Complexity Evolving

an introduction to the new systemic worldview

Francis Heylighen

(unpublished manuscript for a popular science book, May 2000)
# Table of Contents

**Preface**

<table>
<thead>
<tr>
<th>PART I: WORLD OUT OF BALANCE</th>
<th>.................................................................</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>THE ACCELERATION OF CHANGE</td>
<td>.........................................................................................</td>
<td>9</td>
</tr>
<tr>
<td>SCIENTIFIC INNOVATION</td>
<td>.........................................................................................</td>
<td>9</td>
</tr>
<tr>
<td>TECHNOLOGICAL ACCELERATION</td>
<td>.........................................................................................</td>
<td>13</td>
</tr>
<tr>
<td>THE PACE OF LIFE</td>
<td>.........................................................................................</td>
<td>20</td>
</tr>
<tr>
<td>GLOBAL CHANGE</td>
<td>.........................................................................................</td>
<td>23</td>
</tr>
<tr>
<td>LOSING CONTROL</td>
<td>.........................................................................................</td>
<td>27</td>
</tr>
<tr>
<td>THE ADAPTATION SYNDROME</td>
<td>.........................................................................................</td>
<td>27</td>
</tr>
<tr>
<td>INFORMATION OVERLOAD</td>
<td>.........................................................................................</td>
<td>30</td>
</tr>
<tr>
<td>RUNAWAY PROCESSES</td>
<td>.........................................................................................</td>
<td>32</td>
</tr>
<tr>
<td>WALKING THE TIGHTROPE</td>
<td>.........................................................................................</td>
<td>35</td>
</tr>
<tr>
<td>PROGRESS</td>
<td>.........................................................................................</td>
<td>39</td>
</tr>
<tr>
<td>DEVELOPING NATIONS</td>
<td>.........................................................................................</td>
<td>39</td>
</tr>
<tr>
<td>THE RIDDLE OF INCREASING IQ</td>
<td>.........................................................................................</td>
<td>44</td>
</tr>
<tr>
<td>THE SINGULARITY</td>
<td>.........................................................................................</td>
<td>48</td>
</tr>
<tr>
<td>FRAGMENTED WORLD VIEWS</td>
<td>.........................................................................................</td>
<td>53</td>
</tr>
<tr>
<td>VISIONS OF THE FUTURE</td>
<td>.........................................................................................</td>
<td>53</td>
</tr>
<tr>
<td>POST-MODERN FRAGMENTATION</td>
<td>.........................................................................................</td>
<td>56</td>
</tr>
<tr>
<td>THE CAUSES OF DISCONTENT</td>
<td>.........................................................................................</td>
<td>60</td>
</tr>
<tr>
<td>THE BREAKDOWN OF VALUES</td>
<td>.........................................................................................</td>
<td>63</td>
</tr>
<tr>
<td>A NEW WORLD VIEW?</td>
<td>.........................................................................................</td>
<td>67</td>
</tr>
<tr>
<td>THE SEVEN COMPONENTS OF A WORLD VIEW</td>
<td>.........................................................................................</td>
<td>67</td>
</tr>
<tr>
<td>A SCIENCE OF COMPLEXITY AND CHANGE</td>
<td>.........................................................................................</td>
<td>69</td>
</tr>
</tbody>
</table>

**PART II: THE CLASSICAL WORLD VIEW** .................................................................................. | 73 |

| FATALISM AND THE ETERNAL RETURN | ......................................................................................... | 73 |
| LIFE AND DEATH IN AGRICULTURAL SOCIETY | ......................................................................................... | 73 |
| THE BIRTH OF THE GODS           | ......................................................................................... | 76 |
| CYCLICAL TIME                   | ......................................................................................... | 79 |
| THE CLOCKWORK UNIVERSE          | ......................................................................................... | 83 |
| FROM LAWS OF GOD TO LAWS OF NATURE | ......................................................................................... | 83 |
| ANALYSIS: MAKING DISTINCTIONS   | ......................................................................................... | 86 |
| CAUSALITY: CONSERVING DISTINCTIONS | ......................................................................................... | 90 |
| THE GHOST IN THE MACHINE        | ......................................................................................... | 94 |
| INDUSTRIAL SOCIETY              | ......................................................................................... | 98 |
## PART III: THE EMERGENCE OF A NEW WORLD VIEW ............................................... 104

### THE BREAKDOWN OF CLASSICAL MECHANICS ......................................................... 104

- The Heat Death of the Universe ............................................................................. 104
- The Theory of Relativity ......................................................................................... 109
- The Uncertainty Principle ....................................................................................... 114
- Chaos ...................................................................................................................... 118

### WHOLE, PARTS AND SYSTEMS .............................................................................. 125

- The Wholeness of Nature ....................................................................................... 125
- The Systems Approach ........................................................................................... 127

### MIND OUT OF MATTER ....................................................................................... 132

- Measuring Information .......................................................................................... 132
- Cybernetics: the Science of Purpose ...................................................................... 137
- Artificial Intelligence .............................................................................................. 141
- Construction and Deconstruction ......................................................................... 144

### FROM BEING TO BECOMING .............................................................................. 150

- Self-Organizing Systems ....................................................................................... 150
- Natural Selection ..................................................................................................... 155
- The New Darwinians ............................................................................................... 157

### THE NEED FOR INTEGRATION ........................................................................... 162

- Failed Unifications .................................................................................................. 162
- The Principia Cybernetica Project .......................................................................... 165

## PART IV: FIRST PRINCIPLES ................................................................................. 169

### THE DIFFERENCE THAT MAKES A DIFFERENCE .................................................. 169

- Ockham’s Razor ..................................................................................................... 169
- Order and Disorder .................................................................................................. 175
- Complexity ............................................................................................................... 180
- Bootstrapping the Universe .................................................................................... 187
- Causality: Micro or Macro? ...................................................................................... 191

### THE TRIAL-AND-ERROR METHOD .................................................................... 197

- Variation and Selection ......................................................................................... 197
- The Ratchet Effect ................................................................................................ 200
- Maxwell’s Demon .................................................................................................. 207
- Order from Noise .................................................................................................... 210
- The Stepping-Stone Principle ............................................................................... 216

### FITTING IN ........................................................................................................... 223

- The Survival of the Fit ......................................................................................... 223
- Fitness and Causality ............................................................................................ 226
- Self-Organization or Selection? ............................................................................ 230
- Co-Evolution: Fitting Together ............................................................................ 235
- Feedback as Fuel .................................................................................................... 238
- The Cycle of Life .................................................................................................... 241

### THE GROWTH OF COMPLEXITY .................................................................... 246

- Closing the Loop ................................................................................................... 246
- The Architecture of Complexity ............................................................................. 250
The book I present here proposes a survey of the new scientific world view. This view is being developed in a variety of new approaches centering around complexity, self-organization, evolution, systems and cybernetics. On the one hand, the book will review in a simple and understandable language the basic new ideas developed in various 20th century sciences, from quantum mechanics and molecular biology to artificial intelligence and complex adaptive systems. On the other hand, it will try to integrate these different concepts and principles into an encompassing view of the world in which we live. This broad, philosophical framework should provide a 21st century answer to the age-old questions: Who are we? Where do we come from? Where are we going to?

The basic principle of this new world view is that complex organization evolves spontaneously. In other words, any chaotic, unstructured universe will tend to develop highly organized phenomena, including matter, life, intelligence, culture and society. Self-organization is the essence of evolution. I hope to convey the same sense of awe and wonder for this phenomenon that I have experienced time and time again when observing the appearance of organization in a seemingly chaotic environment.

Perhaps a most striking example is the gradual emergence of a natural landscape out of an old industrial terrain flattened by bulldozers. At first, the terrain is just a large muddy basin, some pieces of brick and rusty metal, reminders of a destroyed factory building, sticking out between the deeply cut-out tracks left by the tires of heavy machines. The first transformation is the smoothing of the landscape by wind and rain, the gradual erosion of the sharp tracks, sand piles and holes left by the bulldozers. Then the mud starts acquiring greenish colors, caused by algae and mosses. Soon, some grasses and herbs, such as chamomile and dandelion, take root. After about a year, the plain is mostly covered by a grassy layer dotted with small flowers, which attract butterflies, bees and a multitude of flying and crawling insects. Then, different seeds of bushes and trees start arriving, the lightest ones, such as willow, birch and poplar first. Dense thickets and a few isolated trees appear locally. They quickly become a refuge for rabbits and birds. In the meantime, erosion continues and the rusting metal virtually disappears. The bricks become soft and crumble under the effect of rain, heat and frost. Eventually they get covered by new plants. If you wait 20 years or more, the former industrial wasteland has changed into an open wood, buzzing with animal and plant life. I see this transformation
as a metaphor for the evolution of the universe: the gradual and spontaneous creation of order out of chaos, leading to ever more complex organization.

Several authors, such as Paul Davies, have similarly pictured the world as a self-organizing system. They see the universe unfolding from an undifferentiated state into a complex, organic whole. Their preferred metaphor is an embryo, which develops from a mere seed through different phases to a fully grown individual. Such a process assumes some kind of plan or blueprint that guides the process of development. However, this does not explain where that blueprint might have come from. Since science has found no trace of anything resembling an evolutionary plan, the only recourse is to postulate mysterious, supernatural forces that steer evolution to some predetermined goal. But this again begs the question where these forces come from. In the end, nothing has really been made any clearer. Moreover, such a philosophy ignores the inherent unpredictability, creativity and novelty of the evolutionary process.

The originality of my approach is that I will try to show that no blueprint is needed to explain the emergence of complex systems out of simple components. If the analysis of the underlying mechanisms is carried through deep enough we get to a level where the principles become so simple and obvious that they do not need further explanation.

A number of evolutionary biologists, such as Richard Dawkins, follow a similar approach. They try to show that the Darwinian principle of natural selection is sufficient to explain the evolution of all complex life-forms. However, their explanations are restricted to purely biological systems, and thus fail to explain complexity on the levels of matter, mind and society. Moreover, they tend to reduce all creation of novelty to changes in gene frequencies because of selection by an external environment. The latter position is now challenged even by biologists as being unduly limited, since it fails to appreciate the importance of phenomena such as self-organization, symbiosis and co-evolution.

My view of evolution synthesizes the principles of natural selection and self-organization. I apply these principles to the evolution of systems at all levels: from particles, atoms and molecules, through cells, plants and animals, to people, culture and society. This vision will be presented in the form of a story, recounting the history of the universe from its beginning until now. Telling that story is a bit like reconstructing an old legend: many pieces of information about what happened have been lost, and different authors give in general different interpretations of the remaining data. My work as a historian of the universe then consists in gathering as many data as possible, trying to put
them together like the pieces of a puzzle, while using my own insights to fill in gaps, to resolve inconsistencies, or to choose between mutually exclusive scenarios.

The resulting picture, by necessity, will be personal. It cannot claim to represent consensus, let alone objectivity. Yet, I have tried to remain as objective as possible, and to take into account as much as possible the consensus existing in the scientific community. As such, this work is an attempt to reflect as clearly as possible the implicit world view of contemporary science. Although I am a scientist by education, by character I am rather a philosopher attracted towards practical problems. Therefore, I have been led less by technical precision and comprehensiveness, than by the motivation to convey a sense of what all this really means. Rather than providing lists of all elementary particles that take part in the creation of matter or of the main chemical compounds involved in the origin of life, I try to make the readers understand the underlying mechanisms. I focus on the “what”, “why” and “wherefore”, while paying less attention to the precise “where”, “when” and “how”. By formulating these ideas in a sufficiently general way, while illustrating them with plenty of concrete examples, I try to create a really deep understanding. I hope that afterwards the reader would be able to apply these concepts in his or her personal life, as an aid to solve problems, to anticipate future developments, or to make decisions between alternative courses of action.

Let me conclude this introduction with a quick survey of the book’s contents. The first major part of the book, “A World out of Balance”, sketches the fundamental transitions our present society is going through. It is meant to attract the attention to the importance of complexity and evolutionary changes for our every-day life. It then motivates the construction of a new world view that could guide us through these turbulent times.

The second part reviews the main principles underlying past pictures of the universe. The world view characterizing ancient civilizations and agricultural societies centers around the idea that an individual has very little control over his or her fate, and that nature is unpredictable, except to the degree that there are recurrent cycles. Newtonian science and the industrial society produce a very different picture, in which the world’s evolution is completely deterministic, and can be predicted by linear and reductionistic methods. The accompanying mechanistic science gave humanity an unprecedented control over nature, but produced many harmful side-effects.

The third part sketches the gradual erosion of the Newtonian view during the 20th century, because of scientific revolutions such as relativity theory, quantum mechanics and the theory of chaos. It then discusses the emergence of new ideas in various disciplines. Most importantly, these new philosophies replace reductionism by a focus on
wholes, systems and interactions; they reintroduce mind and knowledge in a science hitherto restricted to matter; and they emphasize the essential creativity of evolutionary processes.

The fourth part is the most personal one. In it I try to synthesize the ideas of these different new approaches into a comprehensive framework, based on a few, easy to understand principles. These principles show how the spontaneous emergence of ever more complex systems is unavoidable, and how the mechanism of self-organizing evolution produces subsequent levels of complexity, from atoms to cells, brains and societies.

F.H. Oct. 1998
Part I: A World out of Balance

The Acceleration of Change

Scientific Innovation

Perhaps the most striking characteristic of our age is the speed of change and innovation. No day seems to pass without some major new discovery, invention, cultural revolution, or political breakthrough being announced. Everything is in a flux of endless transformation. Nothing can be expected to remain the way it was for more than a few years.

This continuous innovation is most salient in science and technology. One might argue that scientific discovery is the real motor of the process: it produces technological and cultural developments, which in turn affect people’s daily lives, the society to which they belong and finally the ecosystem at large. Let us then first try to understand how and why scientific output has increased so much.

A first observation is that there are many more scientific researchers active today in universities, research laboratories and firms than ever before. It is said that the number of scientists alive now exceeds the total number of scientists that ever lived before. This number continues to augment. More people in more parts of the world get university degrees. Of these, a larger number get involved at some stage with research: as part of a PhD program, and perhaps of the following academic career, or simply as one of the activities their business employer needs to stay ahead of the competition.

On its own, increasing employment is not a good measure of increasing production: after all, it might be that all the easy problems have been solved, and that more and more people are needed to tackle the difficult ones. At the end of the 19th century, many physicists believed that their theories were complete, and that there was nothing much left to be discovered in their field. However, the unexpected discoveries of quantum phenomena and of the theory of relativity proved them wrong. The beginning of
the 20th century was characterized by perhaps the greatest burst of scientific innovation ever. At present, innovation in physics seems again to have slowed down. To get more information about elementary particles, physicists need to build ever-larger particle accelerators, producing ever-higher energies. The largest of these accelerators extend over many miles, and have been compared in investment and complexity to the cathedrals built in the middle ages. The cost of building the next generation of accelerators is similar to the cost of the space program. Given the budget deficits of most governments, it is no surprise that many programs have been scrapped. Physicists are starting to worry whether they will ever get the chance to test some of their hypotheses.

The difficult situation in particle physics is not the common rule, though. It is true that research programs that were moving ahead swiftly sometimes unexpectedly stall. Yet, the more common phenomenon is that suddenly new areas emerge in which progress is very fast. In science, you generally cannot predict well which approaches will turn out to be successful. A common pattern is that new discoveries arise as happy side effects of inquiries looking at something completely different.

For example, the World-Wide Web hypermedia system, which is responsible for the huge popularity of the Internet computer network, was developed at CERN, the European center for particle physics. Tim Berners-Lee, responsible at CERN for computer communications, noticed that particle physicists in different parts of the world need to exchange huge amounts of data about the results of their experiments. But different people were in general using different computer systems, which had no simple way to communicate. So he set out to develop an easy-to-use, universal protocol that would allow all these computer systems to exchange information, without the users needing to worry about technical details. The resulting system was so successful that it was quickly adopted, not only by particle physicists, but by scientists in other disciplines, and finally by practically anyone using computers.

I will have more to say about the World-Wide Web later. Here, I just want to note that different fields of science progress at different paces, and that that pace can change very quickly depending on the discoveries that are made. In the beginning of this century, the fastest progress seemed to be made in physics. This gush of development culminated in the discovery of nuclear energy. In the latter half of the century, though, molecular biology and computer science seem to be far more successful in the number of innovations they produce. In the 21st century, their role might have been taken over by the sciences studying mind and brain.

Since we cannot predict which approaches will generate most discoveries, we can only use a rule of thumb: more people investigating more diverse domains will produce
more new knowledge. If a domain, such as physics, lacks opportunities, the bright young researchers will simply move to a different one. They may perhaps use their mathematical skills to study the chaotic movements of quotations on the stock exchange. Or they may try to develop computer programs that compress the amount of information in an image, so that it is more easily transferred over a wire.

There certainly is no lack of challenging new questions, or promising new investigation methods. Every problem that is solved seems to call up new problems. In that respect, scientific investigators are similar to the great explorers, like Columbus or Cook. Each newly discovered land pushes the frontier of the unknown back. Yet, it also provides a glimpse of further lands ahead, waiting to be explored by a next generation of pioneers. It is said that the really wise is the one who is aware how little he or she really knows. Similarly, the more science advances, the more scientists become aware of how much they still don’t know.

This continuous widening of the area covered by science is perhaps most clearly illustrated by the increase in the number of scientific publications. Publications are supposed to report new findings. However, scientists are motivated to publish as much as possible, even if they haven’t discovered anything really new. This is because of the “publish or perish” rule: researchers who don’t get enough articles published, are considered to be unproductive; therefore, they may not get the funding they desire, and may even lose their job. The resulting surge in the number of articles submitted for publication is kept in check by a filtering system, based on so-called peer review. Reports submitted for publication to a scientific journal are read and evaluated by scientists working in the same specialty. In principle, the journal accepts only the best papers, which satisfy the highest standards of thoroughness, clarity and novelty. The higher the standing of the journal, the more selective it is. Rejection rates of over 95% aren’t uncommon. In practice, the evaluation by peers is not a watertight system. On the one hand, researchers often manage to publish basically the same results in different journals. On the other hand, the really original, non-conformist ideas are often unduly rejected, because none of the evaluators is competent to judge about an approach very different from their own. Still, the system works rather well in practice, and the number of accepted publications seems to reflect the number of new results relatively well. The creation of new journals reflects the emergence of new domains. Articles that previously had difficulty being accepted and that were scattered over various traditional journals may now fit nicely together in a journal led by a sympathetic board of editors.

If we look at the total number of scientific papers published, and at the total number of scientific journals we see a very quick increase. The number of publications
doubles about every fifteen years. This kind of growth pattern, characterized by a constant time needed to double (or triple or quadruple) a quantity, is very common. It is called exponential growth. It leads to an explosive development: now you have one, next time you have 2, then 4, then 8, then 16, 32, 64, 128, ... After ten time steps, you have gone from 1 to 1024. However, it takes only one more step to add another 1024, and reach 2048. And another one to add twice that amount and reach 4096. The amount added at each time step becomes larger and larger, the sign of a true acceleration.

This pattern is typical of self-fuelling growth: the more there already are, the more new ones will be added. In other words, the increase is proportional to the number already present. For example, the more people there are, the more children will be born. But the more children are born, the more adults there will be to have further children. This leads to the population explosions we know so well.

A similar mechanism seems to drive scientific production: the more new results are published, the more points of departure there are for making further discoveries. On average, we might say, every new result produces two further new results within the next 15 years. It is as if every discovery is the child of a previous discovery. After a few years, the child has become an adult, and is ready to give birth to its own children. Now, when does a scientific discovery reach “adulthood”? A discovery is ready to spawn new discoveries once a sufficient number of scientists have had the chance to study it, to think through its implications, to test its usefulness, to apply it in a different domain, or simply to combine it with other ideas and see what comes out. How quickly this happens depends on different factors.

The most immediate limiting factor is the speed of publication. The publication process we sketched takes a lot of time. A typical sequence is the following. A scientist writes up the results of an investigation in the form of a paper, submits it to a journal, waits for the referees to send in their reports, rewrites the paper to take into account critical comments, resubmits the corrected text, waits for the final approval, waits for the printer’s proof, corrects typesetter’s mistakes, sends back the corrections, and finally waits for the issue of the journal to be printed and distributed to libraries and institutions. The whole process typically lasts two years or more. That is in the fortunate situation where the paper is accepted for publication. In the more common case where the paper is rejected, the author will generally resubmit a reworked version to one or more other journals, until one of them accepts it, or until she gets convinced that the paper is not publishable. Some of my own papers have lingered like that for eight years until they finally got published (and started receiving very favorable reactions!).
To these publication delays we must add the time it takes for a scientist who has a new idea, to investigate it and gather supporting data, so that she can prepare a convincing case for an article. This generally takes several years. We must also take into account the time needed by other scientists to find out about the new idea. Few researchers read a paper immediately after its publication. Usually, a researcher will only discover that a relevant paper has appeared by searching specialized indexes or databases that list all papers about a specific subject, or through references made by other researchers in their papers. Again, there is an average delay of several years between the moment an idea is published and the moment an interested scientist reads it. Only then could he start extending the idea with his own insights.

