2

GENERAL APPROACH

The object of this introductory notice is not, however, solely to draw attention to the importance and greatness of the physical history of the universe, for in the present day these are too well understood to be contested, but likewise to prove how, without detriment to the stability of special studies, we may be enabled to generalize our ideas by concentrating them in one common focus, and thus arrive at a point of view from which all the organisms and forces of nature may be seen as one living active whole, animated by one sole impulse. … The physical history of the universe must not, therefore, be confounded with the Encyclopedias of the Natural Sciences, as they have hitherto been compiled, and whose title is as vague as their limits are ill defined. In the work before us, partial facts will be considered only in relation to the whole. (Alexander von Humboldt in *Cosmos* (1845), p. 55)

Introduction

Following the approach outlined above by the illustrious German scientist more than 150 years ago, in this chapter a general explanatory scheme for big history is proposed. Any claim to explain all of history must sound very audacious. So let me be clear about my aims and claims. First of all, explaining the past always implies striking a balance between chance and necessity. This point of view was expressed by the natural philosopher Democritus of ancient Greece (460–370 BCE), while French biochemist Jacques Monod said essentially the same more recently (with proper reference to Democritus).[1] My explanatory scheme is about necessity. It consists of general trends that not only make possible certain situations but also constrain them. Yet within these boundaries there is ample room for chance. Although I will not systematically focus on chance in this book, the reader should keep in mind that chance effects do influence the course of history.

Big History and the Future of Humanity Fred Spier
 ISBN: 978-1-444-33421-0

Everything that cannot be explained sufficiently is usually seen as the result of chance. This approach relegates chance to a rather unsatisfactory residual category. However, one may wonder whether pure chance actually exists. Whereas physicists claim that statistical chance rules in nature, most notably in quantum mechanics, in my view pure chance does not exist in reality, because everything is influenced by everything else either directly or indirectly.[2] In other words, as soon as the first regularities emerged, that was the end of pure undiluted chance. Yet within these emerging regularities, a great deal of chance effects do occur, in the sense of events that are so chaotic they cannot be seen as a direct result of those regularities. From the viewpoint of big history, it may therefore be argued that the increase of complexity over time would have led to a corresponding decrease of pure chance events. If correct, this might be a major trend in big history.

Even though a great many events have taken place in big history in which chance has played a role, a large number of unmistakable regularities and trends can be discerned. Apparently, these chance effects have jointly produced structured patterns of many different kinds. For instance, the collisions of all the molecules within an ocean are to a considerable effect based on chance. Yet such an ocean exhibits clear patterns, including currents, waves and varying degrees of salinity. While acknowledging chance effects, it is my first aim to explain such larger emergent properties.

While most processes are extremely complicated in their details, their overall structures may sometimes be surprisingly simple, if considered with the aid of a top-down approach (as exemplified by the Earthrise picture). By starting at the beginning of history, the big bang, the analysis is by necessity top down. By subsequently focusing on our galaxy, then on our solar system and finally on our home planet, it is relatively easy to recognize general patterns that would have been very hard to distinguish had we followed a bottom-up approach, by starting with our own societies today and then widening the view. Such an approach would soon become overwhelming. Because the details are already very complicated, widening the view only leads to more complications, which would be way too hard for even scholarly minds to handle. Yet by starting the analysis at an elevated level, it is relatively easy to see general patterns that might escape one's attention if one were to follow the bottom-up approach.

This does not mean that I think bottom-up approaches are unimportant. Indeed, if one wants to paint a reasonably reliable picture of what developments looked like at a local or regional level, it is essential to immerse oneself into a great many details, as I discovered myself while doing research into religion and politics in the Peruvian Andean village of Zurite. But if one wants to understand how these events were embedded into larger processes, the combination with a top-down approach is indispensable.

Because my explanatory scheme deals with everything ranging from the smallest particles to the universe as a whole, it needs to be formulated in very general terms. It must consist of those general aspects of nature that galaxies, solar systems, human societies, bacteria, molecules and even the tiniest particles all share. As will be shown, this includes the terms 'matter,' 'energy,' 'entropy' (disorder) and 'complexity.'

Before we can explain history, we need to discern those major regularities that we seek to explain. This raises the profound question of whether such regularities can be detected at all. Whereas many traditional accounts of human history consist of major events that are placed within a chronological time frame, I am following the approach to history in which important processes play a major role. These include the agrarian revolution, state formation, globalization and industrialization. Within these larger processes, a great many smaller-scale processes can be distinguished, such as the establishment of the Catholic Church in colonial Peru (which I studied myself in more detail).

All the events that historians consider important must, of course, find their proper place within these larger processes. The industrial revolution, for instance, can be interpreted as a process that first began in England, while it has now spread all around the inhabitable world. Within such a general framework, one can fruitfully study the industrialization of specific countries such as South Korea. While many historians have not yet embraced the process approach, all natural scientific accounts of big history, ranging from cosmic evolution to Earth history, are phrased in such a way. As a result, the process approach to human history advocated here fits very well within this larger context.

If we want to explain big history, we must inventory the major processes that have taken place. In my book *The Structure of Big History* (1996), I explored this theme by proposing the term 'regime' as the general key concept for indicating all the processes that make up big history. With the aid of this concept, the most important regimes were discussed, including their interactions. I placed great emphasis on human history, because this was the only discipline still lacking a central paradigm in Thomas Kuhn's sense. This approach provided a general structure for big history that, at the time, felt like a major theoretical step forward. About six years later, it dawned on me that regimes would be very useful for not only structuring big history but also explaining it.

In October of 1996, I visited the Santa Fe Institute in New Mexico, which is dedicated to the study of what they call 'complex adaptive systems.' As the term suggests, these are forms of complexity able to adapt to the prevailing circumstances. During that visit, I began to wonder what regimes and complex adaptive systems had in common. It seemed to me that all complex adaptive systems are regimes of some sort. Yet because in big history many regimes are not adaptive, including stars, galaxies and black holes, complex adaptive systems should be regarded as a subset of all the regimes that have existed in the universe. As a result, in big history there are at least two types of regimes, complex adaptive systems and complex nonadaptive systems. Interestingly, the term 'regime' appeared to cover all forms of complexity that have ever existed.

