chandra X-ray Obeservatory.
Stephen Hawking	Portrait extracted from a photograph on Wikimedia Commons of Hawking at NASA's StarChild Learning Center, NASA.
Georges Lemaître	Portrait extracted from a photograph on Wikimedia Commons showing Lemaître in the 1930's.
Fritz Zwicky	Portrait of Fritz Zwicky at a 1959 astrophysics conference. Photo courtesy of California Institute of Technology Archives.
Vera C. Rubin	Portrait photo of Vera C. Rubin courtesy of Mark Godfrey.

Our Home - Planet Earth

Paintings:

Colliding Worlds	Painting by Brigitte M. Wurm, courtesy of Sevencatsgalerie, © Sevencatsgalerie LLC.

Figures:

Fig. 2.2	Adapted from tectonic map data on Wikimedia Commons, courtesy of the U.S. Geological Survey.
Blue Marble Earth	Figure extracted from a photograph on Wikimedia Commons taken on December 7, 1972 by the crew of Apollo-17.
Fig. 2.4	(a)-(e) modified and adapted from Ref. [12], (f) modified and adapted from a combination of public projections.
Fig. 2.7	Modified and adapted from data in Refs.[17–20].
Fig. 2.10	Adapted from a map published on the NASA website, courtesy of NASA/Robert Simmon, adapted from [21,22].

Portraits:

James Hutton	Portrait extracted from a photograph on Wikimedia Commons of a painting by Sir Henry Raeburn in 1776, Scottish National Portrait Gallery, Edinburgh, Scottland.
Charles Lyell	Portrait extracted from a photograph on Wikimedia Commons of an engraving by George J. Stoddart for the frontispiece of Volume II of the *Life Letters and Journals of Sir Charles Lyell, BART* by John Murray publishers in 1881.
Eduard Suess	Portrait extracted from a photograph by Peter Geymayer on Wikimedia Commons of a lithograph by Josef Kriehuber, circa 1869.
Alfred Wegener	Portrait extracted from a photograph on Wikimedia Commons of Wegener dated between 1924 and 1930, Bildarchiv Foto Marburg Aufnahme-Nr. 426.294.
Louis Agassiz	Portrait extracted from a photograph on Wikimedia Commons of Agassiz dated circa 1865, source unknown.
James Croll	Photo courtesy of Archives and Special Collections, University of Strathclyde Library.

Milutin Milanković	Portrait of Milanković in middle age taken from public a domain copy of N. Pejović, *Serbian And International Honors Awarded To Milutin Milanković*, Преглед НЦД 28 (2016), 79–86.

Life On Earth

Paintings:

Spring Life	Painting by Brigitte M. Wurm, courtesy of Sevencatsgalerie, © Sevencatsgalerie LLC.

Figures:

Fig. 3.2	Image extracted from a photograph on Wikimedia Commons of the drawing in *Generelle Morphologie der Organismen*.
Fig. 3.4	Figure reproduced from Ref. [39] with permission from The American Association for the Advancement of Science.
Fig. 3.6a	Photograph taken by Heidi Allen, image courtesy of the Geological Survey and Resource Strategy, Department of Mines, Industry Regulation and Safety. © State of Western Australia 2019.
Fig. 3.6b	Photograph taken by Arthur Mory, image courtesy of the Geological Survey and Resource Strategy, Department of Mines, Industry Regulation and Safety. © State of Western Australia 2019.
Fig. 3.11	Modified and adapted from Ref. [12].
Fig. 3.12	Photograph by John Reader, image no. E437/0018 Science Photo Library, © John Reader / Science Photo Library.

Geological Period Temperature and CO_2 graphs:

The Cambrian, Ordovician, Silurian, Devonian, Carboniferous, Permian, Triassic, Jurassic, Cretaceous, and Neogene period and the Pleistocene and Holocene epoch graphs show details taken from the data shown in fig. 2.11 and fig. 2.12.

Human Fossil Pictures:

Sahelanthropus tschadensis	Black and white adaption of a photograph courtesy of the Human Origins Program, Smithsonian Institution.
Ardipithecus ramidus	Image extracted and modified from a photograph by T. Michael Keesey after Ref. [51] on Wikimedia Commons; Attribution 2.0 Generic (CC BY 2.0).

Kenyanthropus platyops	Image extracted and modified from a photograph by Pavel Švejnar on Wikimedia Commons of a skull cast at the Moravian Museum – Pavilion Anthropos, Brno, Czech Republic; Attribution-ShareAlike 4.0 International (CC BY-SA 4.0).
Australopithecus africanus	Black and white adaption of a photograph courtesy of the Human Origins Program, Smithsonian Institution.
Paranthropus boisei	Black and white adaption of a photograph courtesy of the Human Origins Program, Smithsonian Institution.
Homo habilis	Black and white adaption of a photograph courtesy of the Human Origins Program, Smithsonian Institution.
Homo erectus ergaster	Black and white adaption of a photograph courtesy of the Human Origins Program, Smithsonian Institution.
Homo heidelbergensis	Black and white adaption of a photograph courtesy of the Human Origins Program, Smithsonian Institution.
Homo neanderthalensis	Black and white adaption of a photograph courtesy of the Human Origins Program, Smithsonian Institution.
Homo sapiens	Black and white adaption of a photograph courtesy of the Human Origins Program, Smithsonian Institution.