Taking all these delays together, 15 years seems about right for the average duration of a “generation”, the time a discovery needs to produce offspring. However, our earlier assumption that every discovery will engender two new results in that interval is a little too simplistic. Some parents have many children; others have none. However, it suffices that every couple has more than two children on average to see the population explode. Similarly, some scientific papers seem to have no impact at all. A study of publications in geology found that half of the published papers are never referred to by any scientist afterwards. It is as if they never existed. On the other extreme, a paper may become a “citation classic”: thousands of scientists refer to it in their papers, and thus acknowledge the influence it has had in shaping their own thoughts.

We have sketched the traditional process whereby a scientific innovation, after spreading through various channels, gives rise to further innovations. As long as there is an increasing number of scientists ready to follow the new leads, this will automatically produce an ever more quickly growing number of discoveries. However, this acceleration of scientific production is itself accelerating. There is reason to believe that the 15-year period needed to double the number of results is becoming shorter. How that can happen will become clear when we examine the interactions between science and technology.

Technological Acceleration

The effect of scientific innovation on technological progress is fairly straightforward. A new scientific idea makes us understand how something functions. In general, this means that we can now predict what will happen in which circumstances. We can then choose those circumstances that make things happen the way we want them to happen.
“Technology” is simply the knowledge that tells us how to make things work the way we want.

For example, it has recently been discovered that many ulcers are caused by bacteria living in the stomach (and not by stress, as the earlier theory suggested). This means that in general ulcers will be absent in circumstances where those bacteria are absent. Therefore, if we want to prevent or cure ulcers, we simply need to find a way to eliminate the bacteria. Some simple experimentation with different cocktails of antibiotics has led to a cure, which is effective in 95% of the cases. Thus, we now dispose of a practical “technology” for healing ulcers.

The distinction between “basic” and “applied” sciences is much less relevant than people tend to assume. “Pure” or “basic” science investigates problems for the sake of understanding itself, regardless of possible applications. However, basic science will sooner or later produce insights that can help us to solve practical problems. For example, Einstein’s famous formula equating matter and energy followed from his theory of relativity, a very deep and abstract set of ideas, apparently without any technical applications. However, just a few decades later this same formula was applied to calculate the amount of energy that could be produced by a nuclear reactor. This marked the advent of atomic energy. In the 1950’s, Watson and Crick unraveled the structure of the DNA molecule, which carries our genetic information. This was a truly magnificent achievement, offering a first glimpse of the basic organization of living matter. Few people at the time would have guessed that hardly twenty years later this insight would have led to techniques for manipulating the genes of plants and animals. This opened the field of biotechnology, one of the hottest domains at present.

There are plenty of examples of unexpected applications. Linguistic theories about grammar and semantics are now being used to make computers understand everyday language. Without the puzzling insights from quantum mechanics, a theory describing the behavior of very small particles, we would never have developed an electronics industry. Philosophical investigations of mind and knowledge support the development of intelligent computers. Number theory, one of the more arcane parts of mathematics, studies questions such as how many large prime numbers there are. This has led to reliable methods for encrypting information. These methods now make it possible to do secure business transactions over public computer networks.

These examples illustrate a recurrent pattern. The most innovative applications usually come from research that did not directly focus on a practical problem. When you aim at a concrete target, you will pass by all avenues that don’t seem to directly bring you closer to that target. However, by exploring those side streets and back-alleys you might
discover riches that you never knew existed. Generally, you will get many more results by just following hunches, or explore anything that looks interesting. Targeted research is only useful if the road to be traveled is fairly straightforward, that is, if you know which methods and techniques will get you where you want to go. But a more free-spirited exploration may lead you to the conclusion that your initial goal was not really worthwhile, and that there are much more important things to be found around the corner.

Whatever the philosophy of investigation, basic or applied, targeted or free roaming, if the results can be used to make life better in some respect, they will be used. “Better”, of course, is a subjective judgment. Many people will not see nuclear energy or genetic manipulation as improvements. The moral dilemmas are most obvious for discoveries with military applications. But there is a more practical sense of “better”, about which almost everyone agrees: doing more with less. If a new method or technology can get the same result with less effort, someone is bound to introduce it. (That is, unless nobody would like to get the result in the first place.)

If you aren’t inclined to switch over to the new technique, your competitor will, and you will have to follow suit. Otherwise, your competitor will gain an edge, producing more than you for the same investment. This competitive drive leads to a frantic chase for more efficient methods, for “optimization” or “rationalization” of existing systems. It pushes researchers to develop ever more powerful technologies. The visionary designer Buckminster Fuller coined the word “ephemeralization” for this drive to progressively do more with less. Gradually smaller amounts of materials and effort will accomplish more and more useful functions.

Diminishing effort means in the first place diminishing the time needed to get the desired result. Why spend days on a task that can be performed in hours? You can use the time gained to produce more of the same, to tackle different problems, or simply to relax. Thus, technological innovation will lead directly to a speeding up of existing processes. With improved technology, houses will be built, motorcars will be assembled, food will be produced, and diseases will be cured in less time than it used to take.

This increased speed is most clear in travel and transportation. In pre-industrial societies, people moved by walking, on horseback, in horse-drawn carriages, or by ships driven by wind or rowing. The typical speed was a few miles per hour. This changed radically with the introduction of the steam engine, first in ships, then in locomotives and primitive automobiles. These first motorized vehicles reached tens of miles per hour. The next major jump was the invention of aircraft, moving at hundreds of miles per hour. Finally, space ships travel at thousands of miles per hour. Once they have left the Earth’s atmosphere, their further acceleration is limited only by the speed of light, the absolute
limit according to physics. In a mere 200 years, maximum speed of movement has increased by several orders of magnitude.

Velocity has continued to augment within each major category too. Ships, trains, cars and planes move much more quickly now than they did 50 years ago. Although there are practical limits on the maximum speed of each type of vehicle, the average speed of transport continues to increase, thanks to better roads and traffic infrastructure, and more efficient methods of navigation, loading and unloading. The net effect is that people and goods need a much shorter time to reach any far-away destination. In the 16th century, Magellan’s ship needed two years to sail around the globe. In the 19th century, Jules Verne gave a detailed account of how to travel around the world in 80 days. In 1945, a plane could do this trip in two weeks. Present-day supersonic planes need less than a day. Satellites circle the planet in one hour.

More important than speed records is bulk transport: the amount of goods moved per unit of time. In Magellan’s time, only precious, small volume goods, such as spices, china, and jewellery, were transported over large distances. The duration and risks of the travel were simply too large. Presently, we find it normal to get most of our oil, coal, ore, and other raw materials from thousands of miles away. Giant container ships, freight trains and highways have made it economical to continuously move billions of tons around the world.

Not only matter, but energy is distributed more and more speedily. Pipelines and tankers carry oil and natural gas from other continents. For shorter distances, the electrical grid system is the most efficient, delivering power to industry and households whenever it is needed. Presently, the national power grids of neighboring countries are starting to cross-connect, creating an international network. Thus, a factory in Denmark can receive instantaneous energy from a power station in France, or from wherever there is a high capacity available.

The acceleration is even more striking for the distribution of information. In pre-industrial times, people communicated over long distance by letters, carried by couriers on horseback. We can estimate the average speed of information transmission by counting the number of characters in a message. Technically, one character corresponds to one byte, or 8 bits, of information. If we count an average letter to contain 10,000 characters, and assume that a journey on horseback to a neighboring country takes one month, we get a transmission rate of 0.03 bit per second. The first major revolution in communication technology was the invention of the telegraph in the 19th century. It could transmit a signal virtually instantaneously. However, it would still take quite a while to transmit an extended series of signs. If we estimate that it takes a little over two
seconds to punch in the Morse code for one character, we get a transmission rate of 3 bits per second. This slow speed explains the typical “telegraph style” of messages sent this way: it pays to minimize the number of words and characters that need to be transmitted. The first data connections between computers in the 1960’s would run at speeds of 300 bits per second, a dramatic improvement. Present-day modems, through which computers can communicate over telephone lines, reach some 30,000 bits per second. However, the most powerful long distance connections, using fiber optic cables, transmit some 300 million bits per second. In a mere 200 years, the speed of information transmission has increased 10 billion times! And this is just the beginning: experiments with even higher transmission rates are going on.

The acceleration of data transmission is not only much larger but also more significant than the acceleration of transport. That is because it boosts all further scientific and technological progress. Swift transmission of information eliminates the major bottleneck of scientific innovation: the long delay between the moment an idea is written down and the moment it is read by another scientist. With the present electronic networks, a researcher can make a document, including all relevant data, illustrations and references, available on a public computer, announce its availability to hundreds or thousands of scientists working in the same domain, and start getting their reactions within the next few hours.

Of course, part of the reason why the traditional publication process takes so long is because of the time needed for evaluation, to make sure the ideas and results are sound. But this too can be speeded up using network technology. A new idea proposes that people would attach an electronic “seal of approval” to the documents on the network they find worthwhile. If a document has gathered a sufficient number of such seals from well-respected individuals or institutions, scientists might consider it to be reliable, just like a paper that has passed the refereeing process for a scientific journal.

Still, since it is much easier and cheaper to “publish” a document on the network than to publish it on paper, meager or controversial findings need not be rejected. Making them available on the net, even without “seals of approval”, can still be useful. A hypothesis that lacks evidence is perhaps all that is needed by a scientist who has evidence but no theory to explain it. Similarly, scientists often perform “failed” experiments, where the data don’t fit any of the proposed hypotheses. Most of these results are never published. Yet, if these data were available on the net, someone might be able to interpret them in a different way and derive a new theory.

Better communication media diminish not only the delay between a discovery and its use by other scientists, but the delay between an invention and its acceptance in the
marketplace. Studies have shown that the gap between the development of a new product and its diffusion throughout society has been steadily decreasing. The first patent for a typewriter was issued in 1714, but it was not until the end of the 19th century that typewriters were commonly used. For inventions introduced in the beginning of this century, such as vacuum cleaners and refrigerators, it would typically take some thirty to forty years before they would reach peak production. Recently, new technologies such as CD players or video recorders have swept through society in a mere ten years.

The speed of diffusion is understandably highest in the domain of communication technology itself. I was a close witness to the epidemic spread of the World-Wide Web (WWW) communication system, whose invention by Tim Berners-Lee I mentioned earlier. The first sketchy ideas were written down in 1989. 1991 saw the first working prototype. At the end of 1992, I heard about its existence on the Internet, and started eagerly to set up my own WWW system. It would become the first one in my country, Belgium. By the time it was in working order, in July 1993, some two hundred similar WWW “hosts” had appeared on the Internet. One year later, the number had increased to several thousands. At the moment I write, hosts number in the hundreds of thousands.

At the same time, the number of users has grown beyond all expectations. Initially, the only people using WWW were scientists and students at universities and research institutes. But 1995 was the year in which Internet and WWW, in the guise of the “information superhighway”, reached the public. At first, the many enthusiastic newspaper articles met with understandable skepticism. Early that year, an acquaintance tried to convince me that “not in the next ten years would ordinary people start using the Internet”. I did not have to wait long to prove him wrong. One year later, publicity leaflets offering inexpensive Internet access were part of the junk mail, and advertisements sporting company WWW addresses were appearing in family magazines. Dozens of millions worldwide are using WWW now, and their number continues to expand at a staggering rate. I am still waiting for the greengrocer behind the corner to have his own WWW page, but I am sure I will not have to wait much longer.

Technological advances accelerate not just the diffusion, but the creation of new knowledge. This is most obvious for the collection of data on which theories are based. Satellite images can help scientists to understand the development of weather patterns. Electron microscopes make it possible to observe the structure of atoms and molecules. The humble tape recorder has enormously facilitated the work of anthropologists and linguists studying how people speak.

The second stage of theory development is the processing of the data. Computers make tasks such as selection, classification and analysis of data much easier. Advanced
computer programs can find subtle patterns in masses of data that are too large to be grasped by the human mind. This is called “data mining”. These patterns or regularities may express new laws. Automatic discovery of laws further accelerates the explosive development of knowledge. Computers can moreover help with the visualization of the results. They make it possible to explore the myriad consequences of a proposed theory, so that nothing is overlooked. They support every stage of the process of scientific discovery.

Computing itself is a domain where speed has increased spectacularly. Present-day desktop computers are several orders of magnitude faster than the room-size monsters of the fifties. They can make complex calculations or search through extensive data in a fraction of the time needed by earlier models. Unlike transportation, the speed of computers increases at a regular, constant rate, without apparent upper limit. According to Moore’s law, the speed of new computer processors doubles every one and a half year. This is the same type of exponential growth that we encountered for scientific publications, except that the doubling time is ten times shorter. This acceleration is mainly due to miniaturization of the electronic components: the smaller the size of the circuits, the less time electrons need to move around, and the faster the electronic signals can go through the processor.

Miniaturization is a direct implementation of the “more with less” philosophy. Smaller machines need less material, energy, and time to carry out their functions. They can be produced in larger numbers, for a lower cost. The cutting edge of research focuses on nanotechnology: a set of techniques to build structures atom by atom. It would allow the mass production of molecular size appliances. Such microscopic tools might for example be injected into the human body, where they could repair damage done by diseases, cancer or ageing. Swarms of tiny, robot like machines might be able to assemble practically any kind of material with an incredible precision. This is still science fiction at present, but progress in miniaturization is very fast.

All these developments point in the same direction: not only will scientific and technological progress continue at a breakneck speed, it will accelerate more and more. This is because innovation feeds on itself: every discovery, in whatever domain of science or technology, provokes and facilitates further discoveries. We should not be surprised if the time needed to double the scientific output has become much shorter than the fifteen years we estimated earlier. The effects, of course, extend far beyond science and technology. Let us discuss some of the revolutionary changes brought about in daily life.
**The Pace of Life**

The most visible effects of technology occur in the objects that surround us. New appliances are continuously appearing in our homes and offices: hand-held computers, mobile phones, microwave ovens, high definition TV sets, video cameras, laser printers, processor-controlled washing machines, ... Some of these fulfill wholly new functions. Because of inexpensive video cameras millions of parents have started to make home movies, featuring children, pets and neighbors. Microwave ovens have led to a quite different way of cooking, with less fat and more frozen ingredients. Other new technologies merely enhance existing functions. CD players do essentially the same thing as the older record players: reproduce recorded music.

But there is one constant: all appliances are continuously undergoing change. No single device seems to remain the same for more than a year. Last year’s model is never quite as good as the new one. Changes in the things we use imply changes in what we do. New functions create wholly new activities, like videotaping your children’s play, or “surfing” the World-Wide Web. But even minor enhancements demand a change in behavior. The new washing machine with its more sophisticated laundry programs may simplify your life in some respects, but it will also force you to rethink your old habits about which buttons to push.

The increased speed and ease of transport too have an immediate effect on the way people live. Bulk transport makes it economical to buy goods produced in faraway countries. Supermarkets and specialty shops offer the most diverse foods. A few decades ago, bananas and oranges were still considered luxury goods outside the regions where they are cultivated. Today, assortments of the most exotic fruits, like lychees and mangosteens, are readily available. Newly imported products provide the consumer with an ever-increasing choice. They lead to changes in diets and habits, as when traditionally wheat-eating populations start consuming large quantities of rice, or when beer drinkers increase their consumption of wine.

Easier movement affects people more directly. In the Middle Ages, few individuals would ever venture more than twenty miles from the village where they were born. Today, many people travel thousands of miles every month. In the last thirty years, taking a plane has become as ordinary as taking a train. An increasing section of the public finds it normal to pass yearly vacations on a different continent. Moreover, companies, institutions, schools and universities are increasingly sending employees or students to other countries. They can thus gather experience, or represent their organization abroad. This means that people no longer need to learn about exotic climates.
and cultures from books or movies; they can now experience these strange environments firsthand. This creates a different, broader outlook. It makes people more aware of the similarities and differences between their own culture and that of the rest of humanity.

Even the ones that stay at home will more often come into contact with foreigners. Emigration has become much easier. Especially the affluent countries of the West come under pressure from an increasing stream of would-be immigrants. Often these immigrants try to escape poverty and political oppression at home. In other cases, the immigrants belong to the higher, well-educated classes, which are lured by more advanced education systems and richer opportunities abroad. The resulting “brain drain” may weaken the home country, while unjustly strengthening the country of immigration. In any case, both the immigrants and the local people, whose neighbors they become, need to adapt to a drastically changed situation.

On a more local scale, easy travel means that people are more likely to work in a different city than the one they live. They are more likely to visit many different places for their work. They will shop in one city and go out in another one, while living in the countryside or suburbs. They will also more frequently move houses. All this adds to the increasingly broad range of situations, persons and environments that people come into contact with.

The ever more powerful communication media reinforce this trend. Since the advent of television, the world has become a “global village”. People now see events happening in other parts of the world with the same vivid details as if they were present themselves. Any new development, whether it concerns a scientific discovery, a political revolution, an earthquake or the outbreak of a disease, is instantaneously broadcasted over the globe. Following the “news” has become a universal addiction.

Private communication too has transcended the limitations of space and time. Whether you are in a plane flying above the clouds, on a ship in the middle of the ocean or in farthest reaches of the Siberian forest, you can pick up a phone and immediately get in touch with someone on the other side of the world. The rapidly developing computer networks offer an unprecedented control over this torrent of information. You no longer need to wait for the 9 o’clock news to get the latest information about the subject you are most interested in. Neither do you need to wait until your loved one gets near the phone to send him or her an urgent message. Smart programs will find information for you, or deliver your messages, wherever the source or recipient are located.

Sophisticated networks even allow you to work in different places without actually traveling. Business leaders in different countries hold videoconferences to discuss marketing strategy, without getting out of their office. Speculators buy and sell
huge amounts of stocks through their computer terminals. An increasing number of people are becoming telecommuters: instead of driving the car to the office every morning, they stay home, and do all their communications and transactions through their telephone and computer link. They thus lose much less time traveling, while having access to the whole world, rather than just to the people at the office.

The growing ease of travel and communication has a profound effect on the people themselves. They are constantly bombarded with new ideas and facts, new lifestyles, cultures and fashions. They try new sports or hobbies, such as Tai Chi Chuan gymnastics, African dancing or Japanese flower arrangement. They start listening to Arab “Rai” music, Gregorian chant or the newest computer-produced dance music. They develop an interest in Third World development, endangered species, or Eastern Europe’s transition to a market economy. While the world around them is changing ever more rapidly, they feel they need to change their old ideas and habits in order to stay abreast.

To keep up with the new technologies and developments, at home or at work, people need continuing re-education. Training programs, seminars for career development, educational CD-ROMs and “How to” books become increasingly popular. Different kinds of psychotherapies, relaxation techniques and mental health programs even promise changes of personality. “Are you anxious, shy, afraid to express your feelings? Do you avoid contacts with people you don’t know? Our program will teach you how to be assertive, and boost your self-confidence!” New medical developments constantly advise us to change our life style: to do more sports, to eat more of this and less of that, to avoid certain drugs or chemicals, to take additional vitamins, minerals or antioxidants, to stay inside during periods of high ozone concentration, to avoid exposure to the sun, etc.

Changes in individual life bring about changes in relations between people. Frequent travel means that we come into contact with more people, but also that it becomes more difficult to maintain existing contacts. Friendships have a tendency to become fleeting: a close relationship needs much time to build and keep up, and this is difficult to achieve if people move every few years. The difficulty of maintaining contact is to some degree compensated by the ease with which we can start new relations. The communication networks now make it possible for people to get to know each other—even to fall in love—without ever having been in the same place. Friendships are started on the basis of common interest rather than geographical nearness. However, just because it is so easy to get interesting new acquaintances, people tend to have many rather than few. This means that each acquaintance will only get a fraction of the attention that previously was reserved for the few good friends.
This short attention span affects even the ones that are most close to us. Husbands and wives, parents and children are so busy developing their own personality, career or relationships that they easily drift apart. There are so many things to do, persons to meet, skills to learn, that people just lack time to spend with their family. Moreover, the acceleration of personal development means that the person you live with now may be quite different from the one you married ten years ago. You yourself may have developed interests, values and attitudes very different from the ones that seemed to match with your partner so well. It is not surprising then that marriages fall apart much more frequently and that the divorce rates continue to increase in all parts of the world. In fact, in a world characterized by transience, many people tend to question any arrangement based on permanence. This explains the increasing reluctance to get married at all. In many countries, more children are now born out of wedlock than ever before.

The general effect of all these changes is a staggering increase in the pace of life. Everybody and everything just moves much faster. Losing time has become a cardinal sin. This follows directly from the principle of ephemeralization: if more can be produced in a given time, then the value of time itself increases. It does not matter much if you lose an hour when doing a task (like assembling a car) that takes several days to complete. However, if the same task could be performed in half an hour, one hour less would mean two cars less to come out of the factory, and a considerable loss of money. Even if you decide for yourself that production is sufficient, and can stand a slightly slower pace, you could only diminish the tempo if all people around you do the same. Your competitors will gain directly if you work more slowly than them. Because of ephemeralization a small difference in speed will result in a considerable difference in productivity. But even your colleagues and collaborators will put pressure on you to do things more quickly. The time they spend working at a slower pace with you is time they could have used for others things if you had reacted more quickly. The result is that people all around the world are put under pressure to achieve ever more in ever-shorter time. The pressure is strongest in the technologically most advanced countries, where ephemeralization is most pervasive. This explains the surprise at the slow pace of life experienced by visitors to less advanced countries, where “mañana” (tomorrow) is always soon enough.