I prefer the term 'regime,' rather than 'system,' because there are no forms of complexity that are completely stable over time. This is especially important within the social sciences, where the term 'system' often bears the connotation of a static entity.[3] Because we need to bridge the gap between 'the two sciences' in big history, we must make an effort to find terms that are acceptable to all sectors of academia. In my usage, the term 'regime' is a shorthand expression for conveying both the structure and the change of processes. Given the remarkable variety of regimes found in the modern scientific literature, ranging from celestial regimes to regimes of the tiniest particles, I have some hope that the term 'regime' may actually become more widely accepted as an analytical term.[4]

The shortest summary of big history is that it deals with the rise and demise of complexity at all scales. As a result, the search for an explanation boils down to answering the question of why all these different forms of complexity have emerged and flourished, sometimes to disintegrate again. Here I will argue that the energy flowing through matter within certain boundary conditions has caused both the rise and the demise of all forms of complexity. Right now, this may sound very abstract, and I can only hope that the elaboration below will bring this formulation alive. Before exploring this concept in any further detail, we will first examine the scientific meaning of the key terms 'matter,' 'energy' and 'complexity.'

Matter and Energy

It is surprisingly difficult to find a satisfactory answer to the simple question of what matter and energy are. Eric Chaisson, for instance, defines matter as 'anything that occupies space and has mass,' while he describes mass as 'a measure of the total amount of matter, or "stuff," contained within an object.'[5] In my opinion, this is a circular argument. Yet I have found no physics textbooks that provide any further clarity. Apparently, it is very difficult to define matter unambiguously. A similar problem appears while trying to define energy.[6] Why would that be?

In my opinion, this problem is first of all caused by the nature of defining things. Inevitably, any definition involves a short description of a concept in terms of other concepts that are considered to be unproblematic. In doing so, the often tacit assumption is made that there are unproblematic concepts. Yet as soon as we start probing these supposedly unproblematic concepts, we find that they are problematic also. The second problem is that if one wants to define concepts that are considered basic, or fundamental, such as matter and energy, there are no even more fundamental concepts available that can be used for these definitions. This explains why basic concepts can probably never be defined satisfactorily.

In the second place, like almost all scientific terms, matter and energy were first used as everyday concepts. When these concepts began to be employed as scientific terms, their meanings were narrowed, first by specific language and later by mathematical formulas. Although this approach has led to a great many deep insights, one may wonder whether there are limits to the application of terms derived from everyday human experience to either the smallest particles or the largest possible structures in the universe. This has led to, for instance, some confusion about questions such as the dual character of light as a wave and a particle (though without mass). It may well turn out to be that in the next century, scientists will design more detached terms that would make our current terms and theories look hopelessly old fashioned. Yet we live here and now, and we have to make do with the best possible scientific terms currently at our disposal.

The first scientific use of the term 'matter' can be traced back to at least c.400 BCE in ancient Greece, when Democritus of Abdera theorized that all the everyday stuff we could observe was composed of extremely tiny, and therefore invisible, *atomoi*, portions of matter that could not be split up any further. These ideas re-emerged during the rise of modern science in Europe.

The first emergence of the term 'energy' may be similarly ancient. Greek philosopher Aristotle would have coined the term *energeia* around 350 BCE, while arguing that 'every object's existence is maintained by *energeia* related to the object's function.'[7] A more modern scientific use of the term 'energy' appears to date back only to the early nineteenth century. This was the period of the industrial revolution, which was driven by steam engines. Because these machines were used by commercial enterprises to make money, there was a premium on any invention that could improve their efficiency. Over the course of time, this led to a new branch of science, now known as thermodynamics, in which terms like 'energy' and 'entropy' (disorder) began to figure prominently. During the same period, scientists also investigated both the domain of the very small particles and the largest discernable structures in the sky. A few outstanding scientists, such as Lord Kelvin and Ludwig Boltzmann, soon realized that the new thermodynamic concepts could be applied to the universe as a whole. Yet a fully fledged application of thermodynamics to living matter only emerged in the 1970s.

Let us now return to the question of how to define 'matter' and 'energy.' Given the fact that our scientific understanding of matter and energy has evolved from everyday concepts, and given the issues related to defining these things, I propose to tackle the definition of matter and energy in the following way. Here, 'matter' is defined as anything that we humans in principle can touch: an everyday concept that hopefully makes some sense. Touching also includes scientific measurements. For instance, we usually measure mass with the aid of other masses, often with a scale of some sort. Of course we are unable to touch any matter beyond our reach, including most of the matter that exists in the universe. The presence of matter far away from us is inferred by the light it has emitted or by its gravitational effects on forms of matter that do emit light.

In big history, light plays a major role. The light we observe with our eyes is, in fact, only a small portion of a whole range of wavelengths that scientists call 'electromagnetic radiation.' In this book, the shorthand term 'light' will often be used for indicating electromagnetic radiation. According to natural scientists, light can be described as waves with a particle-like character, in this case particles without mass (whatever that means). Because light supposedly has no mass, it would not be matter. Yet its effects on matter, for example on our eyes or another type of light detector, are clearly visible. We can only measure light through its interactions with matter and through our subsequent interactions with that matter. If there were no matter at all in the universe, it would be impossible to detect any light. Thanks to the effects of light on matter, we can infer the masses of structures far away, such as planets, stars and even entire galaxies. We do so by measuring the light that was emanated from such structures that hit detectors mounted within our telescopes. The resulting pictures are interpreted in terms of established scientific theory. In this way, scientists have estimated the masses of things far beyond our direct reach.

In our current scientific thinking, light is considered a form of energy. There are many other forms of energy, including kinetic energy and nuclear binding energy, all of which have in common that we can detect them as a result of their effects on matter. The effect of light on a detector is such a case, while a collision between two moving cars – two chunks of matter that in their violent encounter convert kinetic energy into a change of matter – presents another example of the same process. A closer examination of the effects of energy on matter has led scholars to the profound insight that it is energy – and energy alone – that can make matter change. It makes sense, therefore, to define 'energy' as anything that can change matter, either its structure or its movements, including making it more, or less, complex.