Portraits:

Andreas Vesalius	Portrait extracted from a photograph on Wikimedia Commons of a posthumous portrait painted by Pierre Poncet, Musee des Beaux-Arts, Orleans, France.
William Harvey	Portrait derived from a photograph on Wikimedia Commons of a circa 1627 portrait attributed to Daniël Mijtens, National Portrait Gallery London, England.
Antonie van Leeuwenhoek	Portrait derived from a photograph on Wikimedia Commons of a portrait by Jan Verkolje circa 1680-1686, Rijksmuseum Amsterdam, The Netherlands.
Maria Sibylla Merian	Portrait derived from a photograph on Wikimedia Commons of a 1679 portrait by Jacob Marrel, Merian's step-father, Kunstmuseum Basel, Switzerland.

Antoine-Laurent de Lavoisier	Portrait extracted from a photograph on Wikimedia Commons of a 1788 painting by Jacques-Louis David, The Metropolitan Museum of Art, New York City.
Carl Linnaeus	Portrait derived from a photograph on Wikimedia Commons of a 1739 painting by Johan Henrik Scheffel, Uppsala University Library, Sweden.
Jean-Baptiste Lamarck	Portrait taken from a photograph on Wikimedia Commons of a lithographic portrait by Jules Pizzetta published in Pizetta J., Galerie des Naturalistes, Histoire des sciences naturelles depuis leur origine jusqu'a nos jours, 2 edit. Paris, Hennuyer, 1893.
Charles Darwin	Portrait derived from a photograph on Wikimedia Commons of Darwin from 1877 by Lock & Whitfield, National Portrait Gallery London, England.
Ernst Haeckel	Portrait extracted from a photograph on Wikimedia Commons taken by an unknown photographer, U.S. National Library of Medicine Digital Collections.
Gregor Mendel	Portrait taken from a photograph on Wikimedia Commons, source unknown.
Thomas Hunt Morgan	Portrait extracted from a photograph on Wikimedia Commons showing a portrait of Morgan taken from the 1891 Johns Hopkins yearbook, U.S. National Library of Medicine Digital Collection.
Rosalind Franklin	Portrait extracted from a on flickr Flickr: The Commons, source unknown.
James Watson	Portrait extracted from a color photograph on Wikimedia Commons of James D. Watson, date unknown.
Francis Crick	Portrait extracted from a color photograph on Wikimedia Commons of Francis Crick taken 1981, Attribution 4.0 International (CC BY 4.0).
Maurice Wilkens	Portrait extracted from a photograph on Wikimedia Commons of Maurice Wilkins in his laboratory, date unknown.
Carl R. Woese	Photo courtesy of Charlie Vossbrinck showing Carl R. Woese around 1982.
Lynn Margulis	Photo courtesy of the National Science & Technology Medals Foundation showing Lynn Margulis in 1999 as the recipient for the National Medal of Sciences in Biological Sciences.

The Human Endeavor

Paintings:

Last Rays — Painting by Brigitte M. Wurm, courtesy of Sevencatsgalerie, © Sevencatsgalerie LLC.

Figures:

Fig. 4.4 — Modified and adapted from a combination of public projections.

Epilogue

Paintings:

Rising Sun — Painting by Brigitte M. Wurm, courtesy of Sevencatsgalerie, © Sevencatsgalerie LLC.

Figures:

Earthrise — Figure extracted from a photograph on Wikimedia Commons taken on December 24, 1968 by Apollo-8 astronaut William Anders.

List of Abbreviations

Acronyms

ATP	Adenosine Triphosphate
BCE	Before Current Era
BIF	Banded Iron Formation
CDS	Centre de Données astronomiques de Strasbourg
CE	Current Era
CERN	Conseil européen pour la recherche nucléaire
CFC	Chlorofluorocarbon
DNA	Deoxyribonucleic Acid
GOE	Great Oxygenation Event
GPS	Global Positioning System
GWA	Genome-Wide Association
KBO	Kuiper Belt Object
LHB	Late Heavy Bombardment
LUCA	Last Universal Common Ancestor
NASA	National Aeronautics and Space Administration
NGC	New General Catalogue (of Nebulae and Clusters of Stars)
RNA	Ribonucleic Acid
SSSB	Small Solar System Body
UV	Ultraviolet

Symbols

	Creative Commons license symbol for Attribution (BY)
	Creative Commons license symbol
flickr	Flickr: The Commons website
	flickr No known copyright restrictions
	Creative Commons Public Domain Mark
	Creative Commons license symbol for Share-alike (SA)
	Wikipedia website
	Wikimedia Commons website

Metrics

°C	Celsius
AU	Astronomical Unit
cm^3	Cubic centimeter
g/cm^3	Gram per cubic centimeter
GPa	Giga-Pascal
K	Kelvin
km/s	Kilometer per second
pc	Parallax-second
ppm	Parts per million
µs	Microsecond

Measure & Number Conventions

Million	1,000,000 or 10^6
Billion	1,000,000,000 or 10^9
Trillion	1,000,000,000,000 or 10^{12}
Millennium	A period of thousand years
Millennia	Several periods of thousand years
Nanometer	One billionth of a meter
Micrometer	One millionth of a meter
Millimeter	One thousandth of a meter
Centimeter	One hundredth of a meter
Kilometer	One thousand meters
Nanosecond	One billionth of a second
Microsecond	One millionth of a second
Millisecond	One thousandth of a second

References

[1] Jet Propulsion Laboratory - California Institute of Technology. *Voyager Mission Status*. https://voyager.jpl.nasa.gov/mission/status/. 2019 (cit. on p. 38).