Global Change

The changes brought about by technology affect not only individuals, but organizations, society, and the world at large. Easier movement of information and goods across
national boundaries has created a global market, in which everyone competes with everyone. The drive to optimize production, the increased competition, and the need to continuously develop new products put a heavy burden on companies. Reorganization has become a permanent process. It may include downsizing, mergers, expansions, cutting middle management, creating special task groups and adapting to local cultures. Management thinkers preach flexibility and innovation as the highest virtues. They continuously propose new organization models to better cope with a changing society. The employees are at the mercy of these unpredictable waves of reorganization. They constantly move from one department or firm to another, are promoted or demoted, hired or fired. By the time they get settled in a particular job, they need to be retrained for something quite different.

Many firms don’t survive the increasing pressures. The record highs for bankruptcies, however, are compensated by record highs for new firms starting up. Every new technology or social trend creates a new opportunity, a “niche” in the market, which may bring fortune to the entrepreneur clever enough to propose the right product at the right time. Though many are called, few are chosen. But those who fail will simply try again in another domain. As a result, the average life span of companies has become much shorter.

The turmoil in the corporate world is mirrored in the world of politics. Globalization puts pressure on the old boundaries separating countries or groups of countries. Some countries, like Yugoslavia or the old Soviet Union, have fallen apart, while others, like those of the European Union, are busy eliminating their mutual borders. On the one hand, the disappearance of the old opposition between the Eastern and Western political blocks has created new opportunities for transnational cooperation. On the other hand, an increasing number of peoples demand regional autonomy, threatening to sever ties with their national governments. As a result, maps of the globe need regular redrawing. The “new world order” is still a far-away dream. Although the political struggle between communism and capitalism has all but ceased, the contradictions within the victorious “capitalist” ideology loom large. Tensions between ecological and economical concerns, between freedom and security, between competitiveness and solidarity, pull society in opposing directions. An “end of history” is certainly not in sight.

The changes aren’t limited to the social sphere. The acceleration of industrial production has transformed the physical environment. Because of ephemeralization, the volume of raw material, energy and goods being processed has become so large that it affects the state of the planet. Large parts of the Earth’s surface are covered by human-
made structures: buildings, highways, airports, ... The illumination of roads and cities during the night is visible from outer space. Forests the size of countries are being cleared in months. Natural resources, like ores, minerals or oil, are being exhausted at increasing rates.

Industrial production and consumption of energy and goods create side-effects so large that they shift natural equilibria. A case in point is the destruction of the ozone layer, which shields the Earth from ultraviolet radiation, by gases released from industrial processes. The greenhouse effect provides an even more drastic example of the effects of human intervention. The burning of fossil fuels for heating, transport and energy production increases the concentration of carbon dioxide in the atmosphere. This gas, together with a few other “greenhouse gases”, makes it more difficult for heat from the sun to escape back into space. This leads to a gradual warming up of the Earth. Average temperatures have noticeably increased since the beginning of this century. Larger increases are predicted for the next century, leading to radical changes in climate patterns. Wet regions may become dry, while dry regions may be flooded by rains. The level of water in the oceans is likely to increase to such a degree that many ports and coastal towns will be submerged.

Biological changes accompany the physical transformations. The on-going replacement of virgin territories by cities and agricultural land drives out many native species of plants and animals. On the other hand, travel and agriculture have introduced many species to regions where they are foreign. In some cases, like the rabbits in Australia, they adapt so well that they displace the indigenous species. The combination of these factors with pollution and hunting produces record extinction rates.

The disappearance of natural species contrasts with the creation of artificial breeds through genetic manipulation. New tomato races are immune to diseases. Genetically manipulated bacteria produce human insulin. Human intervention creates new breeds in more indirect ways too. In a classic example of Darwinian adaptation, moths in 19th century England became darker, so that they would be less visible on soot-stained walls. Nowadays, insects become resistant to the pesticides that are used to exterminate them. Bacteria develop resistance to antibiotics. Radioactivity and mutation-inducing chemicals accelerate the spontaneous appearance of new strains.

The changes induced by technology affect every aspect of our world. Changes produce more changes, snowballing into an avalanche of transformations that no one can stop. Although in most cases the intentions leading to innovation are good, the second and third order effects can be extremely harmful. While we don’t seem to have control
over the rate of change, we may still have some control over its direction. The next chapters will address the question in how far we can steer change in a positive direction.
Losing Control

The Adaptation Syndrome

Whether things evolve in a positive or in a negative way, change itself constitutes a problem. It is intuitively obvious that too much change will put a strain on people and organizations. This intuition can be examined more extensively. The futurologist Alvin Toffler has made a detailed study of the acceleration of change and its psychological effects. He concluded that it could lead to a set of severe physical and mental disturbances, which he called the “future shock” syndrome. Just like people exposed to war or disaster may develop a nervous breakdown (“shell-shock”), people exposed to the rapid changes of modern life may develop a state of helplessness and inadequacy, where they simply cannot cope with new developments anymore.

In most cases, the effects of change are less extreme. Yet their presence is very real. They can be understood by noting that people, like all living organisms, are adaptive systems. Such systems survive by maintaining their equilibrium in the face of changes in their environment. They compensate external disturbances by internal changes, so that the equilibrium is preserved. For example, if a warm-blooded animal is exposed to cold, its body will generate more heat, in order to keep up the ideal temperature. If it is exposed to heat, it will start to sweat. Specific disturbances lead to specific counteractions or adaptations that restore the equilibrium. The more changes occur in the environment, the more actions the organism will need to take, and the more time and energy it will spend adapting.

However, things get more complicated when the system does not recognize the perturbation. In such novel situations, the organism does not a priori know how to react. In studies of animals and people, psychologists have observed a recurrent pattern of behavior, which they call the “orientation response”. It occurs when an organism is confronted with something new in its environment, and tries to establish what action it should take. For example, when a dog notices a strange creature, like a toy robot, entering its familiar surroundings, it will prick up its ears, turn its head towards the intruder, approach it prudently, and perhaps sniff to catch its smell. All these responses are aimed at gathering more information about the situation at hand. Similarly, when we are
suddenly confronted with something unusual our senses will be stimulated: the pupils of the eyes dilate, hearing becomes more acute, and the muscles involuntarily turn our body towards the source of the disturbance.

If the disturbance is not immediately recognized as harmless, the body gets ready for a possible counteraction by making more energy available. The heart starts to beat faster. The lungs pump in more oxygen. Muscles are tensed. Blood is diverted away from places where it is not urgently needed, like the digestive system. Sweat appears on the skin. Different hormones, such as adrenalin and ACTH, circulate throughout the body, initiating a host of physiological processes. This set of interconnected reactions prepares us for extreme eventualities, where we might have to fight or run away from danger. It is called the “stress response”, or, more precisely, the “General Adaptation Syndrome” (GAS).

The problem is that the GAS will be triggered by every situation that is novel, unexpected or uncertain. If the novelty eventually turns out to be innocuous, the GAS will subside, but that may take a very long time to happen. In the meantime, the body is subjected to stress, which disturbs all normal physiological processes. Most importantly, during the GAS energy is taken away from background processes like digestion, fighting infections, or repairing damaged tissues. This leads to a general weakening of the body and a lowered resistance. The characteristic diseases caused by stress are well known: high blood pressure, heart failures, stomachaches, etc. These negative effects are triggered by all situations characterized by a recurrent or prolonged GAS. Therefore, we may expect that they will be experienced more frequently by people confronted with continuous novelty.

Researchers have indeed found a positive correlation between illness and change. The “Life Change Scale” is a psychological tool which measures the amount of change experienced by a person over a given time interval. The “Life Change” questionnaire asks people to mark on a list which important changes they recently underwent: move to a new home, a new job, marriage, divorce, birth of a child, death of a family member, travel, promotion, etc. These different events are scored according to their relative importance. For example, the death of a spouse carries a much higher weight than a journey abroad. The total score for a person is calculated as the sum of all changes that the person experienced, multiplied by their relative weights. This gives a good estimate of how important the total change in someone’s life has been. Using this scale, it was shown that individuals with high life change scores are significantly more likely to fall ill. More surprisingly, it turned out that illness correlates with all changes, positive (like marriage or promotion) as well as negative (like divorce or job loss).
Rapid change affects not only our physical but our mental state. The emotional reaction associated with the orientation response and the adaptation syndrome is first of all arousal. This *a priori* neutral state may develop either into a positive feeling, as when novelty elicits curiosity, excitement and wonder, or into a negative one, as when lack of understanding triggers confusion, tenseness and fear. However, the longer the GAS is sustained, the more likely it is that interest will wear off and fatigue will set in. If a person does not manage to find an adequate response to the novel stimuli, he or she will experience loss of control. Distress will be the resulting feeling.

The instinctive reaction of an animal to stressful situations falls into three main categories: fight, flight or fright. The same inherited reactions seem to underlie our negative emotions. The “fight” reaction is associated with anger and aggression. “Flight” corresponds to fear and anxiety. “Fright” is the reaction of an animal that freezes or “plays dead” in the face of danger. The corresponding human emotions seem to be apathy, depression and despair, which are all characterized by utter helplessness. Aggression is a short-term reaction, which cannot be sustained very long. Anxiety and fear, however, can be constantly present. Fear is directed at a specific, frightening target, whereas anxiety is a generalized expectation that bad things may happen. This seems the most likely response to the continuing experience of unpredictable and uncontrollable change. Apathy and depression are more likely to be the outcome of a long process of failed attempts to control the stressful situation.

Not only individuals but society as a whole is likely to undergo these negative effects of too rapid change. It has been observed that mental illness increases in societies undergoing rapid changes, such as forced industrialization. The three basic attitudes are easily recognized in current patterns of social behavior. Aggression directed at no one in particular seems to underlie phenomena like vandalism and hooliganism. Individuals running amok and shooting dozens of innocent bystanders may suffer from a more extreme version of this condition. Helplessness and despair can be recognized in the increasingly common “burn-out” syndrome, and in the ever so frequent depressions. Drug addiction may be another of its symptoms.

But perhaps the most common neurosis in present society is anxiety. This is illustrated by the record use of anxiolytic drugs, like Valium or Temesta, which suppress the typical anxiety symptoms, such as sleeplessness, worrying, irritability, tension and stomach upsets. Anxiety also shows in the many irrational fears and scares, where far-away threats trigger disproportionate reactions. For example, the 1991 Gulf War should have worried only the countries neighboring Iraq and Kuwait. Yet, the world, and the USA in particular, recorded a spectacular drop in air traffic: people got afraid to travel.
anywhere. In my own country, a continent away from the Gulf, people started hoarding basic foodstuffs, like sugar, coffee and flour.

On the socio-economic level, anxiety is apparent in the growing feeling of insecurity. Fear of aggression tends to increase more quickly than the actual crime rates. Even when economic conditions are good, governments lament the absence of the "feel-good factor". Uncertainty makes people save money for later rather than invest it now. The "consumer confidence" needed to boost sales remains elusive. The main cause seems to be the continuing threat of job loss. Although households in the Western countries have a much higher average income than twenty years ago, the turmoil in the labor market makes future earnings very unpredictable.

This lack of knowledge about what will happen next is at the core of the problem, producing "future shock" and a persistent adaptation syndrome. But perhaps this problem could be solved by giving people more information about possible consequences? Let us see how information affects this state of permanent uncertainty.

**Information Overload**

The first reaction to a novel situation, the orientation response, is aimed at gathering information. This information may help the organism decide how to react. Ideally, it allows the individual to select the appropriate action (or possibly no action at all), which will bring things back to normal. This stops the GAS, and brings about relaxation. However, information can have just the opposite effect.

Imagine that you are wondering what to do about a particular situation, and that you get advice of the following kind: if you do A, you will get result B. If B is what you want, the problem is solved. However, assume that you get a list with millions of "if ... then..." rules. Each rule spells out a different possible action, with a different result. Even if you could immediately recognize the one result that you want, you would spend an enormous amount of time and energy finding the needle in the haystack. In a more realistic situation, there would be many results that might satisfy your requirements, but none of them would be obviously perfect. So you would need to examine and compare all the different results in order to select the best one. Assume moreover that a horde of competitors are involved in the game, trying to find a better option than you in a shorter time. If they succeed, you have lost...

This nightmarish scenario provides a perhaps too recognizable picture of the way people make decisions nowadays. Information facilitates decision-making up to a point. If you get too much information, decisions become more difficult. You may ignore part
of the information, but at your own peril: the data you discard might contain just that one item that makes the difference between success and failure... Of course, there are many different kinds of problems, decisions, and types of information. Yet, the general principle is clear: people have a limited capacity for processing information.

The most direct limit is the “magical number 7”. The psychologist George Miller discovered that people can only keep 7 (plus or minus two) items at once in their short-term memory. If you plan to buy ten or more things in the grocery, you will need to write them down on a list, or you are bound to forget some. Now, remember the task of comparing millions of possible outcomes to select the best one. The magical number tells you that you will never be able to compare more than about seven results at a time. It is simply not possible to read the whole list, think about it, and decide about the best option. You can only read a few items at once, choose the provisionally best one, compare it with the next few items, and so on.

Computers can help, as they process information in some respects more efficiently than humans. Their “magical number” is limited only by the size of their RAM memory. But they have other limits. They are much less good than people at selecting information, at separating the important from the less important. On the other hand, they make it easier to collect and distribute information. Overall, computers seem to contribute to the information overload rather than alleviate it.

A primary cause of the problem is the greatly increased efficiency of the communication media. In earlier times, very little information about far away places and events would actually reach the people. Hundred years ago, nobody would have worried about a war in another continent, because nobody would have heard about it. There would be no need for decisions on which action (such as hoarding food or canceling journeys), if any, to take. Nowadays, we are bombarded with information about the most diverse events, situations or discoveries. Most of these facts have at least potential relevance for our life. Perhaps a rebellion in some far-away African country may trigger an influx of refugees to your country, or boost the value of your shares in copper mines. The discovery that the use of a certain type of vitamin decreases the chances of getting colon cancer may save your life. So, it seems worthwhile to pay attention to the news coming in.

However, most reported events have very little, if any, effects on your situation. Taking them into account just burdens your mind. How do you separate the chaff from the wheat then? Couldn’t you just pretend that the information revolution had not happened, and ignore everything that was not told to you in person? After all, you might argue, if something is really important, everybody will be talking about it. Instead of
actively selecting what is important, you would let society filter information for you. That is the way information was distributed before the advent of mass media and computer networks. It worked relatively well then, so why shouldn’t it work now? As I will argue now, technological acceleration has not only boosted the availability of information, but also the need for using it.

Runaway Processes

Overall acceleration facilitates all developments, positive as well as negative. Its general effect is similar to the elimination of friction. Normally, objects are difficult to move because friction creates a force opposing the movement. That is why carmakers and airplane designers put so much effort in aerodynamic shapes, which minimize air friction. Friction functions as a kind of resistance, which dissipates or wears off the energy of motion, and thereby slows down the movement, until complete standstill. Noise plays a similar role in information transmission: over noisy lines, parts of the message get lost on the way. For obvious reasons, communication engineers try to minimize noise, just like airplane engineers try to minimize air resistance.

The elimination of noise or friction is positive for desired movements. However, it can be dangerous when there is a possibility for unwanted movements. For example, ice as a surface produces much less friction than earth or asphalt. That is why you can reach higher speeds and sustain them for a longer time when skating than when running. However, walking on ice is much more difficult and potentially dangerous than walking on asphalt: once you start slipping there is very little to stop the movement getting out of control.

In a similar way, technology smoothens or oils all the mechanisms of change. Movements of information and matter run freely, with very little loss or resistance. But this applies to unwanted movements too. It has become much easier to distribute arms, drugs or poisonous materials, like plutonium or pesticides. Once such a movement gets started, it can develop very quickly, making it difficult to counteract it in time. For example, an infectious disease can spread much more quickly in a world where people travel frequently. In antiquity, an epidemic might need decades to reach a neighboring country. Recently, AIDS needed just a few years to spread over the entire globe. Computer viruses are a more modern variant of the same principle: the easier and faster the exchange of information between computers, the more quickly viruses can spread. The accelerated spread of information applies similarly to rumors, scares and urban legends.
The genocide in Rwanda provides a particularly poignant illustration of the associated dangers. After about a million Tutsi people were killed in a few weeks time, observers wondered how something which seemed like a spontaneous movement could have developed so quickly, leaving no time for outside intervention. It turned out that the mass murders were triggered by hate propaganda broadcasted by a regional radio station. Radio “Mille Collines” (Thousands Hills) told the Hutu farmers that the Tutsis were going to kill them, and that they could avoid being massacred only by killing their rivals first. In a primitive agricultural society like Rwanda, had a dictator ordered troops to carry out the genocide, they would have needed months, if not years, to comb out Tutsis from all the remote villages and farms. Now, the farmers in the hills were killed by their next-door neighbors, in a synchronized action, which would have been impossible without broadcasting technology.

The reduction of friction is particularly dangerous for self-reinforcing processes, like epidemics. In such processes, the more there already is of something (e.g. infected people), the faster it will grow (the more people will get infected). Such positive feedback processes typically exhibit the exponential growth pattern we discussed earlier. Exponential growth is measured by its doubling period, the time needed to double the amount (of infections, of publications, etc.). Decreased friction will lead to a decreased doubling period, and therefore to faster growth.

Imagine a new computer virus being transmitted from machine to machine via the networks. Suppose that every day it infects another computer from an already infected computer. The doubling period is one day, since after that interval there will be two infected computers instead of one. After two days, there will be four, after three days eight, and so on. If nothing is done to stop the spreading, after ten days there will be about a thousand infected computers, and after twenty days a million. Thousand infected computers seems about the stage where a directed intervention becomes likely, since there are enough cases to get a clear diagnosis of the problem, but the problem has not yet gotten out of hand. At the million infections stage it is likely that large parts of the network, and the economy as a whole, have shut down and become incapable to react. Ten days seems like a short, but reasonable time to set up an intervention. But imagine now that the computer network would be more powerful, transmitting information, including viruses, at a much higher rate. If it would be twice as fast, the million infections mark would be reached after ten days instead of twenty, making a successful intervention unlikely. If the net would be ten times as fast, two days would be sufficient to infect a million computers, and any concerted action would be impossible.
This example may be hypothetical, but the dangers of fast information transmission boosting explosive developments are very real. In 1987, on “Black Wednesday”, stock exchanges all over the world witnessed a spectacular collapse of the share prices. Though the crash was as important as the Wall Street collapse of 1929, the economy suffered remarkably little. After a detailed study of what happened, it turned out that the causes had little to do with the state of the economy. The main contributing factor seemed to be the new phenomenon of computer trading. Specialized computer programs would monitor the prices of different stocks. If prices fell below a certain value, the computer was programmed to offer the shares for sale, before they would lose even more value. But the more shares were on sale, the lower their prices became, triggering yet more selling. This development reinforced itself, producing ever more selling for ever lower prices. Since networked computers would get data about prices from all around the world, and react immediately, this led to an extremely fast, worldwide collapse of the share indexes.

The positive feedback process, where buying triggers more buying (causing a “boom”) and selling triggers more selling (causing a “bust”), was not new. What was new was the absence of friction. Normally, a speculator would need time to get information about the price, and to give the order to buy or sell. This very much slowed down the process. It would provide a delay during which speculators would have time to assess whether the change in price was caused by a change in the intrinsic value of the stocks, or just by a temporary fluctuation. In the latter case, they could decide to buy undervalued stocks, under the assumption that the downward movement would soon be reversed. In the computer-controlled market, there simply was no time for such corrective action. Shares would lose most of their value before anybody could assess what they were really worth.

The investigation of the Black Wednesday crash concluded with the recommendation that delays should be built into the computer trading programs. Such artificially added friction seems paradoxical in a world where increased speed is the highest good. Economists are already planning for the coming “superliquid”, frictionless economy, where the slightest increase in demand would be immediately satisfied by a corresponding increase in supply. The Black Wednesday example reminds us that we should be very careful when implementing frictionless mechanisms.

Most exponential growth processes are much slower than these examples. They have doubling times of the order of decades rather than days or hours. Growth of the world population is a key example. Economic growth and the associated increase in production and consumption is another one. But no exponential growth process can go on
forever. It exhausts resources at an increasing rate. When resources become scarce,
growth must slow down. Usually, it reaches a plateau where supply of resources balances
consumption. But when growth is too fast, resources may be exhausted so rapidly that
reserves are destroyed and the process collapses catastrophically. Population explosions
in the animal world often follow this pattern: when there are too many rabbits, all the
grass will be eaten, and most of the population will starve before a new equilibrium can
be reached.

Friction may be essential to avoid such overshooting. Frictional forces, being non-
linear, increase more rapidly than the movement they are resisting. They thus naturally
reduce acceleration and impose a plateau or limit to further expansion. For example,
foxes eating the rabbits may constitute a frictional force, which keeps the rabbit
population from exploding, and subsequently collapsing. Paradoxically, they may thus
maintain the rabbit population at a higher overall level than if they were left to grow
without restraint. (The actual dynamics of rabbit populations interacting with grass and
foxes is much more complicated, but you get the idea).