Complexity

As was mentioned, big history deals with the emergence and decline of complexity. In the beginning, there would not have been any complexity at all. The further the universe evolved, the more complex some portions of it could become, most notably galaxies. Yet after a rather stormy beginning, most of the universe became, in fact, rather empty and therefore not complex at all. Today, after almost 14 billion years of cosmic existence, the human species is arguably the most complex biological organism in the known universe.

Unfortunately, no generally accepted definition of 'complexity' appears to exist.[8] As a result, there is no established way of determining different levels of complexity. Yet it surely makes sense to call certain configurations of matter more complex than others. Who, for instance, would be willing to argue that a bacterium is more complex than a human being, or that a proton would be more complex than a uranium nucleus? It is often said that a system (I would prefer 'regime') is more complex when the whole is greater than the sum of its parts.[9] This idea was coined in the 1890s by two German founders of gestalt psychology, Christian von Ehrenfels and Max Wertheimer. In modern complexity studies, this difference is expressed in terms of emergent properties: characteristics of a certain level of complexity that cannot be derived from a lower level. Life, for instance, is such a characteristic, because it cannot be derived from the molecules that constitute a living entity. French founding father of sociology August Comte and, in his footsteps, German sociologist Norbert Elias characterized these properties in terms of relative autonomy: different levels of complexity that cannot be reduced to lower levels.[10]

Because no generally accepted definition of 'complexity' appears to exist, I decided to tackle this problem by making an inventory of its major characteristics. First of all, there is the number of available building blocks. As more building blocks become available, structures can become more intricate. The same is the case when the variety of the building blocks increases. Clearly, with a greater variety of building blocks, more complex structures can be built. The level of complexity can also increase when the connections and other interactions between and among the building blocks become both more numerous and more varied. On the whole it appears, therefore, that a regime is more complex when more and more varied connections and interactions take place among increasing numbers of more varied building blocks.

At different levels of complexity, different types of building blocks can be discerned. The basic building blocks of ordinary matter are protons, neutrons and electrons. These elementary particles can combine to form chemical elements, which are building blocks on a higher level of complexity. The chemical elements, in their turn, can combine to form molecules, which can be seen as building blocks on an even higher level of complexity. They may jointly form stars, planets and black holes, which are the building blocks of galaxies that, in their turn, may be the building blocks of galaxy clusters. Chemical elements may also combine to form molecules. At a higher level of complexity, a great many different molecules may jointly form cells, which may combine to form individuals that, in their turn, may be the building blocks of society. All these different levels of complexity should be considered relatively autonomous with regard to one another, which simply means that such a particular level of complexity exhibits emergent properties that cannot be sufficiently explained from the properties of a lower level of complexity.

There is another important aspect to complexity, namely sequence. Digital computer information, for instance, consists of only two elementary building blocks, namely ones and zeros. Yet by using enormous amounts of ones and zeros in specific sequences, humans have been able to generate a great deal of complexity. Apparently, the sequences in which these building blocks are organized can produce considerable levels of complexity, while only a slight change in sequence can wreck this complexity entirely. The sequence of building blocks, and thus information, mostly matters in life and culture. In life, the genetic information is organized in long strands of DNA molecules, in which the sequence of the building blocks is of overriding importance for determining what happens inside cells. In a similar way, sequence is also important for all cultural information and communication.

One may argue that lifeless nature can also exhibit certain sequences and can thus carry information. Sediments, for instance, may consist of a great many layers, each containing fossils of many different kinds, which are interpreted by scientists as clues to a more or less distant past. Yet there is an important difference between such things and genetic or cultural information. Sediments and fossils do not perform any functions for the regime as a whole – they are just there. The information stored in genetic molecules and in cultural depots, such as books and computer hard drives, by contrast, can always be interpreted as having some function for the individuals they belong to.

While comparing different forms of complexity, one has to take into account their complexity per unit mass (kilogram). Otherwise, a piece of rock weighing a few kilograms, just by its sheer size and consequently its large number of atomic building blocks, would have to be considered much more complex than a tiny microorganism. Yet as soon as we compare rocks and microorganisms per unit mass, then this little living thing suddenly appears much more complex, thanks to its greater variety of building blocks and connections.

The approach of defining complexity in terms of building blocks, connections and sequences should in principle allow us to determine to what extent the whole is greater than the sum of its parts. Yet this is very difficult in practice. For how would we rate the different aspects and which equations would we use? What would count for more: a greater variety of building blocks, more and more varied connections, or perhaps a longer and more varied sequence? Right now, I find it impossible to rate all these aspects in a way that would allow us to compute levels of complexity reliably. If possible at all, achieving such a goal even in terms of a first order approach could well constitute an entire research agenda. And even if we could achieve this, would this lead to a sufficiently precise characterization of the emergent properties of that particular level of complexity? As a result, for the time being, we have to rely on qualitative, rather subjective, statements of how to assess all the levels of complexity in the known universe. This may be unsatisfactory, yet to my knowledge this is the best available approach today.[11]

The terms 'order' and 'complexity' do not always mean the same thing. A crystal consisting of sodium chloride (ordinary salt), for instance, may be extremely regular and orderly, because it is made up of alternating positively charged sodium ions and negatively charged chloride ions that are located in a very orderly fashion. Yet such a crystal should not be considered extremely complex, because it has only a few building blocks that interact with one another in very simple ways. I prefer to reserve the term greater complexity for biological organisms, in which a great many molecules of different kinds interact in myriad ways. As a result, the opposite of disorder consists of two types of order: on the one hand a type of very regular order that is not by necessity very complex, and on the other hand a type of order that consists of a great many structured compounds that interact with each other.

Forms of greater complexity never suddenly emerge all by themselves out of nothing. Instead, they always develop from forms of lower complexity. Human societies, for instance, emerged out of groups of primates, which, in their turn, developed from earlier, less complex, life forms. This is just one example of a very general rule. Such a process usually takes large amounts of time. The destruction of great complexity, by contrast, can go very quickly, while it may revert to very low complexity without passing through a great many intermediate stages. This happens, for instance, when humans are cremated after having passed away.