[2] NASA. *New Horizons*. https://www.nasa.gov/mission_pages/newhorizons/main/index.html. 2019 (cit. on p. 40).

[3] Albert Einstein. "Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt". In: *Annalen der Physik* 17.6 (1905), pp. 132–148. DOI: 10.1002/andp.19053220607 (cit. on p. 47).

[4] Albert Einstein. "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen". In: *Annalen der Physik* 17.8 (1905), pp. 549–560. DOI: 10.1002/andp.19053220806 (cit. on p. 47).

[5] Albert Einstein. "Zur Elektrodynamik bewegter Körper". In: *Annalen der Physik* 17.10 (1905), pp. 891–921. DOI: 10.1002/andp.19053221004 (cit. on p. 47).

[6] Albert Einstein. "Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?" In: *Annalen der Physik* 18.13 (1905), pp. 891–921. DOI: 10.1002/andp.19053231314 (cit. on p. 47).

[7] M. Wenger et al. "The SIMBAD astronomical database." In: *Astronomy and Astrophysics Supplement* 143.1 (2000), pp. 9–22. DOI: 10.1051/aas:2000332 (cit. on p. 64).

[8] Centre de Données astronomiques de Strasbourg. *The SIMBAD astronomical database*. http://simbad.u-strasbg.fr/simbad/. 2018 (cit. on p. 64).

[9] Arthur Stanley Eddington. "The Internal Constitution of the Stars". In: *The Scientific Monthly* 11.4 (1920). Source: The American Association for the Advancment of Science, https://www.jstor.org/stable/6491, pp. 297–303 (cit. on pp. 66, 67).

[10] Michiel Lambrechts and Anders Johansen. "Rapid growth of gas-giant cores by pebble accretion". In: *Astronomy & Astrophysics* A32 (2012), pp. 1–13. DOI: 10.1051/0004-6361/201219127 (cit. on p. 125).

[11] Anders Johansen and Michiel Lambrechts. "Forming Planets via Pebble Accretion". In: *Annual Review of Earth and Planetary Sciences* 45.1 (2017), pp. 359–387. DOI: 10.1146/annurev-earth-063016-020226 (cit. on p. 125).

[12] Wolfgang Frisch, Martin Meschede, and Ronald Blakey. *Plate Tectonics – Continental Drift and Mountain Building*. Berlin: Springer Verlag, 2011 (cit. on pp. 137, 277, 446, 447).

[13] Kenji Ohta et al. "Experimental determination of the electrical resistivity of iron at Earth's core conditions". In: *Nature* 534 (2016), pp. 95–98. DOI: 10.1038/nature17957 (cit. on p. 141).

[14] Zuzana Konôpková et al. "Direct measurement of thermal conductivity in solid iron at planetary core conditions". In: *Nature* 534 (2016), pp. 99–101. DOI: 10.1038/nature18009 (cit. on p. 141).

[15] Richard K. Bono et al. "Experimental determination of the electrical resistivity of iron at Earth's core conditions". In: *Nature Geoscience* 12 (2 2019), pp. 143–147. DOI: 10.1038/s41561-018-0288-0 (cit. on p. 141).

[16] Peter Driscoll. "Geodynamo recharged". In: *Nature Geoscience* 12 (2 2019), pp. 83–84. DOI: 10.1038/s41561-019-0301-2 (cit. on p. 141).

[17] J. Imbrie et al. "On the structure and origin of major glaciation cycles 2. The 100,000-year cycle". In: *Paleoceanography and Paleoclimatology* 8.6 (1993), pp. 699–735. DOI: 10.1029/93PA02751 (cit. on pp. 156, 157, 446).

[18] André Berger and M.F. Loutre. "Insolation values for the climate of the last 10 million years". In: *Quaternary Science Reviews* 10 (1991), pp. 297–317. DOI: 10.1016/0277-3791(91)90033-Q (cit. on pp. 156, 157, 446).

[19] M. Crucifix, M. F. Loutre, and André Berger. "The Climate Response to the Astronomical Forcing". In: (2006). Ed. by Y. Calisesi et al., pp. 213–226. DOI: 10.1007/978-0-387-48341-2_17 (cit. on pp. 156, 157, 446).

[20] Qui Zhen Yin and André Berger. "Individual contribution of insolation and CO2 to the interglacial climates of the past 800,000 years". In: *Climate Dynamics* 38 (2012), pp. 709–724. DOI: 10.1007/s00382-011-1013-5 (cit. on pp. 156, 157, 446).