Similar mechanisms govern global change. Most of the important growth
processes, such as population increase, production of goods, depletion of minerals and
land, release of pollutants, etc., are exponential. The “Limits to Growth” report to the
Club of Rome has popularized the idea that such growth cannot be sustained. According
to the 1995 “State of the World” report, the present symptoms of unsustainable growth
include collapsing fisheries, shrinking forests, eroding soils, falling water levels and
expanding deserts. The economic effects include falling incomes, rising unemployment,
scarcity, and decreased investment. This leads to social problems such as hunger,
environmental refugees, ethnic conflicts, riots and breakdowns of government.

Soon, we will reach the boundaries of viability. That the trend will halt is certain.
But the question is how that will happen: with a smooth transition to a comfortable
equilibrium value, or in a catastrophic manner characterized by wild, erratic swings? The
decrease of frictional forces has made the latter alternative more likely.

**Walking the Tightrope**

The overall effect of these developments is a loss of balance. Equilibria have disappeared
or become unstable. Tiny fluctuations, caused by the most intricate and indirect
interactions, may be magnified by positive feedback and snowball into explosive
developments.
Low-friction, exponential growth affects many processes, but not to the same degree. Small differences in initial state are amplified. Gaps in development may appear between countries, companies or cultural groups. Perhaps the one has a slightly shorter doubling time than the other, or has started growing a little earlier. Such lags create tensions and disbalances. A group that is noticeably less successful than a rival group may feel discriminated. In a competitive climate, small differences in the speed of innovation or growth lead to bankruptcies. In a more slow-paced economy, a firm that lags behind its competitors may have sufficient time to develop and implement a restructuring plan. In the present economic climate, yesterday’s giants may go out of business without even getting the time to understand what went wrong. The same may apply to countries. Until the 1960’s, the communist countries managed to keep up relatively well with their capitalist rivals. Their more rigid political system did not cope as well with the general technological and economic acceleration in the seventies, though. Hardly a decade later, this led to the collapse of their political system.

Another example of the problems created by lagging is the population explosion in the Third World. Parents in poor countries tend to have many children, because they assume that most of them will die before reaching adulthood. Improved medical care has drastically reduced child mortality. However, the parents need some time to adapt their traditions to that new situation. Although they are starting to plan for smaller families, the decrease in childbirths still lags behind the decrease in child deaths. The result is a larger number of children reaching adulthood, an acceleration of exponential growth, and an associated increase in the depletion of land and resources needed to sustain a larger population.

A more subtle effect of reduced friction is the lengthening of cause-and-effect chains. Imagine a row of billiard balls, each ball a short distance from the next one. If you hit the first ball with your cue (cause), it will hit the second ball (effect), which will itself hit the third ball (second effect), and so on. Because of friction, energy is lost, and each next ball will move more slowly than the previous one. At some point the ball will start so slowly that it will stop before it has reached the next one: the causal chain has broken; your initial move will have no further effects. However, if friction would be lower, the chain would be much longer, and a much larger number of balls would be affected by your initial move.

The same principle applies to our low friction society. In earlier periods, an event happening in one country would have little or no effect on events happening in another country. The chain of effects would have died down long before it would have reached the national border. Nowadays, the world as a whole has become interdependent. A poor
harvest in one country will affect the production in another country, which will affect the stock market in a third country, which will affect the employment in yet another country, and so on.

Imagine now that the billiard balls aren’t neatly aligned in a row but scattered over the billiard table. The ball you hit with your cue will now hit several balls rather than one. Each of those hits several further balls, and so on. Balls coming from different directions collide, contributing to the overall chaotic movement. The lower the friction, the more collisions there will be and the longer it will take for the movement to settle down. The resulting configuration will be unrecognizable, no ball remaining in the place it was.

Again, this example can be extended to present society. The poor harvest will not only affect production in another country, but a thousand other things in different places. Each of those will again affect a thousand other phenomena, and so on. Everything interacts with everything. Because of diminished friction, these interactions are more important, more numerous and more complex in their interdependencies than before. That is why we need to keep informed about a multitude of far-away events. The noiseless streams of information merely complement the frictionless streams of causes and effects, allowing us to anticipate possible interventions. But longer cause-and-effect chains mean that we need to take into account many more factors that might influence the outcome. The world is becoming more complex and changeful every day. As a result, prediction of future developments has become practically impossible.

All these effects together create a situation that is very difficult to control. Individuals, organizations and governments are confronted with an ever-increasing variety of new developments. These developments may be positive or negative, but that is hard to tell, since their effects and side effects extend so far. If they are positive, they should be taken up and stimulated. If they are negative, they should be counteracted before they have snowballed out of control. In any case, it is dangerous to ignore them. But reacting adequately means making decisions based on information. The problem with information is that there tends to be too much of it. Or at least there seems to be more information than we can process given our present level of human (and computer) intelligence.

Even if we had adequate information, we would need resources to tackle the problem. But resources spent on a new problem need to be taken away from some existing problem. It is not surprising then that old problems, such as crime, poverty, homelessness and infectious diseases have resurfaced in the developed countries. These problems too must be counteracted before they become uncontrollable. The result is that
governments need to perform a precarious balancing act, moving a little to the right now, a little to the left then. If they are a little too slow in responding (or if their response is a little too enthusiastic) the resulting misbalance may be amplified by positive feedback and grow out of control. As these misbalances increase, while stabilizing forces like friction further erode, and the complexity of decision-making becomes ever larger, it seems that we are heading towards an eventual collapse. At some point, it seems, we will have lost all control over our own fate, and be at the mercy of unstoppable, destructive forces.

But is the situation really so gloomy? Perhaps a different point of view may show a more optimistic picture.
Progress

Developing Nations

Until now, we have been looking at a host of specific processes, recognizing recurrent trends such as ephemeralization, reduction of friction, and increase of complexity. To better understand the overall effect of these developments, though, we need to take a step backward and look at the larger picture. In this picture, the different effects are aggregated over the long term. Instead of examining the violent swings of a particular country’s stock market, we might consider the aggregate gross product for the world. This index of overall production of goods and services shows a remarkably stable growth over the past four centuries. Economic crises in certain countries are usually counterbalanced by accelerated growth in other countries. Of course, there are temporary slowdowns, like the 1930’s depression, but the overall trend is unmistakable.

Though it facilitates many things, economic growth is not necessarily good. Because of its exponential character, it can lead to rapid exhaustion of natural resources. That is why the emphasis has shifted from growth per se, to sustainable growth. This means that all resources consumed should be replenished in some way, by nature or by society. For example, the depletion of ocean reserves by overfishing may be compensated by “farming” fish in fjords or freshwater basins. Although depletion has made the average price of fish increase over the last decade, the price of farmed species, such as salmon, has dropped spectacularly.

This example illustrates the general principle that when resources become rare, new technologies are developed that diminish the need for these resources. Such technologies either produce more out of the same amount of resources (ephemeralization, e.g. fuel-saving engines), use wholly different resources (e.g. solar energy), or even start producing the resources themselves (e.g. fish farming). The rarer a resource gets, the more research will be devoted to developing such technologies, and the more likely a breakthrough becomes.

That is why simple quantitative extrapolations of growth or exhaustion processes often turn out to be wrong. The famous 1972 “Limits to Growth” model, which proposed a gloomy view of the world’s future, is a case in point. Twenty-five years later, its basic premise that unbridled exponential growth leads to collapse is still eminently applicable.
However, its concrete predictions were often far off the mark. In many respects, the world is in a much better shape than the one predicted by the model. This is because the model did not take into account either technological innovation or changing attitudes. No model can predict the effect of discoveries that did not yet exist when the model was made. The only thing we can predict is that discoveries will be made, and that they will be applied to solve problems. The extent to which they will succeed, however, is beyond anybody’s guess.

Again, the best we can do is to look at the widest possible picture, where the effects of multiple events and trends are averaged out. Not just wealth, but health, shows a remarkably steady progress for the world population. Both rich and poor countries experience a continuing decrease in mortality. In the developing nations, child death rates have fallen, over the past half-century, from 300 per 1000 births to less than 100. Average life expectancy has risen with 50%, from 40 to over 60 years. In the wealthy nations too, average life expectancies of 75 to 80 years continue to increase, in spite of new “civilisational” diseases, such as stress-induced heart problems and cancers caused by smoking or pollutants. At present, life expectancy increases with some three years for every ten years that pass. Together with a decrease in the birth rate this brings about an overall ageing of the population.

Of course, there are exceptions to this general pattern. Because of the social and economic crisis accompanying the transition to a market economy, the countries of the former Soviet Union have witnessed a decrease in wealth and life expectancy during the last few years. But it is likely that that will be merely a temporary blip in the statistics, which will be redressed as soon as society has adjusted to a new mode of functioning.

The steady increase in life expectancy is perhaps the most reliable measure of overall progress. Good health and long life are universal values, to which all people and cultures aspire. The increase is caused by a combination of factors: antibiotics, better hygiene, better nutrition, increased safety, stronger awareness of dangers to health, better education, extended health care, more free time, advances in medical technology, etc. Each of these factors itself results from a combination of causes. For example, the number of deadly accidents in the USA has decreased with some 50% over the past half-century. This is the result of a multitude of safety improvements, including seat belts, better highway design, smoke detectors, and window guards in apartments.

Why does such a diverse array of interacting factors lead to such a regular and predictable outcome? We have seen that well-intended innovations often have unpredictable and damaging side effects. For example, the pesticides used to increase crop yield leave residues in the human body. Yet, until now the lives gained by increased
food production certainly outweigh the lives lost by pesticide poisoning. Calculations show that the elimination of all cancer deaths due to pollutants, additives and nuclear materials would increase life expectancy by a mere 90 days, which is negligible compared to the gains of over 25 years made since the beginning of the twentieth century. The increasing life expectancy seems to imply that this is a general trend: statistically, the positive effects of innovation are much larger than the negative ones.

The reason may simply be that once a negative effect appears, it triggers changes in attitude, policy measures and research aimed at redressing the situation. For example, use of pesticides is being replaced by the introduction of disease resistant plants, and other biological technologies. The stress of modern life is being tackled by advanced relaxation techniques, such as meditation or biofeedback. Not all countermeasures will be effective, but on average they seem to be more often successful than not.

What all these developments have in common is that they require a more complex organization. A high level executive job, together with a relaxation program, requires a more complex life style than quiet, predictable farm labor. Countermeasures and countermeasures to the side effects of earlier countermeasures just add complexity. Such an intricately balanced development requires a system of controls and supports, rooted in a reliable infrastructure. A nation’s infrastructure has physical, social and cultural components. The physical infrastructure includes things like roads, railways, hospitals, communication networks, factories, houses, ... The social infrastructure includes institutions such as government, democratic representation, markets, police, the legal system, health care, schools, universities, social services, ... The cultural infrastructure includes the general knowledge, values and attitudes of the population. All of these systems help people to lead healthy and prosperous lives.

These different types of infrastructure, contributing to a nation’s level of organization, tend to develop together. In most countries, when one of these important systems is underdeveloped, other systems tend to be underdeveloped too. Nations with poor health care usually also have low education levels, poor road systems, lack of democracy, unstable or rigid governments, lack of individual ambition, low trade, and deficient communication networks. Examples in different continents include Haiti, Sudan, Somalia and Afghanistan.

That is why institutions such as the United Nations can use abstract development indicators to measure how far a country has advanced. These indicators are statistical measures that combine several of the factors we have noted, such as level of education, GNP per head of the population, number of telephones per inhabitant, availability of running water, etc. Some of the best known are the Physical Quality of Life Index, which
aggregates child mortality, life expectancy and literacy, and the Human Development Index, which moreover includes GNP and average duration of schooling, but excludes child mortality. However diverse or seemingly subjective these factors may seem to be, the combined index is reliable because all these factors tend to go up or down together.

That is because they all measure components or aspects of the general infrastructure or social organization. If one of these components is missing, the rest will not be able to function properly. For example, it does not make sense to try to boost trade in a nation that lacks roads to transport the goods. You cannot teach people how to live more healthily in a region where no one can read and where TV’s and telephones don’t exist. Again, there will be exceptions to the general rule: regions where some aspects are much better developed than others. A classic example is the Indian state of Kerala, where the physical quality of life index is close to the one of the USA, while the income per head of the population is 70 times smaller than the American one. But these exceptions tend to be unstable: sooner or later a corrective process will either push up the underperforming factors, or push down the inflated ones.

Perhaps the most interesting combination of factors is wealth versus education level. A 1979 book by a Brussels colleague of mine, Peter Peeters, provides a chart where income per head of the population for developing countries is plotted against the percentage of people who can read and write. Using data for the early 1970s, Peeters found that almost all countries are situated in a narrow band in which income increases almost proportionally with literacy. Two small groups of countries lie outside that band. A group of oil producing countries, including Iran, Iraq, Kuwait, Libya and Saudi Arabia, lies on the side where income level is much higher than literacy. On the opposite side, where literacy is higher than income level, lie a number of South East Asian and Latin American countries, including South Korea, Taiwan, Thailand, Indonesia, Uruguay and Costa Rica. Peeters ventured the prediction that the latter group would become much wealthier, thus correcting the imbalance. It seems that his prediction has come true. More unexpectedly, some of the outliers on the other side have performed the opposite movement: because of war or international isolation, the relative wealth of Libya, Iran, Iraq and Kuwait has seen a marked decline over the past years.

The development indexes used to classify nations aren’t very informative when applied to the most advanced countries. This is because these indicators in a sense measure the distance between the “developing” and the “developed” nations. The richest nations function as a yardstick against which progress is measured. With measures such as illiteracy or child mortality there is not much progress to be made once these variables are close to 0%. Measures of availability, such as number of cars, telephones or TV sets
per head of the population, also lose their usefulness once a certain level has been reached. Going from one to five cars per person does not look in any way like progress. Even a more abstract measure of wealth, such as Gross National Product (GNP), is questionable as an index of progress. Increased production can be negative if it implies increased resource exhaustion. But GNP does not even represent the actual production of goods or services. It rather measures the amount of market transactions, in which a good or service is paid for by a consumer. This can lead to weird distortions.

Imagine a country in which the people are wealthy and content. They earn quite some money, but spend little, saving most of their income for later, so that they can buy a house or other durable property. In their free time, they work in their garden, have discussions with friends, play musical instruments, or draw pictures and generally get the best out of life without really consuming much. Now, imagine a second country where people are wealthy but stressed. The moment they get money they spend it: they go shopping, buy and sell. Not content with the car they have, they will sell it within a year as a second-hand and buy a more fashionable model. Since they don’t have the savings that would allow them to buy property, they need to pay large amounts for rent and interests on the sums they borrowed from the bank. Because of their stressful life, they are often ill and spend a lot of money on doctors, medicines, hospitals and stress-relieving programs. If something does not go the way they want, they will blame it on the firm or person who delivered the service, and sue them for compensation. They may win or lose the court case, but whatever the outcome, a lot of money will have been spent on lawyers, judges, and simply lost time and energy. In this country, large amounts of money are constantly changing hands, thus boosting its GNP. In the first country, on the other hand, the GNP will be much lower, even though the inhabitants’ actual wealth and overall level of comfort will be much higher, while growth will be much easier to sustain. It is clear that in this case GNP is quite the opposite of a measure of progress.

A better candidate to measure economic progress may be productivity, the amount of goods produced for a given amount of labor. Productivity is rising at a stable rate of over 3% a year worldwide. Technological innovation ensures that this increase will continue steadfastly, undisturbed by the wild upward and downward swings of the capitalist economy that affect GNP. Since productivity indicates the degree to which we can do more with less, it is also a good measure for sustainable growth: increasing production with stable or decreasing use of resources. But the most important variable, quality of life, remains elusive. Although the increase in life expectancy seems to indicate that life is in an objective sense getting better, it is still not clear how or why that
happens. Instead of focusing on material growth, we should perhaps pay more attention to mental development.

**The Riddle of Increasing IQ**

We have seen that literacy is a good indicator of potential for economic development. In the last 50 years, adult literacy rates in the developing countries have doubled, from 35 to 70%. This is just a symptom of a more general phenomenon: the level of education is increasing everywhere in the world. People go longer to school, get higher diplomas, have better access to books, newspapers, TV programs and other carriers of information, and are more stimulated to permanently continue their education, at work or in their free time. For example, in Belgium the average duration of schooling has risen from 6 to 14 years over the past century. This has profound effects on economical, social and physical well-being.

A friend of mine, Jan Bernheim, has called this development the “cognitive revolution”. As a doctor specialized in cancers, he has been seeing thousands of patients in his career spanning thirty years. Especially in the beginning, he had great difficulty explaining the sometimes complicated treatments needed to tackle abstract diseases. He often was frustrated at the misunderstanding of his patients, who ignored important guidelines because they did not see the connection with their particular situation. But that has changed. Patients nowadays seem much better informed and have a quicker grasp of the implications of different treatments and life styles. Bernheim saw the same evolution in his students, who seem much less naive when entering university than their predecessors a generation ago. He concluded that something fundamental is taking place: the general level of knowledge and thinking in the population has risen dramatically. As a result, people are generally more sophisticated, rational, and broad-minded, and less inclined to believe in the simplistic explanations proposed by fanatics and ideologues.

The effects of this cognitive revolution may be illustrated by the 1989 collapse of the former Eastern Block. After the reforms of Gorbachov in the Soviet Union had created a more open and liberal atmosphere, the fall of the Berlin Wall was the starting shot for a rapid succession of revolutions transforming all Eastern European states from communism to democracy. It culminated in a failed hard-line coup against Gorbachov, the replacement of Gorbachov by Yeltsin, and the break-up of the Soviet empire into independent countries. Apart from the revolt against the dictator Ceausescu in Romania, all these transitions were remarkably peaceful. Never before in history have such dramatic changes of regime been accompanied by so little violence. Bernheim’s
interpretation is that both the men in power and the people rebelling against them were generally intelligent, well-educated and reasonable individuals. These participants understood that the course of history cannot be stopped, and that there is little sense either in trying to violently suppress the changes, or in taking revenge on those that maintained the status quo until then.

The often-chaotic developments in Eastern Europe after 1989 show a less rosy picture. Bloody wars have erupted in regions like Bosnia, Chechnya and Nagorno-Karabakh. Yet it is undeniable that there is a correlation between peacefulness and the level of education, information and general development. The regions where violence broke out were also the regions where the level of development was lowest: Yugoslavia, Romania, and the Caucasus. Within the old Yugoslav federation, the most developed region, Slovenia, was spared from all troubles, while the war concentrated in the least developed one, Bosnia-Herzegovina. On the other hand, the transition to a Western style democracy occurred most smoothly in the countries with the highest levels of openness and development: Eastern Germany, Czechia, Hungary and Poland.

This same pattern can be found world-wide: the countries where war breaks out have typically a low level of education and economic development. Examples include Afghanistan, Somalia, Rwanda, Haiti and Angola. The countries involved in the two Gulf wars, Iran, Iraq and Kuwait, were economically rich because of their oil reserves, but intellectually underdeveloped, as I noted earlier. One of the main theses of the book of Peeters, which I mentioned then, is that militarism is characteristic for a low level of development. Once a nation has surpassed that level, the motivation to solve problems by war disappears. That is why the rich nations of the West are so reluctant to involve their troops in any conflict, even when everybody agrees that intervention is needed to avoid worse. Although Peeters puts more emphasis on economic development, arguing that wealthy people have more to lose and less to gain in a war, I would add intellectual development as a primary factor.

The First World War will remain in memory for the naivety, if not downright stupidity, with which young, enthusiastic soldiers jumped out of the trenches and ran to their death. The war itself was started by the Europeans powers after a round of colossal miscalculations, where every party expected quick gains, but instead got into a quagmire that lasted four years, at a staggering cost in human lives, money and equipment. The absurdity of the war sent shockwaves through the intellectual establishment, but it took the Second World War before this awareness would reach the population at large. Since then, involvement in war has become increasingly unpopular in the West, with the Vietnam war as the final example of a war that had to be stopped in part because of
popular resistance at home. This development has certainly been fueled by the fact that people are better informed of how damaging a war really is, and how little there really is to gain. The media, which show the destruction and blood live in people’s living room, have played an important part in this evolution. But another factor is certainly that the naive patriotic, religious or ideological reasons given to start a war have lost their appeal to an increasingly sophisticated public.

Such a cognitive interpretation of world events leads Bernheim to conclude that history does not repeat itself. Though lessons are sometimes forgotten, most of the time knowledge and awareness accumulate irreversibly. Never before have so many people known so much about so many domains. Each of the three components in this expression continues to increase: the number of people being educated and informed; the amount of education and information each of them gets; and the breadth of the domain about which science and the media provide useful information. This may seem trivial if education is seen as mere accumulation of data. However, there is reason to believe that a more fundamental transformation is taking place.