On our home planet, we cannot create any new complexity without destroying existing forms. We simply do not have a new set of building blocks at our disposal that we can use for a new construction within free, empty space. Instead we are surrounded by existing forms of complexity that we reshape. As a result, while creating new forms of complexity, we are also continuously destroying old ones. And we should not forget that humans have also engaged in destroying forms of complexity without creating new ones.

Let us now take a crude qualitative look at the various levels of complexity that can be discerned in big history. According to many scholars, there are three major types of complexity: physical inanimate nature, life and culture. In terms of matter, lifeless nature is by far the largest portion of all the complexity known to exist in the universe. The following example may help to grasp the significance of its sheer size. Let us assume for the sake of simplicity that the whole Earth weighs about as much as an average American car, about 1,000 kg. The combined weight of all planetary life would then amount to no more than 17 mcg. This more or less equals the weight of a tiny paint chip falling off that car. Seen from this perspective, the total weight of our solar system would be equivalent to that of an average supertanker. Because the mass of our galaxy is not well known, it is hard to extend this comparison any further. But even if life were as abundant in our galaxy, or in the universe as a whole, as it is within our solar system, its relative total weight would not amount to more than a paint chip on a supertanker.

All of this cosmic inanimate matter shows varying degrees of complexity, ranging from single atoms to entire galaxies. It organizes itself entirely thanks to the fundamental laws of nature. Whereas the resulting structures can be exquisite, inanimate complexity does not make use of any information for its own sustenance. In other words, there are no information centers that determine what the physical lifeless world looks like. It does not make any sense, therefore, to wonder where the blueprint of our solar system is stored that would help to shape Earth or our solar system, because it does not exist.

The second level of complexity is life. As we just saw, life is a rather marginal phenomenon in terms of mass. Yet the complexity of life is far greater than anything attained by lifeless matter. In contrast to inanimate complexity, life maintains itself by continuously harvesting matter and energy with the aid of special mechanisms. As soon as living things stop doing so they die, while their matter disintegrates into lesser levels of complexity. To achieve these elevated levels of complexity, life organizes itself with the aid of hereditary information stored in DNA molecules. While trying to find out how life works, it does therefore make a great deal of sense to wonder where the information centers are located that help configure it, what this information looks like, how the control mechanisms work that help to translate this information into biological shapes and what the limitations of these mechanisms are in shaping organisms.

The third level of complexity consists of culture: information stored in nerve and brain cells or in human records of various kinds. The species that has developed this capacity the most is, of course, humankind. In terms of total body weight, our species currently makes up about 0.005 per cent of all planetary biomass. If all life combined were only a paint chip, all human beings today would jointly amount to no more than a tiny colony of bacteria sitting on that flake. Yet through their combined efforts humans have learned to control a considerable portion of the terrestrial biomass, today perhaps as much as between 25 and 40 per cent of it. In other words, thanks to its culture this tiny colony of microorganisms residing on a paint chip has gained control over a considerable portion of that flake. To understand how human societies operate, it is therefore not sufficient to only look at their DNA, their molecular mechanisms and the influences from the outside world. We also need to study the cultural information that humans have been using for shaping their own lives as well as considerable portions of the rest of nature.

In contrast to genes, the building blocks of cultural information cannot be defined unambiguously. It is, therefore, even more difficult to rigorously define cultural complexity. Cultural concepts not only are flexible and apt to change very quickly, but also need to be interpreted by people. While genetic information needs to be interpreted unambiguously in living cells by its cellular machinery to function properly, such a lack of ambiguity in interpretation is rare in human societies, if it ever occurs.[12] Nonetheless, cultural information has allowed many animals, including humans, to successfully wage the struggle for life.

The greatest complexity known to us, namely life, may well be a marginal phenomenon, in the sense both that it is exceedingly rare and that, in terms of matter concentration, it can be found on the margins of larger regimes. Life as we know it exists on the surface of a planet situated relatively close to the edge of its galaxy. Most of the planetary matter is below our feet – it is not surrounding us. In the solar system, most of its matter is concentrated in the sun and not beyond the Earth's orbit around the sun. A similar observation can be made for our position within the galaxy. Yet, as Eric Chaisson observed, this is not the case for the complexity within life. The greatest biological complexity, most notably DNA and brains, is found in well-protected areas and not on their edges. These types of greater complexity are there because they need to be protected against matter and energy flows from outside that are too big, which would lead to their destruction. Apparently, life has created a space suit to protect its greatest levels of complexity. In fact, terrestrial life may actually have succeeded in turning the entire biosphere into a space suit. This is, in my view, the essence of James Lovelock's Gaia hypothesis discussed in chapter five, which states that terrestrial life has evolved feedback mechanisms that condition the biosphere in ways that are advantageous for its continued existence.

During the history of the universe, all these forms of physical, biological and cultural complexity would have emerged all by themselves. In the scientific approach, the possible influence of supernatural forces bringing about complexity is not considered to be an acceptable explanation, because we have never observed such forces at work. The major question then becomes: how does the cosmos organize itself? This question becomes even more difficult when we realize that in our daily lives we usually observe the opposite, namely the breakdown of complexity into disorder. Children's rooms, for instance, never clean themselves up, while cities without a trash-collecting regime would soon choke in their own refuse. This tendency is known as the Second Law of Thermodynamics, which states that over the course of time, the level of disorder, or entropy, must rise. In other words, the history of the universe must also be the history of increasing disorder. Any local rise in complexity must, therefore, inevitably have been accompanied by a larger rise of disorder elsewhere. Given this situation, how could complexity have emerged all by itself?

Energy Flows and the Emergence of Complexity

To understand the rise and demise of complexity, it is important to make clear distinctions between the emergence of complexity, its continuity during a certain period of time and its eventual demise. According to the modern view, the emergence of any form of complexity requires an energy flow through matter. Only in this way is it possible for more complex structures to arise. The emergence of life, for instance, must have required a continuous energy flow. But also stars need an energy flow to come into being, while the same happened to planets and galaxies, as we will see in the coming chapters.