[21] J.T. Houghton et al., eds. *Climate Change 2001: The Scientific Basis*. Cambridge: Cambridge University Press, 2001 (cit. on pp. 166, 446).

[22] Stefan Rahmstorf. "Ocean circulation and climate during the past 120,000 years". In: *Nature* 419 (2002), pp. 207–214. DOI: 10.1038/nature01090ID (cit. on pp. 166, 446).

[23] J. Veizer et al. "^{87}Sr/^{86}Sr, Δ^{13}C and Δ^{18}O evolution of Phanerozoic seawater". In: *Chemical Geology* 161 (1999), pp. 59–88. DOI: 10.1016/S0009-2541(99)00081-9 (cit. on pp. 175, 177).

[24] Dana. L. Royer et al. "CO_2 as a primary driver of Phanerozoic climate". In: *Geological Society of America Today* 14.3 (2004), pp. 4–10. DOI: 10.1130/1052-5173(2004)014<4:CAAPDO>2.0.CO;2 (cit. on pp. 175, 177).

[25] James Hansen et al. "Climate sensitivity, sea level and atmospheric carbon dioxide". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371 (2013). DOI: 10.1098/rsta.2012.0294 (cit. on pp. 175, 177).

[26] James C. Zachos, Gerald R. Dickens, and Richard E. Zeebe. “An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics”. In: *Nature* 451 (2008), pp. 279–283. DOI: 10.1038/nature06588 (cit. on pp. 175, 177).

[27] Lorraine E. Lisiecki and Maureen E. Raymo. “A Pliocene-Pleistocene stack of 57 globally distributed benthic Δ^{18}O records”. In: *Paleoceanography and Paleoclimatology* 20.1,2 (2005). Correction for Figure 7 published in issue 2; https://doi.org/10.1029/2005PA001164. DOI: 10.1029/2004PA001071 (cit. on pp. 175, 177, 178).

[28] Jean Jouzel and Valerie Masson-Delmotte. *EPICA Dome C Ice Core 800KYr deuterium data and temperature estimates*. Data set. Supplement to: Jouzel, Jean; Masson-Delmotte, Valerie; Cattani, Olivier; Dreyfus, Gabrielle; Falourd, Sonia; Hoffmann, G; Minster, B; Nouet, J; Barnola, Jean-Marc; Chappellaz, Jérôme A; Fischer, Hubertus; Gallet, J C; Johnsen, Sigfus J; Leuenberger, Markus C; Loulergue, Laetitia; Luethi, D; Oerter, Hans; Parrenin, Frédéric; Raisbeck, Grant M; Raynaud, Dominique; Schilt, Adrian; Schwander, Jakob; Selmo, Enrico; Souchez, Roland; Spahni, Renato; Stauffer, Bernhard; Steffensen, Jørgen Peder; Stenni, Barbara; Stocker, Thomas F; Tison, Jean-Louis; Werner, Martin; Wolff, Eric W (2007): Orbital and millennial Antarctic climate variability over the past 800,000 years. Science, 317(5839), 793-797, https://doi.org/10.1126/science.1141038. 2007. DOI: 10.1594/PANGAEA.683655. URL: https://doi.org/10.1594/PANGAEA.683655 (cit. on pp. 175, 177, 178).

[29] Gavin L. Foster, Dana L. Royer, and Daniel J. Lunt. “Future climate forcing potentially without precedent in the last 420 million years”. In: *Nature Communications* 8 (2017), p. 14845. DOI: 10.1038/ncomms14845 (cit. on pp. 175, 177, 178).

[30] Bernhard Bereiter et al. “Revision of the EPICA Dome C CO_2 record from 800 to 600 kyr before present”. In: *Geophysical Research Letters* 4.2 (2015), pp. 542–549. DOI: 10.1002/2014GL061957 (cit. on pp. 175, 177, 178).

[31] Jean R. Petit et al. “Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica”. In: *Nature* 399 (1999), pp. 429–436. DOI: 10.1038/20859 (cit. on pp. 175, 177, 178).

[32] U.S. Department of Commerce National Oceanic & Atmospheric Administration, NOAA Research. *Earth System Research Laboratory Global Monitoring Division Mauna Loa, Hawai*. https://www.esrl.noaa.gov/gmd/ccgg/trends/global.html. 2018 (cit. on pp. 175, 177, 178).

[33] Fred Griffith. “The Significance of Pneumococal Types”. In: *The Journal of Hygiene* 27.2 (1928). Reprinted in "Reviews of Infectious Diseases", Volume 3, Issue 2, March 1981, Pages 372–395, https://doi.org/10.1093/clinids/3.2.372, pp. 113–159 (cit. on p. 227).

[34] Ron Sender, Shai Fuchs, and Ron Milo. “Revised Estimates for the Number of Human and Bacteria Cells in the Body”. In: *Public Library of Science - Biology* 14.8 (2016), e1002533. DOI: 10.1371/journal.pbio.1002533 (cit. on p. 228).

[35] Alastair Crisp et al. “Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes”. In: *Genome Biology* 16 (50 2015). DOI: 10.1186/s13059-015-0607-3 (cit. on p. 229).