James Flynn, a political scientist working in New Zealand, observed in the 1980’s that the scores of different groups of people on standard intelligence tests had consistently augmented over the past decades. Earlier researchers had failed to pay attention to that trend, because IQ scores are always calculated with respect to the average score for the present group. By definition, the average is set to 100. Someone who scores 20% more than the average would therefore get an IQ of 120. But if that person’s score would be compared with the average for the corresponding group, tested one generation earlier, the final score would be about 130. Flynn was the first to systematically make such cross-generational comparisons.

Since then, the so-called “Flynn effect” has been confirmed by numerous studies. The same pattern, an average increase of over three IQ points per decade, was found for virtually every type of intelligence test, delivered to virtually every type of group. The pattern applied to some 20 countries for which data were available, including the USA, Canada and different European nations. The increase was highest, 20 points per generation (30 years), in Belgium, Holland and Israel, and lowest, 10 points per generation, in Denmark and Sweden. Although the data are limited, it moreover seems that the increase is accelerating. In Holland, for example, scores went up most (over 8 points) for the last measured period, 1972 to 1982. For one type of test, Raven’s Progressive Matrices, Flynn found data that spanned a complete century. He concluded that someone who scored among the best 10% a hundred years ago, would nowadays be
categorized among the 5% weakest. That means that someone who would be considered bright a century ago, should now be considered a moron!

Such a result has unexpected implications for the relation between intelligence and age. Older people tend to have lower scores on IQ tests than younger people. Until now, it was always assumed that this means that intelligence diminishes with age. However, this observation can be explained as well by noting that older people were raised in a period when the general level of intelligence was lower. Flynn showed that if people’s IQ is evaluated with tests calibrated for the period during which they grew up, an old person scores as well as a young one. The reason that older people do less well on IQ tests is not that they have become more stupid with age, but that the younger generation simply got a head start.

One might expect that the Flynn effect would be more clear for tests that emphasize culture or education. The opposite is true, however: the increase is most striking for tests measuring the ability to recognize abstract, non-verbal patterns. Tests emphasizing traditional school knowledge show much less progress. This means that something more profound than mere accumulation of data is happening inside people’s heads. None of the scientists who have studied the effect can offer a simple explanation.

Flynn himself admits that he is baffled by the results, and that he finds it hard to believe that his generation is significantly more intelligent than the one of his parents. He proposes the following argument. Compared to the previous generation, the number of people who score high enough to be classified as “genius” has increased more than 20 times. This means that we should now be witnessing, in Flynn’s own words, “a cultural renaissance too great to be overlooked”. Because he finds this conclusion implausible, he suggests that what has risen is not intelligence itself but some kind of “abstract problem solving ability”. But if we look at the ever-accelerating production of scientific discoveries, technological innovations and cultural developments in general, the “cultural renaissance” does not seem such an absurd idea anymore. And whether you call the factor that rises “intelligence” or “abstract problem solving ability”, the conclusion that people have become intellectually more capable remains the same.

To me, it seems likely that this intellectual progress is caused by a combination of factors, just like the increases in life expectancy and in productivity, which we discussed earlier. Some factors may have a negative influence on intellectual development, but most developments are positive and reinforce each other. The most obvious one is longer schooling, but this cannot explain everything, as Flynn found that the IQs of American children have been rising even during periods when the time spent in school remained the same. Stimulation by the media, and in particular by television, may be another one, but
this cannot explain progress before the advent of television in the 1950s. Generally improved health and nutrition is likely to contribute too. It has been shown that poor nutrition in early age impairs intellectual development, but this only applies to children who have been severely underfed. A factor that may have been overlooked is that parents nowadays tend to pay much more attention to their children, thus stimulating their cognitive development. This is possible because parents tend to have less children to care for, to have more free time, to be more wealthy, to be better educated, and to have a better insight in the needs of their children.

A more general factor is that society as a whole functions at a higher intellectual level, proposing to the curious child more information, more intellectual challenges, more complex problems, more examples to be followed, and more reasoning methods to be applied. Just using everyday appliances, such as VCRs, microwave ovens, and thermostats, demands a more abstract type of reasoning, of which the older generation is often incapable. The increased complexity of life is likely to stimulate an increased complexity of mind. The growing use of computers for education or games at an early age is likely to further boost general knowledge, abstract reasoning and intellectual agility.

**The Singularity**

Although acceleration and complexity have made most concrete developments impossible to predict, we have seen that large scale, statistical factors, such as wealth, life expectancy, intelligence, productivity, speed of information transmission and speed of information processing, increase in a surprisingly regular way. This makes it possible to extrapolate their development into the future. Most of these growth processes are exponential, characterized by a constant doubling period, or a constant increase in percentage per year. This means that the underlying growth mechanism is stable, producing a fixed number of new items for a given number of existing ones.

However, some processes grow even more quickly. For example, population growth in percentage per year is much larger now than it was a century ago. This is because medical progress has augmented the gap between the percentage of births and the percentage of deaths per year. If the growth of the world population over the past millennia is plotted out, the pattern appears to be *hyperbolic* rather than exponential. Hyperbolic growth is characterized by the fact that the inverse of the increasing variable (e.g. 1 divided by the total population) evolves according to a straight line that slopes downward (see Fig. 1). When the line reaches zero, this means that the variable (world
population in this case) would become equal to 1 divided by zero, which means infinity. This is essentially different from an exponential growth process, which can never reach infinity in a finite time.
Fig. 1: a hyperbolic growth process, where the growing number becomes infinite in the singular point, while its inverse (1 divided by the number) becomes zero.

If the line describing the inverse of world population until recently is extended into the future, we see that it reaches zero around the year 2035. Of course, such an extrapolation is not realistic. It is clear that world population can never become infinite. In fact, population growth is slowing down at this moment. However, that slowdown itself is a revolutionary event, which breaks a trend that has persisted over all of human history.

In mathematics, the point where the value of an otherwise finite and continuous function becomes infinite is called a “singularity”. Since traditional mathematical operations, like differentiation, integration and extrapolation, are based on continuity, they cannot be applied to this singular point. If you ask a computer to calculate the value of one divided by zero it will respond with an error message. A singularity can be defined as the point where mathematical modeling breaks down. The inside of a “black hole” is an example of a singularity in the geometry of space. No scientific theory can say anything about what happens beyond its boundary. We can only postulate that different laws will apply inside, but these laws remain forever out of sight. Similarly, the Big Bang, the beginning of the universe, is a singularity in time, and no amount of extrapolation can describe what happened before this singular event.

The mathematician and science fiction writer Vernor Vinge has proposed that technological innovation is racing towards a singularity. Remember that we sketched scientific discovery as an exponentially growing process with a doubling period of about 15 years. We then noted that the doubling period is diminishing because of increasingly efficient communication and processing of the newly derived knowledge. The rate of growth is itself growing. This makes the process super-exponential, and possibly hyperbolic. Vinge would argue that at some point in the near future the doubling period would reach zero, which means that an infinite amount of knowledge would be generated in a finite time. At that point, every extrapolation that we could make based on our present understanding would become meaningless. The world will have entered a new stage, where wholly different rules apply. Whatever remains of humanity as we know it will have changed beyond recognition.

Some data supporting this conjecture can be found in another chart provided by Peeters, in which the time elapsed between a dozen fundamental discoveries and their practical application is plotted against the year in which the invention was applied. The graph shows a downward sloping line, which reaches zero in the year 2000. These are
just a few data points, and the implication that inventions will be applied virtually instantaneously after 2000 does not yet mean that scientific progress will be infinite. Vinge himself would situate the date of the singularity between 2010 and 2040. His reasoning is based on the accelerating growth of computer-aided intelligence. Rather than considering the IQ of an isolated individual, he would look at the team formed by a person and computer. According to Vinge, a PhD armed with an advanced workstation should already be able to solve all IQ tests ever devised. Since computing power undergoes a rapid exponential growth, we will soon reach the stage where the team (or perhaps even the computer on its own) would reach superhuman intelligence. Vinge defines this as the ability to create even greater intelligence than oneself. That is the point at which our understanding, which is based on the experience of our own intelligence, must break down.

A related reasoning was proposed by Jacques Vallée. Extrapolating from the phenomenal growth of computer networks and their power to transmit information, he noted that at some point all existing information would become available instantaneously everywhere. This is the “information singularity”.

These models should not be taken too literally. They are metaphors, proposed to stimulate reflection. It does not make much sense to debate whether “the Singularity”, if such a thing exists, will take place in 2029 or in 2035. Depending on the variable you consider most important, and the range over which you collect data points, you may find one date in which infinity is reached, or another, or none at all. Even if nothing as radical as a “New Age” could be predicted, the zero point method is certainly useful to attract the attention to possible crisis points, where the dynamics of change itself are altered. Peeters has used this method with some success to “predict” historical crises, such as the First World War and the 1930 and 1974 economic recessions. (The Second World War, strangely enough, did not seem to fit the model...).

The point to remember, however, is that abstract, non-material variables, such as intelligence, information, or innovation, aren’t subjected to the same “limits to growth” which characterize the exhaustion of finite resources. Such variables could conceivably reach values, which for all practical purposes may be called “infinite”. Several parallel trends show a hyperbolic type of acceleration which seems to reach its asymptote (the point of infinite speed) somewhere in the first half of the 21st century. This does not mean that actual infinity will be reached, only that a fundamental transition is likely to take place. This will start a wholly new mode of development, governed by laws, which we cannot as yet guess.
Fragmented World Views

Visions of the Future

The preceding considerations have made it clear that predicting the future has become more difficult than ever. If we try to look ahead more than a decade or so, the crystal ball gets cloudy. Yet, people have always felt the need to know where they are going to. Every culture has its own stories and myths about the future, whether it is the coming of the Messiah or the Apocalypse. In our own technological society, this role has been played mostly by the science fiction genre. Since Jules Verne and H. G. Wells, scientists, writers and artists have tried to imagine the world of the future. Their visions fall in between the two broad streams of optimism and pessimism. The optimists believe that progress, fueled by scientific research, will continue to make our life better, conquering all problems. The pessimists, on the contrary, believe that problems are intrinsic to humanity itself, and that science can only aggravate them, unleashing dark forces that may forever escape control. An early and classic example of the latter view is Mary Wollstonecraft Shelley’s novel “Frankenstein”. The scientist Frankenstein, in his investigations of life and death, creates a monster, which he cannot keep under control. The monster escapes and, after terrorizing the neighborhood, finally comes back to destroy its creator.

The optimistic visions have undergone several changes during the past century. The oldest ones are simple extrapolations of technological progress. They assume that all material things will just become bigger, faster, more powerful and more efficient, while culture and society remain basically the same. A typical naive prediction is that after the quick spread of private cars there would be a quick spread of private planes or helicopters. The 1950’s vision of the future pictures a traditional family living in a hi-tech flat, high up in a towering skyscraper. Father goes to work in his personal plane. He calls Mother from the office on the videophone to tell her when he will be back. Mom stays home and cares for the house and children, helped by an anthropomorphic robot, that fulfills the functions of cook, cleaner and babysitter. The children play with futuristic toys, including a robot dog and a weightless top. For vacation, the family goes on a trip to
the Moon. The Hanna and Barbera strip, “The Jetsons”, neatly summarizes this naive picture.

This view is outdated in several respects. On the one hand, it ignores the material limits to growth, which make things like private airplanes and trips in space prohibitively expensive. On the other hand, it fails to appreciate the unlimited capacity for change on the mental and organizational level. Rather than just enhancing or mimicking existing functions like cleaning and telephoning, technological progress will completely redefine the underlying problems. We have already learnt that it is easier and more efficient to build intelligence into a washing machine, than to build an intelligent, humanoid robot that would operate the machine the way we do it. Similarly, we would rather use enhanced telecommunications to work from home, than to travel to the office by plane, and communicate with those who stayed home.

The simplistic belief in purely material progress brought about a strong reaction in the 1960’s. The hippie movement focused on spiritual development, shunning most forms of technology. With the spread of less “materialistic” technologies, such as computers and networks, in the 1980’s, their original vision of “back to Nature” was broadened to encompass new scientific developments. This led to a novel, optimistic picture of the future, the “New Age” vision. Marilyn Ferguson, in her book “The Aquarian Conspiracy”, clearly describes the emergence of this movement and its ideas. The main metaphor is the “age of Aquarius”, the new era that we are entering according to the astrological calendar. This era will be characterized by a more harmonious and loving psychological climate.

From the new sciences, the “New Age” prophets have adopted the emphasis on networks, synergy, self-organization and holism. To this they add ideas from various mystical traditions, including Buddhism, yoga, and shamanism, and from different “alternative” approaches, like parapsychology, tarot, crystal healing and homeopathy. (Frank Capra is one of the best known authors to develop this world view combining science and mysticism.) Their main message is that humanity is quickly moving towards a higher level of consciousness. It will transcend individual awareness and its selfish concerns, and replace it by the “transpersonal” experience of belonging to a larger whole. The resulting synergy between previously competing individuals will release a lot of pent-up energy and creativity. This will solve all the problems of present society, which are caused by interpersonal and intrapersonal conflicts and by disregard for Nature.

Although the optimism of the “New Age” movement is appealing, and they certainly have a point when noting that many difficulties are caused by needless conflicts and contradictions, the methods they propose for transcending these problems appear
rather naive. Their approach seems characterized by the absence of healthy skepticism. From science and technology they merely import metaphors, ignoring the hard work that goes into developing, testing and implementing new ideas. Basically, they propose that if people have enough goodwill, and sufficiently engage in consciousness-raising activities, like meditation and different forms of psychotherapy and “healing”, the transition to the higher level, where all problems are solved, will occur automatically. This looks like wishful thinking more than like a concrete model of future developments.

The pessimistic scenarios too have undergone transformation. Until recently, the most powerful metaphor for the bleak future imagined by the pessimists was “Big Brother”, the all-seeing eye of the totalitarian state. The theme of a technologically controlled, bureaucratic society, where there is no room for freedom or individual expression, returns in numerous novels and movies, from the classics, Zamyatin’s “We” and Orwell’s “1984”, to Terry Gilliam’s satire “Brazil”. These visions were merely extrapolations of existing political systems, like Stalin’s Russia or Hitler’s Germany, with added technology, such as closed camera circuits and computer databases, which increase the control of the regime over its citizens. The last decades have convincingly shown that totalitarianism and technological progress don’t support, but rather oppose each other. On the one hand, communication and computer technology promotes individual expression more than it facilitates government control. On the other hand, the collapse of virtually all police states has made it clear that technological innovation stagnates in totalitarian regimes, making their economies uncompetitive.

A slightly more recent version of the doomsday scenario is the environmental holocaust. The main idea is that nuclear war, the uncontrolled proliferation of pollutants, and/or the exhaustion of natural resources have made the Earth all but uninhabitable. Civilization has collapsed together with the ecosystem. The recurring image is that of a few gangs of survivors, together with mutant rats and cockroaches, fighting for the last remaining resources amongst the ruins of once proud cities. This pessimistic view too has recently become less popular. The reason is the much-diminished likelihood of nuclear war, and the awareness that environmental problems, though serious, are being tackled more and more forcefully.

The most recent pessimistic vision is perhaps the most realistic one. The “cyberpunk” picture combines a focus on increasingly sophisticated cybernetic technology with the desperate anarchism of the 1970’s punk movement (as reflected by their slogan “No Future”). It can be found in the science fiction novels of authors like William Gibson and Bruce Sterling, in movies like Ridley Scott’s “Blade Runner”, or in a TV series like “Wild Palms”. The society they describe is an extrapolation of unbridled
capitalism rather than totalitarian communism. Everybody competes with everybody, and the gap between “haves” and “have-nots” has become much wider. The unimaginable wealth of top business executives contrasts with the abject poverty in which most of humanity lives. Technology is omnipresent, both as a means for control by multinational corporations and as a tool for different forms of theft, sabotage and fraud by criminals and anarchists. Direct brain-to-computer interfaces, global networks and mind altering drugs have become commonplace. Everybody is either vying for control, or trying to escape the harsh reality in computer-generated fantasy worlds. But no one is in control: the technology-driven society is simply too complex for anybody to grasp. The continuing uncertainty and fight for survival have eroded any sense of justice, values or ethics, replacing them by a high-tech variant of the law of the jungle.

Post-Modern Fragmentation

Both cyberpunk and New Age, the presently most fashionable visions of the future, are basically popular movements without strong theoretical underpinning. Professional futurologists generally limit their predictions to relatively specific, and short-term trends, such as ageing of the population, impact of biotechnology, or “cocooning”. Until recently, the wider picture was discussed by academics who got their inspiration from the theories proposed by philosophers and ideologues, such as Marx, Marcuse or Hegel. However, the failure of such theories to produce good predictions has brought about a remarkable change in the intellectual climate. Academics have become aware of the deep uncertainty and complexity of present-day society. Many of them have concluded that it is inherently futile to try and build universal theories, which would explain everything. They have lost their belief in progress, in the idea that evolution advances in a single and unambiguous direction. The rather fuzzy cluster of ideas that accompany this movement is called “postmodernism”.

Postmodernism defines itself in contrast to modernism. Originally, the term comes from architecture, where modern architecture denotes the familiar glass, steel and concrete buildings with their straight, rectangular, geometric shapes. The style of architects such as Le Corbusier and Mies van der Rohe seems to aspire to maximal simplicity, regularity and symmetry. Once the novelty has worn off, the resulting buildings often give an impression of cold, lifeless, mechanical things. This has led in the 1960’s to a reaction by younger architects, who included different decorative elements inspired by earlier periods in their design. This eclectic mixture of styles was called “postmodern architecture”. From there the term “postmodern” quickly spread to art,
where it denoted a departure from the radicalism and abstraction of the old avant-garde, replacing them by a fusion of different popular and traditional elements, like Warhol’s Pop Art or rock music incorporating African and oriental motives. The most important impact of postmodernism was in philosophy, where it was heralded as a new stage in the history of ideas.

In history, the start of “modernity” is usually taken to be the 18th century period of Enlightenment. During that period, a belief in rationality, progress and science spread through the intellectual establishment, preparing the grounds for the industrial revolution. According to the ideology of modernity, rational thought is the key to discovering the truth about the world and ourselves. Such true knowledge will emancipate all individuals, freeing them from the shackles of ignorance, superstition and dogmatism. Moreover, this knowledge will fuel technological progress. By making production more efficient, it will create wealth and make our lives more comfortable and safe. Thus, the on-going accumulation of knowledge through scientific research leads directly to economic and social progress. Since there is only one true picture of the world, there can be only one way to progress: by gradually filling in more and more elements of this picture. The more elements are known, the better science will be able to predict and control nature, and the more the individual will be liberated from the vagaries of fate.

Such a simple, deterministic view of historical progress does not fit in well with the complexity and confusion of the 20th century. Atrocities like Hiroshima or the Holocaust paint a very different picture of the effects of scientific advances. Where is the progress in killing thousands of people in a few minutes? Does it mean killing more people in less time? Questions like these have brought postmodern thinkers to reject the project of modernity and its belief in rationality and progress. They argue that the idea that there is one true representation of the world leads to intolerance and even violence, since it implies the suppression of everyone who disagrees with this picture. Too often, the supposedly superior Western world view has been used to justify the oppression of women, non-Western cultures, and colonized peoples.

Instead, the postmodernists see knowledge as a set of perspectives, where different people have different views, without anyone being “right” or “wrong”. When considering culture, they see a plurality of views, concepts, theories, styles, movements, which are competing, supporting each other or simply existing side-by-side, without any of them being better or worse than the others. They emphasize fragmentations, discontinuities and chaos, rather than the order, coherence and simplicity characterizing the modernist philosophy. Though its aims and background are different, the postmodern
view of society is in many ways similar to the cyberpunk vision. Both see growing complexity and confusion rather than a unification supported by scientific rationality.

When we look at the development of science, the movement towards fragmentation is unmistakable. As more and more concepts, theories and models are developed it becomes impossible for any one scientist to keep informed about all of them. Researchers are forced to focus on smaller and smaller domains, becoming ever more specialized. If they cannot know everything within their discipline, at least they can try to know everything that is important within a subdiscipline or part of a subdiscipline. Present-day scientists know ever more about ever narrower subjects. But as few people are left with an understanding spanning more than one domain, the communication between these domains ceases. Thus, researchers often tackle problems without being aware of similar efforts by people working in a different discipline. Because of that, much effort is lost.

But even if they would know that others have tackled the same problem, they might not be able to use the results. Specialization implies the development of a specialized terminology or language for describing the concepts and phenomena of the domain. When scientists from different domains try to discuss a common problem they use a different language, and have great difficulty understanding each other. The problem is not that they just need to extend their vocabularies: different disciplines often use the same words, but with a different meaning. Though the meanings of a term may be similar, there are subtle differences in its implications when used in different theories.

For example, the terms “time” and “space” mean something quite different in classical mechanics and in the theory of relativity. Though these theories both belong to the discipline of physics, and describe roughly the same range of phenomena, they really propose different views of the world. In classical mechanics, time and space are strictly separate and absolute, in relativity theory they are relative and connected to each other. There is no equivalent for the relativistic time concept in classical mechanics. The concept of time is again quite different in thermodynamics, another part of physics. The philosopher of science Thomas Kuhn introduced the term “incommensurability” for this impossibility to translate the ideas of one theory into the concepts of another one.