As soon as complexity has emerged, it depends on its nature whether energy is required to keep it going. Some forms of lifeless complexity are close to thermodynamic equilibrium, which means that in the prevailing circumstances very little spontaneous change occurs. Rocks swinging through empty space, for instance, do not need an energy flow to keep more or less the same shape for long periods of time, as long as they are not disturbed by outside events. The same is the case for galaxies and black holes. Yet even these relatively simple structures are never completely sealed off from what happens in the rest of the universe. As a consequence, they are undergoing change through energy from outside, such as cosmic radiation, collisions with other celestial bodies or the decay of their atoms over extremely long periods of time. And because they lack an energy flow that would counter these trends, such simple structures will eventually decay and thus lose whatever complexity they had in the very long run.[13]

More complex forms of lifeless nature, most notably stars and planets, are often not very close to thermodynamic equilibrium and can only exist because of an energy flow that allows them to retain their shape. Such objects are said to be in a dynamic steady state. To be sure, stars and planets are continuously changing, yet they may maintain their shapes more or less over long periods of time. Stars, for instance, can shine for as long as they release energy within their cores through the process of nuclear fusion, in which hydrogen is converted into helium. The current, much less dynamic, layered complexity of Earth, by contrast, which consists of its outer crust, mantle and core, emerged as a result of the energy flows acting during its emergence, which are now mostly gone. Today, the dynamic surface complexity of our home planet is determined by the heat released deep within it through processes of nuclear fission as well as by the energy from outside received from the sun.

As Russian-born Belgian scientist Ilya Prigogine argued, all life forms are far from thermodynamic equilibrium. In contrast to lifeless nature, all life forms must harvest matter and energy from outside on a continuous basis. Humans, for instance, have to keep eating, drinking and breathing on a continual basis to keep our complexity going. If we stopped doing so, our complexity would very soon begin to disintegrate. The energy that we ingest serves many purposes: keeping our metabolism going, making plans, moving around etc. During these processes, the ingested energy is transformed from high-quality to lower-quality energy. As a result, we constantly generate heat (a form of lower-quality energy) that we subsequently radiate out into the surrounding environment. This is one of the ways humans get rid of the inevitable disorder (entropy) that is produced to keep our complexity going. If we were unable to radiate this energy, we would soon suffocate in our own heat. Another major way of discarding entropy is to follow the call of nature by excreting wastes. These characteristics apply not only to humans but also to all other living beings.

To sum up, the complexity of humans, Earth and the sun all have in common the need for an energy flow through matter to keep going while producing entropy. Canadian energy expert Vaclav Smil formulated this in 1999 as follows:

> Energy is the only universal currency: one of its many forms must be transformed to another in order for stars to shine, planets to rotate, plants to grow, and civilizations to evolve. Recognitions of this universality was one of the great achievements of nineteenth-century science, but, surprisingly, this recognition has not led to comprehensive, systematic studies that view our world through the prism of energy.[14]

While flowing through matter, energy inevitably changes from a more to a less productive state. This can be caused by the absorption of some of this energy by the matter that is becoming more complex. Many molecules produced by life, for instance, can only be formed by adding energy. Yet as soon as these forms of greater complexity break down, this energy is released again, although always in a lower-quality form. The need for the absorption of certain amounts of energy to make possible the emergence of complexity is a very general principle. It should be seen as a refinement of the earlier-mentioned general approach consisting of energy flows through matter as an absolute requirement for the emergence of complexity.

By flowing through matter, energy always changes from a higher-quality to a lower-quality form. For instance, the energy stored in our food intake is clearly more valuable for keeping our complexity going than the leftover energy in the products we excrete. Apparently, some forms of energy are better able to produce or maintain complexity than others. In the science of thermodynamics, the ability of energy to change matter is expressed with the term 'free energy.' In this book, which offers a first crude look at this new general approach to big history, we will not systematically examine how energy changes while flowing through matter. Instead, we will mostly consider only the energy input. In a more refined analysis, it will, of course, also be important to investigate systematically how energy changes while flowing through matter.

Can we measure and calculate these energy flows through matter during all of history? In his ground-breaking book of 2001 *Cosmic Evolution: The Rise of Complexity in Nature*, Eric Chaisson sought to do so by defining the concept of 'free energy rate density,' indicated with the symbol Φ_m, as the amount of energy that flows through a certain amount of mass during a certain period of time. For human beings, for instance, it is the amount of energy we ingest during a certain period, let's say 24 hours, divided by our body weight. In principle, Chaisson's approach allows us to calculate these values for every form of complexity that has ever existed, ranging from the tiniest particles to galaxy clusters. This makes it possible to compare all forms of complexity systematically. Unfortunately, the term 'free energy rate density' is rather bulky, while it is equivalent to 'power density,' a term that is often used by physicists, as Chaisson noted in his book. Because in 2009 Chaisson began to use the term 'power density' instead of 'free energy rate density,' this will be our preferred term.[15]

Chaisson next showed that a clear correlation exists between the intuitively defined levels of complexity observed in the known universe and the calculated power densities. Surprisingly, perhaps, whereas humans may seem vanishingly small compared to most other aspects of big history, we have generated by far the largest power densities in the known universe.[16] In Table 2.1, Chaisson summarized some of his findings.[17]

For many people, these results are counterintuitive. One would expect, for instance, the power density of the sun to be much greater than the power density of our brains. Yet whereas the sun emits far more energy than the energy that is used by our brains, the power density of the brain is much larger, because the brain is so very small compared to the sun. In general, the power

Table 2.1: Some estimated power densities (reproduced with permission)

Generic Structure	*Approximate Age (10^9 year)*	*Average Φ_m (10^{-4} watt/kg)*
Galaxies (Milky Way)	12	0.5
Stars (Sun)	10	2
Planets (Earth)	5	75
Plants (biosphere)	3	900
Animals (human body)	10^{-2}	20,000
Brains (human cranium)	10^{-3}	150,000
Society (modern culture)	0	500,000

densities of life are considerably greater than those of lifeless matter. Apparently, these tiny living regimes generate much greater power densities than their lifeless counterparts.