[36] Steven L. Salzberg. “Horizontal gene transfer is not a hallmark of the human genome”. In: *Genome Biology* 18:85 (2017). DOI: 10.1186/s13059-017-1214-2 (cit. on p. 229).

[37] Christopher A. Edmonds, Anita S. Lillie, and L. Luca Cavalli-Sforza. “Mutations arising in the wave front of an expanding population”. In: *Proceedings of the National Academy of Sciences of the United States of America* 101.4 (2004), pp. 975–979. DOI: 10.1073/pnas.0308064100 (cit. on pp. 234, 406).

[38] Carl R. Woese, Otto Kandler, and Mark L. Wheelis. “Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya”. In: *Proceedings of the National Academy of Sciences of the United States of America* 87.12 (1990), pp. 4576–4579. DOI: 10.1073/pnas.87.12.4576 (cit. on p. 236).

[39] W. Ford Doolittle. “Phylogenetic Classification and the Universal Tree”. In: *Science* 284.5423 (1999), pp. 2124–2128. DOI: 10.1126/science.284.5423.2124 (cit. on pp. 237, 447).

[40] Geological Survey of Western Australia (GSWA). *Geology and geophysics*. http://www.dmp.wa.gov.au/Geological-Survey/Geological-Survey-262.aspx. 2019 (cit. on p. 240).

[41] David Wacey et al. “Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia”. In: *Nature Geoscience* 4 (2011), pp. 698–702. DOI: 10.1038/ngeo1238 (cit. on p. 241).

[42] Heinrich D. Holland. “The oxygenation of the atmosphere and oceans”. In: *Philosophical Transactions of The Royal Society B* 361 (2006), pp. 903–915. DOI: 10.1098/rstb.2006.1838 (cit. on pp. 248, 249).

[43] Bilal U. Haq, Jan Hardenbol, and Peter R. Vail. “Chronology of Fluctuating Sea Levels Since The Triassic”. In: *Science* 235.4739 (1987), pp. 1156–1167. DOI: 10.1126/science.235.4793.1156 (cit. on p. 257).

[44] Anthony Hallam. *Phanerozoic Sea-Level Changes*. New York: Columbia University Press, 1992 (cit. on p. 257).

[45] John A. Church and Neil J. White. “Sea-Level Rise from the Late 19th to the Early 21st Century”. In: *Surveys in Geophysics* 32.4-5 (2011), pp. 585–602. DOI: 10.1007/s10712-011-9119-1 (cit. on p. 258).

[46] Rachel M. Spratt and Lorraine E. Lisiecki. “A Late Pleistocene sea level stack”. In: *Climate of the Past* 12.4 (2016), pp. 1079–1092. DOI: 10.5194/cp-12-1079-2016 (cit. on p. 258).

[47] Tomasz Sulej and Grzegorz Niedźwiedzki. “An elephant-sized Late Triassic synapsid with erect limbs”. In: *Science* 363.6422 (2018), pp. 78–80. DOI: 10.1126/science.aal4853 (cit. on p. 287).

[48] Gretchen Vogel. "Giant mammal cousin rivaled early dinosaurs". In: *Science* 362.6417 (2018), p. 879. DOI: `10.1126/science.362.6417.879` (cit. on p. 287).

[49] Rui Diogo, Julia L. Molnar, and Bernard Wood. "Bonobo anatomy reveals stasis and mosaicism in chimpanzee evolution, and supports bonobos as the most appropriate extant model for the common ancestor of chimpanzees and humans". In: *Scientific Reports* 7:608 (2017), pp. 1–8. DOI: `10.1038/s41598-017-00548-3` (cit. on p. 291).

[50] Michel Brunet et al. "A new hominid from the Upper Miocene of Chad, Central Africa". In: *Nature* 418 (2002), pp. 145–151. DOI: `10.1038/nature00879` (cit. on p. 298).

[51] Gen Suwa et al. "The *Ardipithecus ramidus* Skull and Its Implications for Hominid Origins". In: *Science* 326 (2009), 68–68e7. DOI: `10.1126/science.1175825` (cit. on pp. 299, 447).

[52] Sonia Harmand et al. "3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya". In: *Nature* 521 (2015), pp. 310–315. DOI: `10.1038/nature14464` (cit. on pp. 300, 328).

[53] Colin Barras. "'Little Foot' hominin emerges from stone after millions of years". In: *Nature* 564 (2018), pp. 169–170. DOI: `10.1038/d41586-018-07651-z` (cit. on p. 303).

[54] Katerina Harvati et al. "Apidima Cave fossils provide earliest evidence of Homo sapiens in Eurasia". In: *Nature* 571 (2019), pp. 500–504. DOI: `10.1038/s41586-019-1376-z` (cit. on pp. 308, 311, 323, 330).

[55] Matthias Meyer et al. "Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins". In: *Nature* 531 (2016), pp. 504–507. DOI: `10.1038/nature17405` (cit. on p. 309).

[56] Jean-Jacques Hublin et al. "New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens". In: *Nature* 546 (2017), pp. 289–292. DOI: `10.1038/nature22336` (cit. on p. 313).

[57] Daniel Richter et al. "The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age". In: *Nature* 546 (2017), pp. 293–296. DOI: `10.1038/nature22335` (cit. on p. 313).