The growing number of incommensurable theories negates the belief in scientific progress as the gradual completion of a true picture of the world. If such a picture would exist, the different theories would function as the pieces of a jigsaw puzzle, fitting together and gradually filling more and more of the empty space until the complete picture emerges. But incommensurable theories simply don’t fit. They look like fragments, but they cannot be integrated into a whole. It is as if you are trying to
reconstruct an ancient clay pot by puzzling together shards that in fact belong to half a dozen different pots. The more science grows, the more fragments are created, and the more difficult it becomes to build a unified picture, even within the same discipline.

The same fragmentation can be seen in culture at large. Because of the media and the increased possibilities for travel, people come into contact with a much larger array of cultures, religions, ideologies, styles and fashions. The present ideal of *pluralism* means the peaceful co-existence of different cultures. But since no culture perfectly fits the complexity of present society, people will tend to pick out those pieces they find personally most appealing. For example, a person may combine Christian church services with Buddhist meditation techniques and a view of the universe inspired by the Big Bang theory. The growing trends towards pluralism and multiculturalism are reinforced by *secularization*: people are less and less dependent for their values and beliefs on traditional churches. Though religious feelings as such don’t necessarily diminish, the authority of religious leaders and religious institutions continues to decrease. Few Catholics still take the views of the Pope on contraception or sex outside marriage seriously. They rather follow those guidelines they feel like following and ignore the rest.

It is as if society presents us with a supermarket full of the most diverse ideas, beliefs, attitudes, codes, guidelines and rules of behavior. We just go shopping along the racks, picking one thing here, another one there, until the cart is full. The things we bring home may each look quite attractive, but they have little in common, and certainly don’t make any coherent whole. The supermarket of culture may provide many thrills, but it does not leave us with a feeling of satisfaction. In the end, we don’t know what we should believe or not.

According to postmodern thinkers, it has to be that way. They reject the idea that there could exist an integrated world view. There are no basic principles or foundations on which to build an encompassing picture of the universe. There is no universal language that would allow scientists from different disciplines to discuss their achievements and reach consensus. There is no unique story, telling us how the world and the different things, plants, animals and people in it were created, and how humanity is progressing from primitive tribes to an advanced technological society. Postmodernity has been defined as “the end of the great narratives”. Neither the accounts of the Book Genesis, nor the ancient Greek or Indian myths about the origin of the world, nor the theories of the Big Bang and of evolution have any absolute value.

Though not everyone will agree with such radical conclusions, our present age undeniably shows fragmentation and an increasing awareness of the relativity of beliefs.
and values. This state of affairs defines the “postmodern condition”. Let us examine which effects this has on society.

**The Causes of Discontent**

When we review the development trends in contemporary society, we are confronted with a paradox. On the one hand, the pattern shown by statistical indicators, such as life expectancy, wealth and intellectual development, is one of unmistakable, continuing progress. On the other hand, the prevailing mood is one of gloom and doom, of uncertainty and frustration, and of a general disenchantment with the idea of progress. A telling sign of this attitude is the recurring talk of “crisis”, “recession” and “depression” during the seventies, eighties and nineties, even though on average the economy continued to grow at a normal rate. Why aren’t people more satisfied with the objective improvements in their situation?

We have already discussed one of the causes, the uncertainty and loss of control caused by accelerating change. The fact that on average things develop for the better provides little comfort for the individual, organization or country thrown around by the waves of change. There are too many examples of catastrophes and failures to forget about the risks. Even if everything goes for the best, the speed of potential developments requires constant vigilance. Add to that the growing complexity of decision-making and the information overload, and it becomes clear why stress is growing. Continued stress caused by unpredictable novelty promotes anxiety, agressivity and despair.

But if stressful situations are successfully overcome, their effect should be stimulating rather than depressing. Why cannot society fully enjoy its undeniable successes? One reason is that failures and catastrophes simply receive much more attention. Most of the time, things go the way they are expected to go, because that is what expectations are based on. In spite of the rhetoric about the relativity of progress, the normal expectation is that things either remain the same or slowly improve. A situation will only attract attention if it deviates from this default expectation, i.e. if it either deteriorates or improves spectacularly. Negative developments are usually the result of a sudden, unexpected disturbance: an error, an accident, a conflict, a natural disaster. Such situations require quick action, and therefore the immediate and full attention of the people involved. Arousal, stress and emotion are the genetically programmed reactions to such unexpected deviations. Psychological experiments have shown that such stressful experiences remain much longer in memory than less emotionally loaded events.
Positive developments, on the other hand, are usually the accumulated effect of the sustained efforts of many people. They merely require further continuation of the activities, without much emotion. Bursts of excitement are more likely to harm than to help. Therefore, they are less likely to be noticed and remembered. Sudden breakthroughs are very rare. In general, when a new scientific discovery is announced in the media, the only thing sudden about it is the decision of the researchers to contact the press. The “discovery” is normally a process of collecting, integrating and consolidating ideas and data that has taken years. Similarly, breakthroughs in negotiations are usually the result of months of sustained bargaining and discussion. Disasters killing hundreds of people, on the other hand, may be caused by a single earthquake, fire or terrorist attack. The mechanism of reduced friction we discussed earlier will intensify both positive and negative developments. However, it are the negative ones that are most likely to catch the attention and to be remembered in the long term. Therefore, people get the false impression that the situation is deteriorating.

This effect is amplified by the media. Something is deemed newsworthy only if it is likely to grab the attention of many people. This excludes most of the slow, predictable processes of improvement, while favoring events like murders, wars, famines, kidnappings, etc. Marshall MacLuhan already noted “good news is no news”. The larger the deviation from the normal situation, the more newsworthy it is. By definition, large deviations are very unlikely. However, since the present media have a global reach, the total amount of reported events from which they can select the most spectacular ones is extremely large. This means that every day they can find a sufficiently great number of exceptional situations to catch the attention of the newspaper or TV audiences. Though the chance that you would fall prey to a serial killer is virtually zero, there will always be a city somewhere in the world where such a killer is active. This will provide journalists with an endless source of stories about serial killings, thus keeping the attention of the public focused on this threat.

Fiction, which is not constrained by the journalist’s requirement that a story should be based on facts, further reinforces this effect. The occurrence of killings, crime and war in movies, novels and TV series is immensely greater than its occurrence in the outside world. This leaves the public with a strongly distorted view of the true situation. Interviews have shown that people who watch TV consistently overestimate the frequency of murder and other spectacular crimes.

Although the bias towards bad news makes people more pessimistic about the world at large, you might expect that they would still be happy with their personal successes. But satisfaction or happiness depends only indirectly on the objective
situation. People experience satisfaction when they reach the goal they have set to themselves. But that goal is a moving target: once you have reached it, you will get used to the situation and start to aspire to even better things. Psychologists have observed that persons choose their level of aspiration to be higher than what they have now, but realistically attainable. People judge attainability by comparing their goal with their previous experience, and with the situations of the people around them. This provides them with a standard or reference level against which progress can be measured.

For example, a poor farmer from a little village somewhere in Ethiopia, will not consider it a realistic goal to get his own three story house with garage, air conditioning and swimming pool. Yet, for an inhabitant of a rich American suburb, this may seem like a minimal requirement. The American may feel continuously frustrated because he cannot afford a swimming pool as big as his neighbor’s, while the Ethiopian may be very happy because a good harvest allows him to extend his hut, and build a shed for his chickens.

According to the “level of aspiration” theory, people feel happy when they regularly make progress towards their self-selected goal, which moves up in step with their successes. Though the net result may be the same, a series of small improvements is likely to be more satisfying than a big step forward followed by a prolonged status quo or deterioration. This explains why people who win big in a lottery often lead miserable lives afterwards. Once they have gotten used to their new wealth, their old goals and aspirations seem petty, and they become aware that they will never be able to make any improvement as significant as the one brought about by the prize money.

Let us apply this theory of satisfaction to contemporary society. The average level of wealth and achievement has increased spectacularly. Ephemeralization has made progress easier. Therefore, people tend to set their levels of aspiration not only high in absolute terms, but high with respect to their present situation. Since they have experienced how much progress can be made in a short time, they expect that in the future too they will be able to advance quickly. When they look around, they will always find examples of people who are more successful than they are. The democratization of society and the increasing reach of the media have strongly diminished the psychological distance between rich and poor, between upper and lower classes. It has become more natural for an ordinary person to compare his or her lifestyle with the one enjoyed by film stars, tycoons or sporting heroes. In earlier times, the serf would not even have dared to imagine that he could live the life of a nobleman. The barrier between them was so large that they might as well have belonged to different species.
Reduced friction and newly created opportunities have now made it possible for the lucky to progress from rags to riches in a few years time. Such achievements are certain to be well-publicized in the media, and to be chosen as examples by many. Thus, the maximum level of success that seems attainable has strongly increased for most people. But at the same time, the overall spread of success levels has become much wider. Reduced friction has increased both potential success and potential failure. In a few hours time, speculators may either gain or lose fortunes. Moreover, the value of any achievement has become much less stable. The firm which just introduced a revolutionary new product that is going to make everybody’s life easier, may discover one year later that another firm has invented an even more powerful solution that makes all their accomplishments obsolete.

In such a climate it becomes practically impossible to estimate what a realistic, “average” level of achievement would be. Many people are likely to fall short of the goal they set themselves. Those who overshoot their target will quickly get used to their success and set much higher goals for the future. Since nobody really can control the developments, sooner or later they too will experience a spectacular failure. In short, the recipe for happiness, the successive attainment of progressively more lofty goals, seems more and more difficult to realize. In spite of the high level of average achievement, many people will feel frustrated because their actual achievement is out of sync with their aspiration. Similarly, the levels of achievement they expect from their children, friends, colleagues, employer or government are likely to be inaccurate most of the time. This will contribute to their overall feelings of frustration, pessimism and cynicism about the world around them.

The Breakdown of Values

The postmodern condition has more serious consequences than a discrepancy between aspiration and achievement. People not only find it difficult to decide how much they want to achieve, they often don’t know what they want to achieve. Should you get a job in a big corporation, or start your own business? Or perhaps continue your studies; stay home and look after the family; work for a charity; or, more radically, leave everything behind and become a monk in a Buddhist monastery? There are so many alternatives available that it becomes extremely difficult to choose. Moreover, the traditional values,
determining what is “good” and what is “evil”, are becoming less and less reliable as guidelines.

The world has changed so much that traditionally good things now often show negative consequences, while formerly “bad” things are seen in a much more positive light. For example, it used to be that transforming wilderness into roads, cities and arable land was a sign of progress. Nowadays, there are so few virgin forests left that most people would agree that they should be preserved at all cost. Similarly, in previous times sex outside marriage was viewed as an inherently dangerous phenomenon, responsible for spreading diseases and destabilizing families. Presently, thanks to the use of antibiotics and contraceptives, extramarital sex is seen as rather innocuous, and a matter of personal preferences rather than a danger to society. Recently, the spread of AIDS has made irregular sex dangerous again, but that might change once more with the development of an HIV vaccine.

Scientific and technological innovations continuously create new questions, which the traditional value systems cannot answer. For example, the preservation of human life seems a universal value, valid for all times and ages. Medical science has discovered many ways to keep people alive that would otherwise have died. But when the suffering becomes too great, would it not be better to let the patient quietly die, rather than artificially prolong a miserable life? There are no ready-made answers to such problems.

Many religious people see genetic manipulation of living organisms as evil, since it implies humans playing the role of God. Others see it as merely a more efficient way of doing what people have always done: selectively breeding plants and animals in order to increase farming productivity. The fact is that traditional religious teachings have nothing to say about the issue. None of the great prophets could have foreseen the discovery of DNA splicing. Therefore, you can hardly expect their writings to guide you in the matter.

Most people agree that drugs that create a chemical addiction are evil, since they tend to destroy individuals and families. Few are aware, however, that new electronic technologies, from virtual reality computer games to direct brain stimulation, have the potential to produce even stronger forms of addiction. Eventually, rules controlling the use of these technologies will have to be established. Again, there is no ethical system that can help us in formulating such rules.

The potential of science and technology to create new ethical issues is infinite. Every new discovery can be used for good or for bad. But how do we decide what is good and what is bad? Postmodern fragmentation has eroded all belief systems: religion, culture, philosophy, even science, have all been reduced to a jumble of bits and pieces
without universal value. Different groups advocate different values. In a democratic, multicultural society the prevailing attitude is one of “live and let live”: everybody has the right to his or her own opinion, as long as it is not imposed on the rest of the population. But this creates a dangerous situation, where selfish, neurotic, or even criminal behavior can be justified on the basis of idiosyncratic rules. In the USA, self-proclaimed “prophets” and preachers commonly start their own churches, with the main intention to extort money from credulous followers. Given the freedom of religion, it is very difficult to control such parasitic organizations. Who is to say whether one religious system is more worthy than another one?

The absence of a shared value system, of a consensus about “dos” and “don’ts”, has profound effects on individuals and society. The absence of goals makes that many people lack a sense of direction. They desperately search for meaning, for a mission they could devote their life to. Different religious sects and cults have arisen to cater for that need. But the world views they propose are severely restricted, and often dangerously detached from reality.

Others look at kings, presidents or religious leaders for guidance. But these leaders too have lost their sense of direction. Moreover, because of the rapid changes and growing complexity of society they have much less control over the course of events than people think they have. The result is a general disenchantment with political and religious institutions. It seems that politicians the world over have reached record lows in popularity. People’s satisfaction with governments and supranational institutions is universally poor.

In reaction, many turn their back on society and its institutions. The increased permissiveness, resulting from the erosion of traditional “don’ts”, has made it easier to experiment with unorthodox activities. Many of those, such as use of soft drugs, computer hacking and hooliganism, take place in a twilight zone, where the border between crime and “normal” behavior has blurred. The climate of nihilism and despair fosters even more radical reactions, like terrorism, drug addiction or suicide. The growing inequality between rich and poor moreover has increased the temptation to make money quickly but illegally, through corruption, tax evasion, theft, or the trafficking of drugs, arms or people. Because of the easier movement of people, information and goods and the increasing complexity of society, law enforcing institutions find it increasingly difficult to control such criminal behavior. It is not surprising then that, after a long period of decrease, crime is again on the rise in many parts of the world.

Often the reaction of society is to try and turn back the clock. Politicians regularly advocate a return to law and order and a re-establishment of traditional family values. In
its more extreme form, this conservative backlash can take the form of religious fundamentalism, or of extreme right, xenophobic movements. The problem is that the clock cannot be turned back. There simply is no traditional value system capable to guide a fast moving and complex society like ours. Artificially imposing an outdated morality is likely to make things worse. It would create a growing class of outcasts: individuals, families, organizations or ideas that simply don’t fit into the model. The danger is that this may appeal to a public looking for scapegoats: if the problems can be blamed on immigrants, one parent families, pornography or the Internet, no deeper explanations seem to be needed. Yet, the only way to really get a grasp on the problem is to develop a new system of values, adapted to a world of complexity and change. The next chapter will try to define more clearly the elements of such a system.
A New World View?

The seven components of a world view

Let us summarize the observations we have made until now. Science and technology are at the base of an accelerating process of change, which transforms individuals, society and nature alike. Ephemeralization, the drive to accomplish ever more with less, leads to a reduction of friction. This facilitates all processes and developments, positive as well as negative. By prolonging cause-and-effect chains, it increases the complexity of interactions. By accelerating runaway processes, it makes it more likely that things would get out of control. Yet, a statistical analysis shows that in the long term positive developments still overshadow the negative ones. But these trends are so abstract that trying to extrapolate them into the future only increases the mystery.

The biggest problem seems to be the effect of this acceleration on human psychology. Neither individual minds nor collective culture seem able to cope with the unpredictable change and growing complexity. Stress, uncertainty and frustration increase, minds are overloaded with information, knowledge fragments, values erode, negative developments are consistently overemphasized, while positive ones are ignored. The resulting climate is one of nihilism, anxiety and despair. While the wisdom gathered in the past has lost much of its validity, we don’t have a clear vision of the future either. As a result, there does not seem to be anything left to guide our actions.

What we need is a framework that ties everything together, that allows us to understand society, the world, and our place in it, and that could help us to make the critical decisions that will shape our future. It would synthesize the wisdom gathered in the different scientific disciplines, philosophies and religions. Rather than focusing on small sections of reality, it would provide us with a picture of the whole. In particular, it would help us to understand, and therefore cope with, complexity and change. Such a conceptual framework may be called a “world view”.

The Belgian philosopher Leo Apostel has devoted his life to the development of such an integrating world view. As he quickly understood, the complexity of this task is too great for one person. Therefore, a major part of Apostel’s efforts were directed at gathering other people, with different scientific and cultural backgrounds, to collaborate on this task. Only in the last years of his life, after several unsuccessful attempts, did he
manage to create such an organization: the “Worldviews” group. It includes people from
disciplines as diverse as engineering, psychiatry, theology, theoretical physics, sociology
and biology.

Their first product was a short book entitled “World views, from fragmentation to
integration”. This booklet is a call to arms, a program listing objectives rather than
achievements. Its main contribution is a clear definition of what a world view is, and
which are its necessary components. The “Worldviews” group continues to work on
different components and aspects of this general objective. Most of its members are also
involved in a new interdisciplinary research center at the Free University of Brussels,
which is named after Leo Apostel. I too work at the “Center Leo Apostel”. Although my
approach is personal, many of the ideas I relate in this book are inspired by the work done
in this center and the Worldviews group.

The book lists seven fundamental components of a world view. I will discuss
them one by one, using a formulation which is slightly different from the one in the book,
but which captures the main ideas. The first component, obviously, is a model of the
world. It should allow us to understand how the world functions and how it is structured.
“World” here means the totality, everything that exists around us, including the physical
universe, the Earth, life, mind, society and culture. We ourselves are an important part of
that world. Therefore, a world view should also answer the basic question: “Who are
we?”

The second component is supposed to explain the first one. It should answer the
questions: “Why is the world the way it is? Where does it all come from? Where do we
come from?” This is perhaps the most important part of a world view. If we can explain
how and why a particular phenomenon (say life or mind) has arisen, we will be able to
better understand how that phenomenon functions. It will also help us to understand how
that phenomenon will continue to evolve.

This extrapolation of past evolution into the future defines a third component of a
world view: futurology. It should answer the question “Where are we going to?” It should
give us a list of possibilities, of more or less probable future developments. But this will
confront us with a choice: which of the different alternatives should we promote and
which should we avoid?

This is the more fundamental issue of value: “What is good and what is evil?” The
theory of values defines the fourth component of a world view. It includes morality
or ethics, the system of rules that tells us how we should or should not behave. It also
gives us a sense of purpose, a direction or set of goals to guide our actions. Together with
the answer to the question “why?”, the answer to the question “what for?”, may help us to understand the real meaning of life.

Knowing what to strive for does not yet mean knowing how to get there, though. The next component must be a theory of action (praxeology). It would answer the question “How should we act?” It would help us to solve practical problems and to implement plans of action.

Plans are based on knowledge and information, on theories and models describing the phenomena we encounter. Therefore, we need to understand how we can construct reliable models. This is the component of knowledge acquisition. It is equivalent to what in philosophy is called “epistemology” or “the theory of knowledge”. It should allow us to distinguish better theories from worse theories. It should answer the traditional philosophical question “What is true and what is false?”

The final point on the agenda of a world view builder is not meant to answer any fundamental question. It just reminds us that world views cannot be developed from scratch. You need building blocks to start with. These building blocks can be found in existing theories, models, concepts, guidelines and values, scattered over the different disciplines and ideologies. This defines the seventh component: fragments of world views as a starting point.

A science of complexity and change

The present book proposes a first sketch of such a world view. It will pay most attention to the first three components, the “what?”, “where from?”, and “where to?” By getting a deeper understanding of the origin and evolution of the world, and the direction in which it is moving, we will get a better view of possible dangers and opportunities. By looking in detail at the history of matter, life and society, we will begin to see which developments lead to dead ends and which produce continuing progress. This will give us a basis for a theory of values. It will also give us hints on which methods and strategies are most likely to succeed, and thus the basis for a theory of action. By studying the evolution of knowledge as a chapter in the evolution of mind, we will lay the groundwork for a theory of knowledge acquisition. Thus, the next three components will naturally flow out of the first three.

But first we must address the seventh component, the existing fragments of knowledge out of which we can build our edifice. We have argued that science is the root of all the changes transforming society. No other part of culture, be it religion, philosophy, art or law, seems to have such a profound impact on the world in which we
live. This has created many dangers as well as opportunities. Therefore, it seems logical
to turn to science as a way to solve the problems, which it itself has created. No other
approach equals science in its capability to tackle complex problems.

Of course, many people will argue that there are problem domains, such as values,
spirituality, love and feelings, where science by definition cannot teach us anything. For
them, science can only explain how things function, not why they are there or what their
purpose is. I would argue that their view of science is old-fashioned and severely limited.
While transforming the world, science itself has changed almost beyond recognition.

At the same time, I need to address the postmodern critique. According to the
postmodernists, science is just one among the many pictures of reality. Moreover, it is
hopelessly fragmented into incompatible points of view, which can never be integrated. I
will try to show that science does acknowledge the relativity of our models of reality, but
that that does not stop it from developing an ever more profound understanding of the
world. The naive, modernist view of science as the uncovering of the true picture of
reality is hopelessly outdated. Yet, by moving to a higher level, science can recover the
idea of progress, while abandoning determinism and absolutism.