For a good understanding of the numbers in this table, we need to consider Chaisson's calculations in more detail.[18] Let us start with the power density for galaxies. Many people may think that galaxies are simply collections of stars. If that were the case, the power density of a galaxy would simply be the average of all the power densities of its individual stars. Yet Chaisson's power density for galaxies (in fact our own galaxy) is considerably smaller. In addition to the fact that a considerable amount of matter in our galaxy consists of gas and dust, the power density for our galaxy is also lower because all the so-called dark matter is included in its total mass. Unfortunately, as will be explained in chapter three, we do not know whether dark matter actually exists. Furthermore, our galaxy is thought to harbor a rather heavy black hole in its center, consisting of extremely dense matter, which would exhibit very little complexity, if any. Because gas, dust, black holes and dark matter do not release any energy, while they may form a considerable portion of the galaxy's mass, they lower its power density, which is therefore smaller than the power density for stars. In fact, Chaisson's value for stars was calculated for our sun, which is an average star.

Whereas the energy flows emitted by stars keep them going, they did not create the overall structure of our galaxy: a large swirling cloud of stars with huge arms. The energy flows that once gave rise to this galactic structure are absent in Chaisson's calculations. The reason for this is that the structure of our galaxy emerged a long time ago, while today it does not need energy any more to keep going. But this can change. As soon as galaxies collide, a flow of kinetic energy is generated that reshapes them. Such a cosmic encounter is expected between our galaxy and its nearest neighbor, the Andromeda nebula, to take place between 2 and 5 billion years from now. Also within galaxies there is constant change, including contracting gas clouds and exploding stars, which releases energy that reshapes these galaxies. Seen in the long run, however, these energy flows and their effects are probably minute compared to the output of all the combined stars and, as a result, do not have to be taken into account for computing a first-order estimate of our galaxy's power density.

Chaisson's power density for galaxies characterizes a relatively stable galactic regime and not a regime in rapid formation or decline. This is actually the case for all of Chaisson's power densities – they all characterize dynamic steady-state regimes. In other words, the energy flows needed for the emergence of these regimes do not play a role in Chaisson's table of calculations for the present.

Let us now consider Chaisson's power density for planets. In fact, this value does not reflect the complexity of any known planet as a whole. It was calculated for only a thin slice of the outer shell of Earth by estimating the amount of solar energy reaching the terrestrial surface during a certain period of time, while using the weight of the atmosphere plus an oceanic layer of 30 m as the total mass. According to Chaisson, this is where most of our planet's complexity resides. Because the geothermal energy generated deep inside Earth is several thousand times smaller than the radiation energy received from the sun, Chaisson did not include geothermal energy in his calculation.

The next power density in Chaisson's table, the average power density for plants, is an average value that includes all living matter, while the value for animals was calculated for the energy used by the human body. This power density was arrived at by calculating the average food intake per body weight. Nonetheless in reality, as Chaisson pointed out, the power densities of vertebrate animals vary by almost a factor of 10.[19] This raises the issue of whether those vertebrate animals that exhibit the largest power densities, namely birds, should be considered the most complex. Chaisson's estimate for human society (modern culture) is based on the current energy use of 6 billion people with an average body weight of about 50 kg (adults and children).[20] In this case, most of the energy does not flow through human bodies. If it did, humanity would cease to exist instantaneously.

The power densities provided by Chaisson for human history exhibit some further problems. Dutch environmental scientist Lucas Reijnders has pointed out that, thanks to their fire use, early humans may have achieved very high power densities. They might have manipulated enormous energy flows by burning large tracts of land, which created desired forms of complexity, such as grasslands, while destroying other forms of complexity, usually woodlands. By stoking fires, they roasted food, while keeping themselves warm and safe from predators. In doing so, the amounts of energy used by recent Australian aboriginals were one to two orders of magnitude larger than those of the average US citizen in 1997.[21] This makes one wonder how large the power densities were that early humans were able to achieve in Australia and elsewhere, wherever nature could be set on fire on a large scale. If one wants to use the power density as a measure of complexity, as Chaisson suggests, Australian aboriginal society would have to be considered more complex than modern industrial societies. This seems unsatisfactory to me.[22]

Today, most of the energy employed by humans is not used for keeping their bodies going or burning the land but for the creation and destruction of what I will call 'forms of constructed complexity': all the material complexity created by humans. These include clothes, tools, housing, engines and machines and means of communication. With the aid of these things, humans have transformed both the surrounding natural environment and themselves. To be sure, not only humans but also many animals have produced a great many forms of complexity. Well-known examples include spider webs and beaver dams. Yet it seems fair to say that humans have developed this capacity to a far greater extent than any other species.

Complexity constructed by humans can be divided into two major categories. On the one hand there are things that do not need an energy flow for their intended functioning, while on the other hand there are things that do need such an energy flow. The first category, which could be called 'passive constructed complexity,' includes things such as clothes, housing and roads. This type of complexity is made by humans as well as by a great many other animals. The second type of complexity, things that do require continuous external energy sources for their intended functioning, will be called 'powered constructed complexity.' This category includes machines driven by energy from wind, water and fossil fuels. To my knowledge, only humans have constructed forms of complexity driven by external energy sources. In this sense, humans are unique in the known universe.

Many forms of powered constructed complexity exhibit much higher power densities than the power densities of human brains (about 15 watt/kg) or human societies (about 50 watt/kg). As Chaisson pointed out, jet engines achieve power densities between 2,000 watt/kg (Boeing 747) and 80,000 watt/kg (F-117 Nighthawk).[23] Relatively high power densities are characteristic not only of jet planes but also of a great many household appliances. While performing a few calculations at home, my son Louis and I found that even our humble vacuum cleaner exhibited a power density of about 180 watt/kg, thus outperforming our brains more than tenfold.[24] This does not imply that jet engines and vacuum cleaners should be considered more complex than human brains. Unlike forms of complexity that emerged spontaneously, forms of constructed complexity are not using this energy for the purpose of achieving greater complexity within themselves. Instead, they were designed to use considerable amounts of energy to perform certain tasks, such as moving heavy objects through the air or achieving a certain degree of order within our living space.