[58] Israel Hershkovitz et al. "The earliest modern humans outside Africa". In: *Science* 359.6374 (2018), pp. 456–459. DOI: `10.1126/science.aap8369` (cit. on pp. 323, 330).

[59] Shannon P. McPherron et al. "Evidence for stone-tool-assisted consumption of animal tissues before 3.39 million years ago at Dikika, Ethiopia". In: *Nature* 466 (2010), pp. 857–860. DOI: `10.1038/nature09248` (cit. on p. 328).

[60] Yue Hu et al. "Late Middle Pleistocene Levallois stone-tool technology in southwest China". In: *Nature* 565 (2019), pp. 82–85. DOI: `10.1038/s41586-018-0710-1` (cit. on pp. 330, 331).

[61] Ewen Callaway. "Oldest *Homo sapiens* fossil claim rewrites our species' history". In: *Nature News* (2017). DOI: `10.1038/nature.2017.22114` (cit. on p. 330).

[62] Chris Stringer and Julia Galway-Witham. "On the origin of our species". In: *Nature* 546 (2017), pp. 212–214. DOI: `10.1038/546212a` (cit. on p. 330).

[63] Francesco Berna et al. "Microstratigraphic evidence of in situ fire in the Acheulean strata of Wonderwerk Cave, Northern Cape province, South Africa". In: *Proceedings of the National Academy of Sciences of the United States of America* 109.20 (2012), E1215–E1220. DOI: `10.1073/pnas.1117620109` (cit. on p. 333).

[64] Matt Kaplan. "Million-year-old ash hints at origins of cooking". In: *Nature News* (2012). DOI: `doi:10.1038/nature.2012.10372` (cit. on p. 333).

[65] Naama Goren-Inbar et al. "Evidence of Hominin Control of Fire at Gesher Benot Ya'aqov, Israel". In: *Science* 304.5671 (2004), pp. 725–727. DOI: `10.1126/science.1095443` (cit. on p. 333).

[66] Melissa A. Toups et al. "Origin of Clothing Lice Indicates Early Clothing Use by Anatomically Modern Humans in Africa". In: *Molecular Biology and Evolution* 28.1 (2011), pp. 29–32. DOI: `10.1093/molbev/msq234` (cit. on p. 335).

[67] Christopher S. Henshilwood et al. "A 100,000-Year-Old Ochre-Processing Workshop at Blombos Cave, South Africa". In: *Science* 334.6053 (2011), pp. 219–222. DOI: `10.1126/science.1211535` (cit. on p. 335).

[68] Kevin Neville Laland, Nicola Atton, and Michael Munro Webster. "From fish to fashion: Experimental and theoretical insights into the evolution of culture". In: *Proceedings of the Royal Society B: Biological Sciences* 366.1567 (2011), pp. 958–968. DOI: `10.1098/rstb.2010.0328` (cit. on p. 344).

[69] Kevin N. Laland. *Darwin's Unfinished Symphony - How Culture Made The Human Mind*. New Jersey: Princeton University Press, 2017 (cit. on pp. 344, 346).

[70] Johannes Krause et al. "The derived FOXP2 variant of modern humans was shared with Neandertals". In: *Current Biology* 17.21 (2007), pp. 1908–1912. DOI: `10.1016/j.cub.2007.10.008` (cit. on p. 349).

[71] David W. Anthony and Jennifer Y. Chi, eds. *The Lost World Of Old Europe: The Danube Valley, 5000 3500 BC*. Princeton, New Jersey: Princeton University Press, 2009 (cit. on p. 353).

[72] Klaus Schmidt. "Göbekli Tepe – the Stone Age Sanctuaries. New results of ongoing excavations with a special focus on sculptures and high reliefs". In: *Documenta Praehistorica* XXXVII (2010), pp. 239–256. DOI: `10.4312/dp.37.21` (cit. on p. 354).

[73] Michael Balter. *The Goddess and the Bull: Çatalhöyük - An Archaeological Journey to the Dawn of Civilization*. London: Routledge, 2016 (cit. on p. 355).

[74] *Extreme Poverty — Data Documentation*. `https://www.gapminder.org/data/documentation/epovrate/`. 2018 (cit. on p. 380).

[75] *Life expectancy at birth.* `https://www.gapminder.org/data/documentation/gd004/`. 2018 (cit. on p. 380).

[76] Amy Goldberg et al. “Microstratigraphic evidence of in situ fire in the Acheulean strata of Wonderwerk Cave, Northern Cape province, South Africa”. In: *Proceedings of the National Academy of Sciences of the United States of America* 114.10 (2017), pp. 2657–2662. DOI: `10.1073/pnas.1616392114` (cit. on p. 405).

[77] Daniel Shriner and Charles N. Rotimi. “Whole-Genome-Sequence-Based Haplotypes Reveal Single Origin of the Sickle Allele during the Holocene Wet Phase”. In: *The American Journal of Human Genetics* 101 (2018), pp. 547–556. DOI: `10.1016/j.ajhg.2018.02.003` (cit. on p. 408).

[78] Andrew Curry. “The Milk Revolution”. In: *Nature* 500 (2013), pp. 20–22. DOI: `10.1038/500020a` (cit. on p. 410).