During the past fifty years, science has been quietly developing a new view of the
world, essentially different from the Newtonian view characterizing 19th century science.
Just like society has become aware of the essential role of complexity and change,
science has started to understand that these same two features underlie all the phenomena
around us. Classical, Newtonian science started from the principle that everything could
be reduced to simple, unchanging elements. Present-day science has concluded that even
things that look simple and static result from an on-going process of evolution, which
produces ever more complexity.

The concepts and principles of this new world view, which sees the universe as an
evolving, self-organizing whole, are scattered over the most diverse disciplines, theories
and approaches. Recently, some of these approaches, such as theories of chaos, artificial
intelligence, neural networks, cybernetics, and complex adaptive systems, have gotten
quite some publicity. However, the quick turnover of these theories may have created the
impression that they are merely passing fashions, fads of the moment, which within the
next ten years are forgotten and replaced by the newest approach. Superficially, this may
be true. However, I believe that at the core all these approaches share the same
philosophy, centering on the evolution of complexity.

In this book I will try to bring together the basic principles of this philosophy, and
apply them to the evolution of the universe, from the Big Bang to our present society.
This is probably the first time that these new ideas have been applied so widely, across
the borders of all the traditional disciplines. Such an ambitious enterprise requires the study of a variety of domains. This I was only able to do with the help of many colleagues, working in different disciplines. Some of these colleague researchers I have met at the Free University of Brussels, and its Center Leo Apostel. Others I have come in contact with through international meetings and organizations, such as the Principia Cybernetica Project. They all have helped me to widen and deepen my understanding. Of course, the knowledge I have gathered is just a tiny fragment of the knowledge that exists. Yet, I believe that I have managed to bring together some of the very basic principles, which explain complexity and change in the most diverse domains. I hope I will be able to convey these insights in the following pages.

Understanding a problem is the first step towards solving it. Therefore, I believe that the understanding brought about by the new sciences of complexity will also help us to practically tackle society’s problems. As we noted earlier, the step from scientific theory to technological application is relatively short. Once we know how complex, evolving systems function, we can start developing methods and tools to manage them.

The problem of information overload will require a new information technology. It should have the same intelligence and creativity as the human mind in deciding what is important, but with the much higher capacity for processing and storage that computers are capable of. The problem of loss of control may be tackled by designing systems that are intrinsically more stable. Such systems should be able to autonomously adapt to rapidly changing circumstances, without requiring constant supervision. This means we will have to delegate authority from the higher levels to the lower levels, and promote self-organization and autonomy at the lower level. The problems of frustration and pessimism and the resulting escape into drugs or cults, will require a new look at the psychology of individuals and society. New methods may be designed to spread positive attitudes and to promote a sense of values and ethics without falling back into old-style authoritarianism.

Although these practical applications aren’t the real subject of the book, we will have a quick look at some of these technologies in the final chapters. They will help us to understand the future development of society, while giving some concrete illustrations of the “action theory” component of the new world view. But first we must lay the foundations for that world view. And that requires that we look at the world views that preceded it, noting their successes, but also their shortcomings. By analyzing just how they fail to grasp the complexity of our present world, we will get an idea of what needs to be changed. Thus, we can build on their strengths, while avoiding their weaknesses. The following chapters propose a quick historical review of scientific and philosophical
thought. After glimpsing at the pre-scientific world views characterizing agricultural societies, we will look in more detail at the Newtonian world view associated with industrial society. Twentieth century science has uncovered different limitations of the Newtonian approach and proposed a number of alternatives. We will then try to synthesize these alternative ideas and develop them into a foundation for a new world view.
PART II: The Classical World

View

Fatalism and the Eternal Return

Life and Death in Agricultural Society

The development of human society can be divided in four major periods: hunting-gathering, agricultural, industrial and “post-industrial”. The futurologist Alvin Toffler has compared the transitions between these types of societies to “waves” of social and technological innovation: the first wave corresponds to the development of agriculture, the second wave to industrialization, and the present third wave to the emergence of an information society.

The first, “prehistoric” period is the one of the hunter-gatherers. This type of social system, which you can still find in a few groups such as the Inuit, Amazon Indians or Papuas, is characterized by a complete dependence on nature. Hunter-gatherers typically live in small nomadic tribes, which roam the land in search of food.

It is only through the domestication of animals and plants that humans got a minimal control over their food supply. This allowed them to build villages and cities, where larger groups could live on a permanent basis. Because agricultural production is more efficient and because its produce, such as wheat and cheese, can be stored for longer periods, it was no longer necessary for everybody to be busy with food production. This facilitated the development of different crafts and specialities, and the appearance of a market for the exchange of goods. It also led to the division of people into castes and classes, and to the emergence of hierarchical organizations, where smaller groups would be integrated into larger city states, counties, kingdoms and empires. The maintenance of trade relations and political structures required a more reliable medium for
communication and storage of information. This led to the development of writing, which is traditionally seen as the beginning of civilisation.

This type of organization can be found in different historic cultures, such as ancient Egypt, Mesopotamia and China, the Roman Empire, the Maya and Aztec civilisations, and the European Middle Ages. It also characterizes most of the countries that are counted as “Third World”, where the majority of people are still involved in agricultural production. All these societies have developed their own views of the world. They differ from the world views of hunter-gatherers basically by being more formalized. The invention of writing led to the appearance of holy books, scriptures, tablets of stone and other means to immortalize a society’s beliefs. Moreover, the division of labor led to the appearance of a specialized class of priests, whose task it was to maintain, develop, and spread these beliefs. This made it possible to develop very complex systems of belief.

The variety of world views was as large as the variety of communities, cultures and sometimes even villages. Because of the difficulty of communication, nearby groups would develop their culture largely independently, with little exchange of ideas. All these communities had their own myths about the origin of people and of the universe, their own local deities responsible for the different forces of nature and society, and their own list of taboos and commandments. Only the building of empires through conquest and the proselytizing drive of the great religions led to belief systems stretching over more than a few cities or counties. Even during the Middle Ages, when the influence of the Catholic Church extended over most of Europe, there was a great diversity of religious interpretations, movements and sects, some of which were violently combated as heresies. All this is just to indicate that there never was something like the agricultural world view.

Yet, in the mosaic of belief systems we can find a number of recurrent patterns. These derive from the way of life all these communities have in common. Although farming is a more reliable way of food production than hunting and gathering, many things remains uncertain. Depending on the vagaries of the weather, pests or diseases, harvests can be rich and plentiful or insufficient to feed the population. The all too familiar pictures of starving African children remind us that famines are common in primitive agricultural societies. The lack of technology for long term food storage or for importing produce from far-away regions meant that a poor harvest would necessarily cause many deaths.

Even when harvests were good, farming populations were regularly decimated by illness. In this respect, they fared even worse than hunter-gatherer groups, where the low population density hindered the spread of contagious diseases. During the Middle Ages,
Europe was ravaged several times by plagues that killed nearly half of the population. The Aztec civilization in Mexico is more likely to have been wiped out by diseases brought in by the Spanish conquerors than by their military might. Even without large scale epidemics, the poor hygiene and lack of medical knowledge meant that there was a high probability to die young. Life expectancies were rarely higher than 30 years. In the Middle Ages, more children lost their parents through diseases than presently through divorce. In short, life in these societies was quite insecure, and the odds that a newborn baby would live until its death from old age were very small.

This meant basically that people had little control over their life. There was little sense in making long term plans, since the odds of a major upheaval, such as bad harvest, disease, death of a close relative, or even an attack by bandits or invaders, were so high. They also had little control over their role in society. The class structure was very rigid, and it was practically impossible for someone to move to a position or occupation different from the one of his or her parents. In fact, people were so dependent on the community in which they lived that they were not really conscious of themselves as individuals, as unique human beings with the freedom to choose the course of their life. The brain theorist Julian Jaynes has even suggested that self-awareness did not exist before Classical Greek civilization.

The lack of communication media and education systems also meant that people had little knowledge about the world beyond their small community. Outside their village, the world was full of mysteries, which were more cause for fear than for curiosity and wonder. This feeling is reflected in old maps, where monsters are drawn in the corners outside the traditional travel routes.

In sum, people did not feel that they could control their fate. The events, positive or negative, that shaped their lives were the results of outside forces, which they did not understand and over which they had very little influence. A common characteristic of agricultural world views is fatalism: the belief that the course of events cannot be changed, and that it is better to just submit to the inevitable, rather than waste precious energy in a hopeless attempt to get your way. The Arabs concisely express this frame of mind in their saying “inshallah”, “if God wills it”. Still, it is part of human nature to try and get control over one’s life. The belief systems of the agricultural age are born out of the tension between attempts to predict and influence the world, and the resigned acceptance that that is not really possible.
Science and technology have given us control over nature by providing dependable knowledge, which allows us to predict what will happen under which circumstances. The problem with ancient agricultural societies was that they did not have any methods to develop such knowledge. There was no reliable way to distinguish true knowledge from fanciful superstitions. Their beliefs were based purely on tradition. The rules people followed were handed down from generation to generation. Their origin was shrouded in mystery. At best, there was a natural selection of traditions: groups carrying dangerously inadequate beliefs (e.g. that you can fly by jumping down from a mountain, or that the best moment to sow is in winter when the ground is frozen) were likely to die out. In the best case, they would be converted to a more successful system of beliefs. This guaranteed that the surviving beliefs were roughly adapted to the circumstances in which the community lived. However, they were certainly not the best possible.

Many traditions were superfluous or redundant, requiring the people to perform complicated but useless rituals. Yet, some parts of the ritual might effectively be useful. For example, the ritual might prescribe that seeds be sprayed with water after being planted in the field. But at the same time it might require that the planting be done wearing yellow clothes. Our modern, scientifically trained mind would immediately distinguish the first rule as practical knowledge, from the second one, as pointless superstition. But to people unaware of the canons of science, the only thing they could ascertain was that the ritual as a whole tends to have the required effect, namely to make the soil fertile. This result is of vital importance. Therefore, they would never dare to experiment by changing parts of the ritual, and thus running the risk of a failed harvest. Hence they would never find out that it was in fact the water, and not the yellow clothes, that fertilized the land.

Even if the ritual would be followed repeatedly by failed harvests, they would not be inclined to question the rules. There would always be good reasons why things went wrong. For example, the gods of the soil might be angry at the villagers and would first need to be appeased by a series of sacrifices. People would always come up with such ad hoc explanations, because they did not see the world as ruled by laws. Because they had no way to predict what would happen, everything seemed contingent, the result of fortuitous coincidences.

Insofar that they tried to find a more general explanation why something happened, they would attribute it to the will of a god. Originally, the gods were personifications of the different forces of nature, such as sun, rain, thunder, fire, the
forest, etc. Such a world view is called “animism”. It sees natural phenomena as
governed by agents or “persons” with a will of their own. If you want to influence them,
you will have to negotiate and bargain. If you want the god of the rain to do something
for you, e.g. water your fields, you will have to do something in return, e.g. sacrifice a
lamb.

Such a tendency to personify nature is understandable, since the only model for
complex behavior that people had were other people. The new discipline of evolutionary
psychology has even suggested that our brains are wired to think in terms of social ties
between people, rather than in terms of logical relationships. This hypothesis was
confirmed by a simple psychological experiment, in which people were asked to solve a
problem that required logical deduction. From two equivalent formulations of the
problem, the one involving people was solved immediately by everyone. The same
problem, formulated in terms of non-personal categories, was solved correctly only by a
small percentage of subjects. Journalists and writers are familiar with a similar
phenomenon: a text with “human interest”, that is, referring to people, is much easier to
read than a text about the same subject couched in more abstract terms. The reason for
the ease with which we reason about people may be that survival in small groups of
collaborating humans required a deep understanding of interpersonal interaction.

Still, people’s reactions are notoriously difficult to predict. If somebody does no
do what you expect or want, you would be inclined to believe that he is fickle or hostile,
and you will try to appease him, rather than develop a theory of how he should behave.
Similarly, if the natural world behaved erratically, people in primitive societies were
inclined to bargain, beg or plead with the underlying forces, rather than try to refine their
predictions. For example, during the Middle Ages, the only method archers used to
increase the accuracy of their shooting was a so-called “ejaculatory prayer”, a quick plea
to God that He would deflect the arrow toward the target while it was still moving
through the air. The fact that such requests to the gods were rarely granted, did not deter
people from trying. After all, you cannot expect a powerful authority to immediately
comply with all your wishes.

During the agricultural transition, the animism of the hunter-gatherers, in which
virtually every rock, stream or tree was personified, got gradually replaced by a more
abstract system of gods representing more general powers. Such a hierarchical system
reflected the hierarchical organization of society, with its kings, princes, counts and
village chiefs. People made little distinction between the gods in heaven and their earthly
rulers, like pharaohs or emperors. Both seemed equally remote, and had the same power
over the life and death of their subjects. Moreover, as everyone who has read about Greek
mythology knows, both gods and kings were similarly involved in the typically human power plays of fight, seduction and conspiracy. The mythical demigods, such as Hercules, who were of mixed human-divine descent, illustrate the fuzzy boundary between people and gods. Similarly, the most powerful of the rulers, such as the Egyptian Pharaohs, Alexander the Great or the later Roman Emperors, were officially pronounced to be gods.

As empires became larger and more centralized, the distance between ordinary people and their god-like rulers increased. At the same time, the power plays between different authorities became less visible. This may help to explain the gradual replacement of polytheist religions, with their mosaic of competing and conspiring gods, by monotheist religions, where power is concentrated in a single, absolute ruler. As we will discuss in the next chapter, that was the first step to a fundamentally new world view.

However, the deep-rooted tendency to personify uncontrollable powers did not vanish so quickly. In the Middle Ages, the official Catholic religion recognized only one God, but that God had three instantiations: the Father, the Son and the Holy Ghost. Moreover, people continued praying to a host of other divine beings, including saints, martyrs, apostles, angels, devils, and, most popular of all, the Virgin Mary. Similarly, although Buddhism is probably the religion that has gone farthest in doing away with personal gods (to the degree that it may be properly called atheism), in practice many Buddhist communities still worship a variety of local deities, saints and prophets.

Let me conclude by tying together the two threads of fatalism and personification. Fatalism is the recognition that the people in agricultural society have very little control over life and death, over the powerful forces that shape their destiny. Personification is a response to the psychological need to have a say in one’s situation. If you cannot predict or control the powers that determine your fate, you would at least like to imagine that they can hear your prayers and may possibly take into account your supplications. The saying “Man proposes, but God disposes” summarizes the situation. The result is that people think of control as something that resides outside themselves. Therefore, it does not make sense to try and improve your lot, or to get a better grasp of the forces that surround you. You can only pray, and hope that it helps.

Let me illustrate this attitude with an anecdote. During a stay in Sri Lanka, a tropical island off the coast of India, I noticed how carelessly drivers negotiated the traffic. I was not surprised to hear that, although the number of cars is small compared to industrial countries, the number of deadly road accidents is quite high. The driver of our bus was clearly aware of the danger. However, instead of increasing his chances by
driving more carefully, he would stop once a day at a small temple or shrine along the road and deposit some coins as an offering to the gods. This relieved him of his anxieties, so that he could continue to drive in the same reckless way.

Let me report another anecdote to show that such an apparently irrational way of thinking is not peculiar to Sri Lanka, but rather typical of agricultural societies in general. A Belgian friend of mine has spent a few years in Honduras, a small country in Central America. When he once announced his plan to travel along a certain road, people warned him that bandits were hiding around a particular crossing on that road. In response to his puzzled looks, they told him that everybody in the region knew that spot, that it was always the same one. He asked them why the bandits did not move, if everybody knew their hiding place. They answered that that is just the way it is, and how it had always been. Apparently, neither the bandits nor their victims felt able to change the situation.

**Cyclical Time**

The permanence of the bandits’ hiding place illustrates a general feature of agricultural society: its apparent changelessness. Compared to our age, the evolution of society in antiquity and the Middle Ages occurred at a snail’s pace. Technological and cultural innovation took centuries rather than years. Without science’s methods for systematic investigation, discoveries would happen purely by chance. The rare new inventions, such as the wheel or the bow and arrow, would take centuries, if not millennia, to spread from culture to culture. Although the Aztec civilization was quite advanced in many respects, Aztecs did not use wheels until the arrival of the Spanish Conquistadors.

As we noted earlier, the rigid maintenance of tradition meant that beliefs, habits and techniques would hardly ever change. The farmers in the fields would sow, irrigate and harvest in exactly the same way as their parents and grandparents had done. They would worship the same gods, listen to the same stories, and follow the same age-old rituals. This does not mean that nothing ever happened in their lives. There certainly were dramatic moments, such as invasions, droughts, famines and plagues. But the way their grandparents had died through disease, war or malnutrition was not essentially different from the way their parents had died, or the way they and their children would die. Though far-away kings and emperors sometimes might be killed in dramatic circumstances, triggering a struggle for their succession and possibly even the start of a new dynasty, this rarely made much difference to the lives of the farmers in the villages. Insofar that there was change, this change was erratic. There was no sense of progress, of continuing, irreversible development.
This experience of immutability was so strong that it even led some Greek philosophers to deny the existence of change. According to Parmenides and his disciple Zeno, change is an illusion. Only absolute, eternal Being is real. In our present changeful times, it seems hard to believe that intelligent people could take such a position seriously. Yet, Zeno proposed a number of sophisticated arguments in order to prove that movement, and by implication change, is impossible. It took more than 20 centuries, and the discovery of calculus, to solve these “paradoxes of Zeno”. It is not an accident that this solution coincided with the creation of the new, Newtonian world view, which we will discuss in the next chapter.

The complete denial of change was not the general rule, though. Some philosophers even took the opposite position. For Heraclitus, all things are in a continuous, ceaseless flux. There is nothing permanent except change. He compared this endless process to a flow, and noted that “you can never step in the same river twice”. A similar awareness underlies the Chinese philosophy of Taoism. It sees the Tao as a continuous process or flow that governs the whole of reality. Although such process-based views may seem very modern, they express the same passivity and fatalism as other ancient world views. You cannot change the course of the river. The best you can do is to let yourself flow with it. There is also no progress or evolution. Although the water changes, and moves turbulently in different directions, the river basically remains the same.

Although Taoism emphasized its transient, erratic nature, the world of the agricultural age was not completely unpredictable. Many things behaved in regular, repetitive ways. There were the 24 hour rhythms of day and night. There were the cycles of the moon, whose 28 day duration happens to coincide with the period between menstruations. Most important for a farmer, there was the 365 day cycle governing the seasons. Understanding this periodicity was crucial when choosing the moment to plant or sow. If the planting was done too early, the chances were that the germs would die from frost. If it was done too late, the crops might not reach full development in time for the harvest.

It is not surprising then, that the calendar was one of the first “scientific” methods which humanity developed to predict events. All agricultural societies had their own calendars, which were based on different systems producing different results, but which were all approximately reliable as a means to predict the seasons. They were based on the periodic movements of the heavenly bodies: the sun, the moon, the stars and planets. People had quickly noticed that the positions of stars and planets changed in extremely regular, predictable ways, unlike the movements of things on Earth. Increasingly
sophisticated mathematical systems were developed to accurately forecast the positions of the planets, for years, or even centuries, ahead. This would be the foundation for the science of mechanics, which in turn would bring about the Newtonian world view.

In antiquity, however, what struck people about these movements in the sky were not the underlying mechanisms, but the cyclicity: sun and moon, planets and stars move in circles, disappear behind the horizon in order to reappear the next day. This mirrored their experience of the cycle of life and death, in which plants, animals and people were born, developed and died, only to be replaced by a new generation that was to go through the same succession. It also matched the periodic decay and renewal of nature during its seasons, and of the moon during its phases. This awareness of endless recurrence was so strong that time itself was seen as cyclical. History was not seen as a progression, but as a mere repetition of patterns of growth and decay. In Hinduism, this recurrence was even extrapolated to life after death, producing the doctrine of reincarnation.

The combination of fatalism and cyclicity provided the motivation for Buddhism, which is perhaps the most sophisticated of the world views to come out of agricultural society. According to Buddhism, people are continuously trying to realize their aspirations, but ultimately doomed to failure. Even if they do manage to achieve their goal, they will soon start to aspire even higher. They will never be satisfied with what they have got, and therefore continue to suffer. This continuous striving to achieve new goals is endless, and will just repeat itself from one life to the next. The only way people can escape from this perpetual cycle of rebirth is to give up all striving or desire. Only then can they reach Nirvana, the enlightened state in which the individual is freed from suffering.

Such Eastern philosophies, and pre-scientific world views in general, appear attractive to many people nowadays. They seem to offer an alternative for the many shortcomings of the mechanistic world view. The New Age movement, which we discussed earlier, draws a lot of inspiration from them. What appeals most to present-day thinkers, is their holism, the fact that they don’t make a sharp separation between people and the different objects and forces surrounding them. This ties in with the present ecological consciousness. The insistence that people should not try to control things that cannot be controlled also strikes a chord with the many restless individuals, who continuously run after the things they don’t have yet, while getting bored by the ones they have.