Although a great many complications emerge on closer inspection, Chaisson's analysis seems fair enough as a first-order approach. With it he provides what US physicist Murray Gell-Mann calls 'a crude look at the whole,' which in the natural sciences is considered perfectly legitimate.[25] Chaisson is well aware of this. As he formulated it:

> A second caveat [the first caveat was the danger of anthropocentrism] concerns the level of detail in our computational analysis; to be honest, we have skirted some of the hardest details. In particular, as noted at the outset of our calculations, the values for Φ_m employ only bulk flow, that is, total energy available to a handful of representative systems. Accordingly, quantity, or intensity, of energy has been favored while largely neglecting measures of quality, or effectiveness of that energy. Clearly, a more thorough analysis would incorporate such factors as temperature, type, and variability of an emitting energy source, as well as the efficiency of a receiving system to use that free energy flowing through it. After all, input energy of certain wavelength can be more useful or damaging than others, depending on the system's status, its receptors, and its relation to the environment. Likewise, the efficiency of energy use can vary among systems and even within different parts of a given system; under biological conditions, for example, only some of the incoming energy is available for work, and technically only this fraction is the true free energy. That energy might benefit some parts of a system more than others is a necessary refinement of the larger opus to come. For this abridgment, our estimates suffice to display general trends; the next step is a more complete (perhaps we should say more "complex") study to examine how, and how well, open systems utilize their free energy flows to enhance complexity.
>
> Even the absolute quantity of energy flowing through open systems needs to be more carefully considered in a detailed analysis. Not just any energy flow will do, as it might be too low or too high to help complexify a system. Very low energy flows mean the system will likely remain at or near equilibrium with the thermal sink, whereas very high flows will cause the system to approach equilibrium with what must effectively be a hot source – that is, damage the system to the point of destruction. … Sustained order is a property of systems enjoying moderate, or "optimum," flow rates; it's a little like the difference between watering a plant and drowning it. In other words, a flame, a welding torch, and a bomb, among many other natural and human-made gadgets, have such large values of Φ_m as to be unhelpful.[26]

All of this should also remind us again of the fact that the data shown in Chaisson's table are about relatively stable matter regimes with relatively stable energy flows, and not about the emergence or decline of specific forms of complexity.

In his approach, Chaisson employed these numbers first of all as a way of measuring different levels of complexity. This was his way of tackling the issue of how to rigorously define and measure different levels of complexity. At the same time, Chaisson also used these numbers as an indication of the energy needed to achieve or maintain certain levels of complexity. This latter approach will be followed in this book. In the next chapters, I will explicitly not employ the concept of power density as the one and only yardstick for measuring different levels of complexity. It will only be used as an indication of the energy that is needed for complexity to emerge and continue to exist.

The Goldilocks Principle

As Eric Chaisson noted but did not elaborate, complexity can only emerge when the circumstances are right. This includes, in the first place, the availability of suitable building blocks and energy flows and, in the second place, a great many limiting conditions such as temperatures, pressures and radiation. Complexity cannot emerge, or is destroyed, when the circumstances are not right. The destruction of complexity is usually caused by energy flows or energy levels that have become either too high or too low for that particular type of complexity. For instance, if biological organisms such as ourselves found themselves without protection in temperatures that were continuously either below 10 degrees Celsius or above 40 degrees Celsius, they would cease to exist. Apparently, there is a certain bandwidth of temperature levels within which humans can live. Such bandwidths exist not only for all living species but also for rocks, planets and stars. In other words, all relatively stable matter regimes are characterized by certain conditions within which they can emerge and continue to exist. In reference to a popular Anglo-Saxon children's story, this will be called the Goldilocks Principle.

For those readers not familiar with the story of Goldilocks, she is a little girl who happened to wander into a house in a forest where one young bear lives with his parents. The bears are, however, not at home. Goldilocks, hungry and adventurous, first tries out the porridge bowls on the counter top. She finds that the porridge in the biggest bowl is too hot and the porridge in the middle-sized bowl is too cold, but the porridge in the little bowl is just right. Then she tries out the chairs: the biggest one is too hard, the middle-sized one is too soft and the little one is just right. And so it goes on until the bears come home and do not like what they see. As a result, Goldilocks flees.[27]

I am not the first to employ the term 'Goldilocks Principle.' Over the past 10 years, a few scientists have begun using this term for indicating the circumstances that limit the emergence and continued existence of various forms of complexity. To natural scientists, the Goldilocks Principle may be obvious, because they perform all their analyses from this point of view. Surprisingly, however, to my knowledge no one has yet elaborated this principle systematically for all of big history.[28]

The Goldilocks Principle points to the fact that the circumstances must be just right for complexity to exist. It is important to see that these circumstances are often not the same for the emergence of complexity and for its continued existence. For instance, Goldilocks circumstances favoring the emergence of the smallest particles only existed during the first few minutes of cosmic history, as we will see in the next chapter. Apparently these conditions were very restrictive. Yet during the billions of years that followed, Goldilocks circumstances have favored the continued existence of these tiny particles, of which everything else consists, from galaxies to human beings. In this book, a great many examples of this general principle will be discussed.

Goldilocks requirements do not exist by themselves, but they always depend on the type of complexity under consideration. Humans, for instance, cannot live below or above certain temperatures, while our direct needs also include sufficient air pressure, oxygen, food and water. The Goldilocks requirements for stars, by contrast, are very different. Stars need huge amounts of closely packed hydrogen surrounded by cold empty space. As a result of gravity, these enormous balls, consisting of mostly hydrogen and helium, create so much pressure in their interiors that nuclear fusion processes ignite, thereby converting hydrogen into heavier (and thus more complex) helium nuclei while releasing energy in the form of radiation. These stellar Goldilocks circumstances are very hard to reproduce on Earth, which explains why nuclear fusion has not yet become feasible as a way of generating electricity.[29] In sum, all Goldilocks circumstances are characterized by certain bandwidths. In the natural sciences, the upper and lower limits of these bandwidths are known as boundary conditions.