Bibliography

Science Texts

Felix V. Kaminsky. *The Earth's Lower Mantle - Composition and Structure*. Basel, Switzerland: Springer International Publishing, 2017.

Hidenori Terasaki and Rebecca A. Fischer, eds. *Deep Earth - Physics and Chemistry of the outer Mantle and Core*. Hoboken, New Jersey: American Geophsical Union and John Wiley & Sons, 2016.

Raymond S. Bradley. *Paleoclimatology – Reconstructing Climates of the Quaternary*. Waltham, Massachussets: Academic Press, 2015.

Michael E. Mann and Lee R. Kump. *Dire Predictions – Understanding Climate Change. The Visual Guide to the Findings of the IPCC*. Second American edition. New York: DK Publishing, 2015.

Charles H. Langmuir and Wally Broecker. *How To Build A Habitable Planet - The Story Of Earth From The Big Bang To Humankind*. Princeton, New Jersey: Princeton University Press, 2012.

Wolfgang Frisch, Martin Meschede, and Ronald Blakey. *Plate Tectonics – Continental Drift and Mountain Building*. Berlin: Springer Verlag, 2011.

Woodruff T. Sullivan III and John A. Baross, eds. *Planets And Life –The Emerging Science Of Astrobiology*. New York: Cambridge University Press, 2007.

Paul R. Weissman, Lucy-Ann McFadden, and Torrence V. Johnson, eds. *Encyclopedia of the Solar System*. San Diego, California: Academic Press, 2003.

G. Brent Dalrymple. *The Age of the Earth*. Stanford, California: Stanford University Press, 2001.

Ron Redfern. *Origins - The Evolution Of Continents, Oceans And Life*. Norman, Oklahoma: University of Oklahoma Press, 2001.

Thomas M. Cronin. *Principles of Paleoclimatology*. New York: Columbia University Press, 1999.

Michael Zellik. *Astronomy - The Evolving Universe*. 8th edition. New York: John Wiley & Sons, 1997.

Anthony Hallam. *Phanerozoic Sea-Level Changes*. New York: Columbia University Press, 1992.

John Imbrie and Katherine Palmer Imbrie. *Ice Ages –Solving The Mystery*. Cambridge, Massachusetts: Harvard University Press, 1986.

Albert Einstein. *Über die spezielle und die allgemeine Relativitätstheorie*. Braunschweig, Germany: Friedrich Vieweg & Sohn, 1981.

Popular Science

Michael Tomasello. *Becoming Human - A Theory of Ontogeny*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 2019.

Robert Plomin. *Blueprint - How DNA Makes Us Who We Are*. Cambridge, Massachusetts: The MIT Press, 2018.

David Quammen. *The Tangled Tree - A Radical New History of Life*. New York: Simon & Schuster, 2018.

David Reich. *Who We Are and How We Got Here - Acient DNA and the New Science of the Human Past*. 1st edition. New York: Pantheon Books, 2018.

Peter Ward. *Lamarck's Revenge - How Epigentics Is Revolutionizing Our Understanding of Evolution's Past and Present*. New York: Bloomsbury Publishing, 2018.

Wallace Arthur. *Life Through Time And Space*. Cambridge, Massachusetts: Harvard University Press, 2017.

John Chambers and Jacqueline Mitton. *From Dust To Life - The Origin And Evolution Of Our Solar System*. Princeton, New Jersey: Princeton University Press, 2017.

Kevin N. Laland. *Darwin's Unfinished Symphony - How Culture Made The Human Mind*. New Jersey: Princeton University Press, 2017.

Jonathan B. Losos. *Improbable Destinies - Fate, Chance, and the Future of Evolution*. New York: Riverhead Books, 2017.

Edward O. Wilson. *The Diversity Of Life*. Cambridge, Massachusetts: Harvard University Press, 2017.

Michael Balter. *The Goddess and the Bull: Çatalhöyük - An Archaeological Journey to the Dawn of Civilization*. London: Routledge, 2016.

Marcus du. Sautoy. *The Great Unknown –Seven Journeys To The Frontiers Of Science*. New York: Viking, 2016.

Frans de Waal. *Are We Smart Enough To Know How Smart Animals Are?* New York: W.W. Norton & Company, 2016.

Jim Baggott. *Origins - The Scientific Story of Creation*. Oxford: Oxford University Press, 2015.

Peter Ward and Joe Kirschvink. *A New History of Life - The Radical New Discoveries About The Origins And Evolution Of Life On Earth*. New York: Bloomsbury Press, 2015.

Clive Finlayson. *The Improbable Primate - How Water Shaped Human Evolution*. Oxford: Oxford University Press, 2014.

Svante Pääbo. *Neanderthal Man - In Search of Lost Genomes*. New York: Basic Books, 2014.

Dimitria Papagianni and Michael A. More. *The Neanderthals Rediscovered - How Modern Science Is Rewriting Their Story*. New York: Thames & Hudson, 2013.

Martin Brasier. *Secret Chambers - The Inside Story of Cells & Complex Life*. Oxford: Oxford University Press, 2012.

Robert M. Hazen. *The Story Of Earth - The First 4.5 Billion Years, From Stardust To Living Planet*. New York: Viking Penguin, 2012.

Addy Pross. *What is Life? How Chemistry Becomes Biology*. Oxford: Oxford University Press, 2012.

Chris Stringer. *The Origin of Our Species*. London: Penguin Books, 2012.

Martin Meredith. *Born in Africa - The Quest for the Origins of Human Life*. New York: PublicAffairs, 2011.

Martin Brasier. *Darwin's Lost World - The Hidden History Of Animal Life*. Oxford: Oxford University Press, 2009.

Gregory Cochran and Henry Harpending. *The 10,000 Year Explosion - How Civilization Accelerated Human Evolution*. New York: Basic Books, 2009.

Ted Nield. *Supercontinent - Ten Billion Years In The Life Of Our Planet*. Cambridge, Massachusetts: Harvard University Press, 2009.

Andrew H. Knoll. *Life On A Young Planet - The First Three Billion Years Of Evolution on Earth*. Princeton, New Jersey: Princeton University Press, 2003.

William James Burroughs. *Climate Change - A Multidisciplinary Approach*. New York: Cambridge Uinversity Press, 2001.

Luigi Luca Cavalli-Sforza. *Genes, Peoples, and Languages*. Berkeley, California: University of California Press, 2001.

Ernst Mayr. *Das ist Biologie - Die Wissenschaft des Lebens*. Heidelberg: Spektrum Akademischer Verlag, 2000.

Kipp S. Thorne. *Black Holes & Time Warps - Einstein's Outrageous Legacy*. New York: W.W. Norton & Company, 1994.

Rudolf Kippenhahn. *100 Billion Suns –The Birth, Life, and Death of the Stars*. Princeton, New Jersey: Princeton University Press, 1993.

Carl Sagan. *Cosmos*. New York: Random House, 1980.

Steven Weinberg. *Die esrten drei Minuten - Der Ursprung des Universums*. München, Germany: Deutscher Taschenbuch Verlag, 1980.

Historical Texts

Erwin Schrödinger. *What is Life? –Mind and Matter*. Cambridge: Cambridge University Press, 2018.

Herodotus. *The Histories*. Trans. by Tom Holland. New York: Viking, 2014.

David W. Anthony and Jennifer Y. Chi, eds. *The Lost World Of Old Europe: The Danube Valley, 5000 - 3500 BC*. Princeton, New Jersey: Princeton University Press, 2009.

Naomi Oreskes. *Plate Tectonics - An Insider's History of the Modern Theory of the Earth*. Boulder, Colorado: Westview Press, 2003.

Jürgen Hamel. *Geschichte der Astronomie*. Basel, Switzerland: Birkhäuser Verlag, 1998.

Galileo Galilei. *Dialog über die beiden hauptsächlichen Weltsysteme, das ptolemäische und das kopernikanische*. German. Ed. by Roman Sexl and Karl von Meyenn. Trans. Italian by E. Strauss. Stuttgart, Germany: B.G. Teubner, 1982.

Cultural Studies

Hans Rosling, Ola Rosling, and Anna Rosling Rönnlund. *Factfullness: Ten Reasons We're Wrong About the World - and Why Things Are Better Than You Think*. 1st edition. New York: Flatiron Books, 2018.

Karin Bojs. *My European Family - The First 54,000 Years*. London: Bloomsbury Publishing, 2017.

Ernst Peter Fischer. *Eine Geschichte der Weltbilder*. Berlin: Rowohlt, 2017.

Jürgen Kaube. *Die Anfänge von Allem*. Berlin: Rowohlt, 2017.

Max Tegmark. *Life 3.0 - Being Human In The Age Of Artificial Intelligence*. New York: Alfred A. Knopf, 2017.

Yuval Noah Harari. *Eine kurze Geschichte der Menschheit*. München: Deutsche Verlags-Anstalt, 2013.

Marija Gimbutas. *The Living Goddesses*. Berkeley, CA: University of California Press, 2001.

Biographies

John L. Heilbron. *Galileo*. Oxford: Oxford University Press, 2010.

Ian Shine and Sylvia Wrobel. *Thomas Hunt Morgan - Pioneer of Genetics*. Lexington, KY: The University Press of Kentucky, 2009.

Janet Browne. *Charles Darwin: A Biography, Vol. 2 - The Power of Place*. New York: Alfred A. Knopf, 2002.

Michael White. *Isaac Newton - The Last Sorcerer*. Reading, Massachusetts: Perseus Books, 1999.

Gale E. Christianson. *Edwin Hubble –Mariner of the Nebula*. Bristol: Institute of Physics Publishing, 1997.

Janet Browne. *Charles Darwin: A Biography, Vol. 1 - Voyaging*. New York: Alfred A. Knopf, 1995.

Abraham Pais. *'Subtle is the Lord...' - The Science and the Life of Albert Einstein.* Oxford: Oxford Universtiy Press, 1983.

Glossary

A

B

C

D

E

F

G

T

U

Name Index

M

N

O

P

X

Z

A·T·I·C·E

ATICE LLC, Albany NY

ISBN 978-1-951894-00-9