More than any other animal, humans have created a great many Goldilocks circumstances that help them to survive. They can have both a social and a material character. Material Goldilocks circumstances include clothes, housing, tools of many kinds and roads, while an example of social Goldilocks circumstances would be presented by traffic rules. The rules are meant to define human behavior in ways that allow members of our species to reach their destination relatively efficiently while at the same time seeking to preserve the complexity of all the participants involved. Those who fail to obey the traffic rules usually do so to reach their destination more quickly at the risk of compromising safety. In fact, all social rules can be interpreted

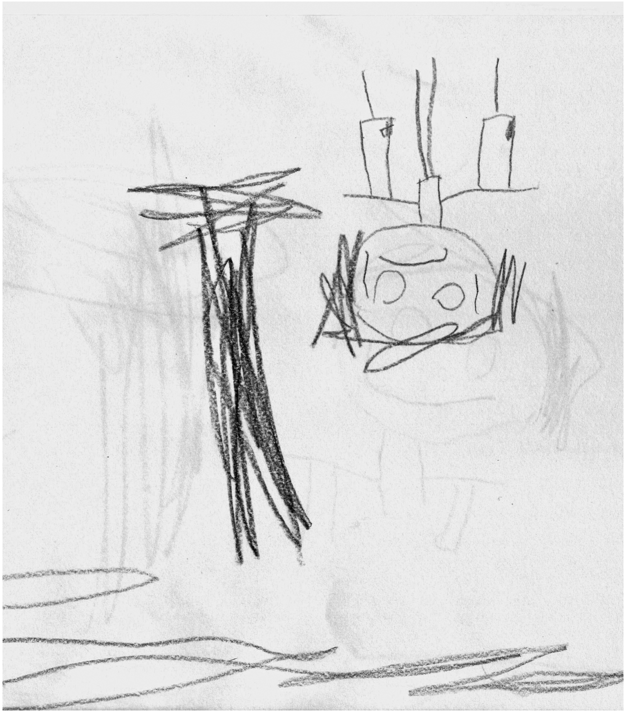

Figure 2.1: Goldilocks falling from a tree. Apparently, she has overstepped her boundaries. Soon, her complexity will be damaged as a result of the impact caused by gravitational energy. (Drawing by Giulia Spier, 4 years old.)

as Goldilocks circumstances that have been created by humans to preserve certain forms of complexity.

Goldilocks circumstances tend to vary both in space and in time. I will call such changes 'Goldilocks gradients.' This concept was first coined as an answer to the question of why the surface of our planet appears to be such a good place for the emergence of greater complexity. Why, indeed, do humans live on the outer edge of our home planet and not deep below its crust? My answer to that question is that the outer edge of our planet exhibits marked differences in the Goldilocks circumstances in space over relatively short distances, in other words: steep Goldilocks gradients. This allows life to capture large amounts of energy while discarding large amounts of entropy. This will be elaborated in the coming chapters. Suffice to say here that among biologists, steep Goldilocks gradients between different ecological zones are known as 'ecotomes' and have been studied intensively.[30]

A better understanding of complexity requires the concept not only of Goldilocks gradients in space but also of Goldilocks gradients over time. Whereas the range of planetary climate zones from the tropics to the arctic can be seen as a Goldilocks gradient in space, climate change happening within these zones can be interpreted as a Goldilocks gradient over time. Climate gradients over time may exhibit more or less regular patterns, such as those caused by regular changes of the Earth's orbit around the sun, the so-called Milanković cycles. This will be explained in more detail in chapter four. Suffice to say here that climate gradients over time have profoundly affected life on Earth for as long as we can detect.

In sum, to understand the rise and demise of any type of complexity, we must not only look at energy flows through matter but also systematically examine the prevailing Goldilocks circumstances. I think that the 'energy flows through matter' approach combined with the Goldilocks Principle may provide a first outline of a historical theory of everything, including human history. While this theory cannot, of course, explain everything that has happened, it does provide an explanation for general trends that have happened in big history.

Because a new, rather unbeaten, track is followed in this book, my effort should be seen as a first attempt at formulating a coherent theoretical framework for big history. This approach may actually constitute an entire interdisciplinary research agenda that, if pursued, would allow scientists ranging from astronomers to historians and anthropologists to collaborate in unprecedented ways while speaking the same scientific language. This may sound idealistic, yet in fact this process has already started.[31]

In the pages that follow, I will offer a simplified overview of big history. For obvious reasons, it is impossible to offer detailed discussions about everything that has ever happened in one book. This problem does not exclusively exist in big history. Any overview of any portion of history is bound to be a simplification of reality, because no historian will ever know all of the details of his or her subject. Furthermore, choices have to be made all of the time about what to include and what to omit. I very much hope, though, that the general trends in my big history account do accommodate most, if not all of the details. This would constitute a major test for my theory. Unfortunately, within the scientific community disagreements exist about a great many aspects of history, while the established scientific theories, most notably perhaps in cosmology, are currently insufficient to explain all of the observations. As a result, a great many choices had to be made concerning the question of which version of history would be presented here. Although I offer conflicting views of history in a number of cases, I found it impossible to outline all of the controversies that I encountered.

Notwithstanding all of these caveats, I hope to persuade the reader that my theoretical scheme does indeed offer the contours of a fresh, integrated approach for looking at images of the past in a way that reunites academic fields and disciplines that have grown apart, while providing general explanations of what we think the past looked like. Whereas we may never be able to explain everything that has happened, I hope to make clear that the opposite position, namely that we are unable to explain historical processes from a general point of view, is untenable. The challenge consists of finding a middle ground between the Scylla of no explanation at all where chance rules and the Charybdis of seeking to explain everything but not allowing for any chance.

My general approach deals with the emergence, continued existence and inevitable decline of complexity in all of its manifestations within big history. Its coherent framework spanning all of time and space helps to justify why it is important to understand human history within its cosmic context.