However, it should be noted that such an attitude of resignation impedes progress. If you believe you cannot or should not control what is happening around you, that also means that you should accept any situation, however miserable, without trying to
improve it. Agricultural societies are characterized by their very slow, almost stagnant, development. Although holism may seem very modern, resonating with a number of recent findings in different sciences, it too hindered scientific and technological advance. In his famous treatise on “Science and Civilization in China”, Joseph Needham argued that, in spite of the many Chinese inventions, a Western style science could never have developed there. The reason is that the Chinese philosophy of the Tao, as an indivisible, pervasive flow, prevents any attempt at analysis, at separating causes and effects. The Tao’s erratic, unpredictable nature also discourages the search for universal laws. As we will discuss in the next chapter, it was just this concept of natural law that led to the birth of modern science.

Let me conclude this chapter by summarizing the different components of the “typical” agricultural world view. First, the model of the world is based on a tissue of interconnected and sometimes conflicting forces, often personified by different gods. The genesis of this world is described by an origin myth, but this is different for the different cultures. The future is seen as similar to the present, since history is essentially repetitive. Sometimes, the end of the world is predicted for a far-away future, but this usually includes the possibility of a rebirth and start of a new cycle. Since individuals anyway cannot control future events, there is no real theory of action. The only way people can hope to improve their lot is by pleading or bargaining with the gods. Similarly, there is no theory of knowledge acquisition. The only valid knowledge is the one sanctioned by tradition. The theory of values and morals too is based on tradition, and as such different for the different cultures. However, a recurring theme is that individuals should submit to the community to which they belong, and should remember their rigidly defined place in society. No particular value is placed on individual freedom.
The Clockwork Universe

From Laws of God to Laws of Nature

The foundation of the mechanistic world view underlying classical science is the belief that events in the world obey universal, unchanging laws. This stands in contrast with the agricultural world views, where events are the result of shifting conflicts and alliances between different forces or gods. The first step in the emergence of the new view was monotheism, the belief in a single God.

Historically, the origin of monotheism was not so much the belief that there exists only one god, but the belief (properly called “henotheism”) that the tribe should worship only one god. Different tribes, and even different towns, had their own particular god, just like different cities and professional groups each have their patron saint. These gods, like the tribes they were supposed to protect, were in competition with each other. Some groups managed to convince others that their god was really more powerful, and therefore more worthy of worship. The most successful of these groups spread their beliefs over large populations, thus effectively creating an area in which their god was the only one being worshipped. With rival gods out of the way, the remaining god really became the only one people believed in.

In parallel with this evolution, there was the development of large, centralized kingdoms and empires, headed by an absolute ruler. The analogy between the king as ruler of the country, and God as ruler of the world was readily made. The first rulers governed purely through personal decree. In more developed societies, such as the Roman empire, however, a formal system of laws evolved, which the emperor could not easily change. Such laws were supposed to be permanent, and to be universally applicable to all people in all parts of the empire.

A similar evolution took place in religion. At first, the God of the Old Testament, like the gods of polytheist religions, was constantly communicating with His subjects, telling them what to do depending on the circumstances. In a later stage, God’s decrees were supposed to have been written down once and for all in the holy scriptures, and His interventions were limited to exceptional circumstances (“miracles”). Since God was supposed to be omnipotent, His decrees should not only apply to the people who
worshipped Him, but to the whole of creation. Gradually, this led to the belief that God had created the universe together with all the laws that would govern its further development. Therefore, He no longer needed to interfere with it. If people would manage to unveil these laws, they would effectively read the mind of God, and be able to predict the further evolution of the universe in all its details.

In the final stage of this development, it became clear that if everything in the universe is determined by these laws, then there is no longer any role for God to play. The mathematician Laplace is famous for his claim that the Newtonian theory of mechanics, which he had helped to develop, would allow him to predict or reconstruct any event in the past or future, given enough data about the present. When asked by Napoleon where God would fit into this theory, he answered “I don’t need that hypothesis!”.

Ironically, the greater the power ascribed to God, the less need there was to explain events by His interventions. Thus, the laws of God had turned into the laws of Nature. This laid the foundation for the mechanistic world view. Let us have a quick look at how the main pieces of this world view fell into place.

To use laws as a method of prediction, you should first be able to clearly formulate these laws, and to deduce their consequences. Mathematics provided the perfect method for that. Initially, mathematics was merely a collection of rules of thumb that allowed you to calculate different amounts and sizes. The Greek mathematician Euclid was perhaps the first to formulate a major part of mathematics, geometry, in a systematic way. He showed how all the rules known at that time could be derived from a few simple, straightforward propositions, which he called “axioms”. These axioms, and the theorems deduced from them, capture the basic properties of size, distance and direction.

The resulting system of geometry could be used to describe and calculate the most complicated shapes. Practically, you could use it to design temples, pyramids and cathedrals. The success of geometry made it look like the most fundamental of all sciences. It even became an object of mystic worship for the followers of Pythagoras, who is well-known for his theorem about right-angled triangles. By the end of the Middle Ages, it inspired the early Freemasons to see God as the Great Architect of the universe. If God has not merely created, but designed, the universe, then this means that He used some kind of a plan or blueprint. And if this blueprint follows the sacred principles of geometry, then by studying geometry you might be able to understand Creation. The first scientific association in the world, the Royal Society of London, would grow out of the secret meetings of Freemasons. One of its members, Isaac Newton, was to lay the foundations of the modern scientific world view.
Geometry wasn’t yet science, though. It did not allow you to predict the events around you. Although mathematics can be applied to the measurement of concrete objects, it really only speaks about abstract shapes, such as circles and squares. The coherence and beauty of these abstract ideas led the Greek philosopher Plato to postulate that only they are real, and that the material objects we see around us are merely imperfect reflections of absolute, eternal Forms. Such a philosophy, however, disregards the study of concrete things. The philosopher Aristotle, on the other hand, understood the importance of observation, but failed to connect such observations with mathematical laws.

Nearly 2000 years after Aristotle, the English philosopher Francis Bacon, another early Freemason, formulated the principle of induction, that is, the method by which universal laws can be derived from concrete observations. The first to systematically use such methods was the Italian physicist Galileo Galilei. By turning observations into mathematical formulas, Galileo in a sense invented modern science. Near the end of the 16th century, he discovered the laws of motion, by deriving them from experimental data rather than from philosophical principles. One of these was the law of falling bodies, which determines the speed of falling as a function of the time since the object had started falling. To everybody’s surprise, his experiments showed that heavy objects fall with the same speed as light objects. (although this is probably incorrect, legend has it that Galileo carried out these experiments by simultaneously dropping a heavy and a light stone from the top of the leaning tower of Pisa.) This contradicted the old theory of Aristotle, which was based on the impression that heavy things fall more quickly.

An essential feature of Galileo’s new science was that it expressed measurable properties, such as position and speed, as a function of another measurable property, time. This was in fact a universal way of describing change. If the situations before, during and after could be measured in some way, this sequence of measurements would provide an objective representation of the process of change. The mathematician and philosopher René Descartes was the first to depict such functions in the now traditional way, as a curve framed by two orthogonal axes. (As a tribute to Descartes, such a system of axes is still called “Cartesian”). Conventionally, the horizontal axis represents the progression of time (independent variable), while the vertical axis represents increasing values for the property that changes as a function of time (dependent variable).

The next step in the mathematical representation of change was calculus, which was discovered, a few decades later, independently by both Isaac Newton and Gottfried Wilhelm Leibniz. Calculus made it possible to compute the speed of change at any moment for a quantity expressed as a continuous function of time. Inversely, given the
speed as a function of time, calculus made it possible to calculate the value of the quantity after a certain duration. With this mathematical tool, Newton could now accurately formulate a complete theory of mechanics, thus generalizing Galileo’s laws of motion. This theory could predict the movements of all bodies on the basis of the forces applied to them. As the most important application, Newton discovered the theory of gravity, which gave a precise expression for the force of gravitational attraction. This force explained both the fact that heavy objects fall to the Earth, and the fact that planets follow an elliptical orbit around the sun.

The cyclicity of astronomical trajectories turned out to be a special case in a whole range of possible movements. The simplest of those was just the uniform, straight motion, in which the distance from the starting point increases proportionally with the time that has passed. Another example was the parabolic trajectory of a cannon ball or arrow, which could now be accurately aimed, without need for “ejaculatory prayers”. But the periodic motion of the planets remained a paradigm case, an exemplary illustration of the regular, deterministic mechanism that was supposed to drive all change. Another way to visualize such a mechanism is a clockwork, a complicated array of interconnected wheels and gears, but which rotates in a perfectly uniform and predictable manner. In the Newtonian vision, the universe as a whole behaved like a clockwork. Perhaps God had originally wound up the spring, but the clock would go on ticking until the end of time without His help.

**Analysis: making distinctions**

In the centuries that followed, Laplace, Lagrange, Hamilton and many others extended Newton’s theory, making it more powerful and general, but without changing its basic assumptions. In a later stage, it became known as “classical mechanics” (thus being distinguished from some definitely “non-classical” theories, which we will discuss in a following chapter). Although its application was restricted to physics and engineering, it became a model for all sciences. The accuracy and reliability of its predictions, the coherence and mathematical beauty of its formulation, the scope of its applications, became the envy of scientists working in other disciplines, from chemistry and biology to sociology and psychology. It seemed that this theory was really the most basic of all, and that other theories in the end would be reduced to special cases of this more general framework. Moreover, although the mathematics had become quite complicated, the basic principles behind classical mechanics remained very simple and intuitive.
Thus, by the end of the 19th century it had largely determined the scientific view of the world. We will call the corresponding world view “Newtonian” or “mechanistic”. It was only in the beginning of the 20th century that some of its basic assumptions came under fire. But before we can discuss these criticisms, and the possible alternatives for the mechanistic view, we need to better understand its principles.

The laws of mechanics can only be used to predict the movement of unchanging, material objects (particles or “rigid bodies”). If you want to describe more complicated phenomena, you somehow must reduce them to collections of elementary material components. For example, if you want to model a living organism, you must find the atoms and molecules that constitute it before you can apply mechanics. If you want to describe changes different from movement, you must reformulate them in terms of changes in the relative positions of these components. For example, although the freezing of water is a change that apparently has nothing to do with movement, physicists can explain this transformation as a rearrangement of the positions of the H$_2$O molecules that make up the water.

This explanation of the properties of a whole (water) by the properties of its parts (molecules) is in fact a general method of mechanistic science. It assumes that all wholes can be analyzed or decomposed into their constituents. Descartes had already advocated such an analytic method a generation before Newton. In his famous “Discourse on Method”, he advised investigators “to divide each of the difficulties under examination in as many parts as possible”. By studying the parts one by one you may notice essential properties which would otherwise have escaped your attention. By systematically listing all the components, their properties, and their interactions, you will get a much deeper understanding of a complex phenomenon, and the way it behaves. It was this insight which lacked in the more holistic philosophies, such as Taoism. In short, the way to understand a phenomenon is to separate or distinguish as precisely as possible its different parts, aspects, and properties.

The mechanistic view assumes that after you have analysed the system into its components, you can get back to the whole by simply adding all their properties together. For example, the mass of a liter of water is the sum of the masses of all the water molecules it contains. In the Newtonian world view, the whole is basically equivalent to the sum of its parts. In mathematics, this property is often expressed as linearity: when you calculate the value of a linear function or the solution of a linear equation, the result you find for a sum will be equal to the sum of the results you find for the elements. Although functions in mechanics are not linear in general, linearity is often a good approximation, which makes the calculations much easier.
The analytic method allows you to reduce the properties of wholes to the properties of their parts. These parts can themselves be reduced to their parts, and so on. Therefore, such a reduction generally happens in stages, where a higher level description is reduced to a description of the level below, which is itself reduced to an even lower level description. The typical sequence is the following: social systems (sociology) are explained as patterns of individual behavior (psychology); behavior is explained by the interactions between the neurons in the brain (biology or physiology), which are explained in terms of chemical reactions between molecules (chemistry), which are explained as rearrangements of atoms (physics). The general philosophy underlying this approach is called reductionism. It assumes that the laws governing atoms determine the laws of chemistry which determine the laws of biology which in turn determine the laws of psychology and sociology. In that sense, progress in physics is supposed to directly lead to progress in all other sciences.

The assumption that physics is the lowest level, supporting all other sciences, does not yet tell us how the physical level itself is supported. Material objects can in general themselves be reduced to combinations of smaller objects. These components might move independently of the movement of the whole. This movement of the parts would affect the movement of the whole, and thus perturb your predictions. The only way to make sure that your theory of mechanics can really explain and predict everything is to apply it to the lowest level of all, to the smallest material components. This means that you should distinguish the parts as finely or precisely as possible. However, it is not obvious that you would find the smallest things: perhaps you could go on splitting up matter into smaller and smaller parts, without ever getting to the lowest level.

The Greek philosopher Democritus was the first to suggest that such a level exists, that matter is composed of parts which cannot be further split. He called these parts “atoms” (meaning “indivisibles”). By the end of the 19th century, it seemed that science had discovered such atoms, as the smallest components of the chemical elements. However, 20th century researchers quickly discovered that atoms contain even smaller particles, protons and neutrons, which themselves consist of quarks, the provisionally lowest level. At the moment it is still not clear whether such “elementary” particles, are really the smallest components of matter. Although the issue has not been completely settled yet, the philosophy of atomism has worked very well in practice. When explaining higher level phenomena, such as planets, rivers and organisms, the levels lower than atoms can be mostly disregarded, since they have very little influence on the behavior of the atoms.
The typical image of particles or atoms in mechanics is that of hard, spherical balls, a little like miniature billiard balls. The characteristic of balls is that they look the same from all directions, because they lack internal structure. Newton’s most successful application, the planets, also looked like balls, at least at the distance from which he studied them. We will regularly use the example of the billiard game to illustrate principle of mechanics.

Associated with atomism and reductionism is the philosophy of materialism. All objects described by Newtonian mechanics have one property in common: they have mass. That mass shows itself in their weight (that is, the way they are pulled down by gravity), and in their inertia (that is, the way they resist acceleration). Mass is additive: the mass of a large object is simply the sum of the masses of its components. Another universal property of objects is that they occupy space. Again, occupied space is additive: the space filled by a composite object is the sum of the volumes filled by its parts. Therefore, it seems that all objects, atoms and particles are merely bits and pieces cut from one and the same massive, spatially extended substance. That substance is what we call “matter”. The idea that all higher level phenomena can be reduced to atoms then also means that they ultimately consist of matter. Thus, matter is the essence, the primary substance, of the world.

However, Newton’s theory of mechanics needed more than material particles to explain movement and change. The second basic component was the concept of “force”. One way to exert force on an object is simply to make another object bump into it, like when you hit one billiard ball with another one. In this case, forces can be reduced to movements of matter. On the other hand, Newton’s celebrated theory of gravitation assumed the existence of a gravitational field, which exerts an “action at a distance” on massive objects. Indeed, the planets that orbit the sun are not kept in place by strings of matter, but by immaterial forces of attraction. Another way to express the concept of force is through the concept of energy: the movement induced by a force can be seen as the transformation of potential energy into the energy of motion.

How can we reconcile these mysterious forces or energy fields with the materialist world view? Twentieth century physics has shown that energy and matter are two aspects of the same thing. The most recent theories of elementary particles moreover treat interactions via force fields as exchanges of a special type of particle. Thus, gravitational attraction can be reduced to an exchange of “graviton” particles. It seems that in this respect at least, 20th century science has vindicated an essential presupposition of the mechanistic world view. As we will discuss later, however, this result came at a considerable price. Some of the most basic assumptions regarding time,
space and causality had to be given up in order to reach it. Let us discuss these assumptions in depth before proceeding.

**Causality: conserving distinctions**

The idea that everything consists of matter did not contribute in itself to the success of the Newtonian world view. After all, when Democritus suggested that everything is made out of atoms, he was not taken any more seriously than his colleagues philosophers who argued that everything is made out of water, out of fire, or out of geometrical forms. What things ultimately consist of is less important than how they behave. If you can predict how something will behave, then you can influence that behavior, and steer it in the direction that you want it to go. That gives you power over it. It was because it offered power over Nature that the Newtonian view had such an impact on society.

Mechanics allows prediction by representing all processes of change as mathematical functions, where each moment in time corresponds with a specific state of an object. In classical mechanics, the state of a particle includes both its position in space and its momentum (velocity times mass). More generally, the state of a system consists of the precise properties that determine its further evolution. The state is for the properties of a system what the atom is for its components: it is the smallest, most fine-grained distinction you can make. If you do not distinguish precisely enough, you may miss properties that will influence the further behavior of the system, and therefore perturb your predictions. For example, if you would only know the position and speed of an object, but not its mass, you would not be able to determine how a collision with another object would affect its movement. Only the state will give you all the information you need to make perfect predictions.

Once you know the present state, you can derive everything else, either past or future, from the state function. The state function describes a trajectory, a line connecting the subsequent states of the object, as it advances in time. Depending on the state the system starts from, it will follow a different trajectory (see Fig. 1)
Fig. 1: different possible trajectories of an object, where its state s changes as a function of time t.

Trajectories in mechanics have a number of peculiar properties. First, they are unambiguously determined. Whatever will happen to the object in the future is already fixed at this instant. There is no freedom to explore alternative trajectories. There is no uncertainty as to the trajectory that the object will follow. Not just the future, but the past is determined. If we know the present state of the object, then we can reconstruct or predict its state at any moment in the past or the future. The belief that all future events are fixed beforehand is called determinism. In principle, it allows no freedom for God nor man to intervene.

In practice, this means that trajectories cannot intersect. Indeed, at the point where two trajectories would cross, there would be a bifurcation, a fork, with two possible future developments (see Fig. 2). An object reaching that bifurcation point would have to choose which trajectory to follow, since it cannot be in two places at once. But the theory of mechanics is based on the principle that there is no choice, no uncertainty, no freedom. So, trajectories must be parallel. They never meet. This means that if two objects start out on two separate trajectories, their trajectories will forever remain separate (see Fig. 1). Another way to formulate this principle is by noting that the separation or distinction
between trajectories always remains. In my book “Representation and Change”, I have called this property *distinction conservation*.

Fig. 2: two intersecting trajectories. In the point p, the system has the choice to follow either trajectory s1 or trajectory s2. (This is not allowed in classical mechanics.)

Distinction conservation is equivalent to *causality*, to the principle that causes completely determine their effects. The traditional way to express causality is the statement that “equal causes have equal effects”. If two objects start out from the same initial position (cause) on the same trajectory, after a given duration of time, they will end up in the same final position (effect). But the principle of distinction conservation does not care whether you look towards the future or towards the past. Trajectories can intersect neither in the past nor in the future. So you might as well note that two objects ending up at the same final position must necessarily have started out at the same initial position. Otherwise, it would mean that different trajectories somehow have come together in the same end point, which mechanics does not allow. Causality in the mechanistic view therefore also means that “equal effects have equal causes”.

There is another, equivalent way to formulate causality. Causality is often associated with *covariation*: if you change something about a cause, the effect will vary as well. For example, imagine that you take part in a battle. Suppose that you fire a cannon, and see an explosion half a mile further. You might assume that your firing has
caused the explosion. However, the cannon might have been empty or wrongly adjusted, and the explosion caused by one of the many other cannons taking part in the battle. If you repeat the procedure and see another explosion in the same spot, it becomes more likely that it is your firing that caused the explosion. Indeed, it confirms your expectation that equal causes (firings) have equal effects (explosions). However, this could still be because different cannons are all aiming at that same spot (equal effects but with different causes). To make sure that it is really your firing that causes the explosion, you might turn the cannon a little to the left. If the explosion now also moves to the left, you have established that firing and explosion covary: they change together. This can no longer be explained by the firings of other cannons, since their cannoneers do not know that you changed your aim. Therefore, it establishes a causal relation between your firing and the explosions in a particular direction.

This application of covariation can also be expressed as “different causes (cannon orientations) have different effects (explosion areas)”’. This is logically equivalent to “equal effects have equal causes”. Together with the “equal causes have equal effects” which you established earlier, you may conclude that “equal causes have equal effects, and vice versa”. Equivalently, you may conclude that “different causes have different effects, and vice versa”. All these expressions are equivalent formulations of the principle of distinction conservation: the distinctions between causes are reflected in the distinctions between their effects and vice versa. And this is again another way of saying that trajectories never intersect.

The example of the cannon shows that such reasoning is of more than philosophical importance. If you know that there exists a distinction conserving relation between a “cause” event and an “effect” event, then you also know how you can change the effect by influencing the cause. Thus, you get control over the effect. Newtonian mechanics provides a wonderful tool for finding out which precise cause-and-effect relations there are.

The “equal causes have equal effects” requirement in fact ensures predictability: if you know the cause, you can predict its effect. On the other hand, the complementary “equal effects have equal causes” requirement ensures “retrodictability”: if you know the effect, you can reconstruct its cause. (the difference is that you cannot influence the cause by changing its effect). In fact, in classical mechanics the laws that determine the trajectories don’t care about the direction of time. Any movement or change in a particular direction might as well have happened in the opposite direction. The movement of a stone thrown upwards is simply the mirror image of the movement of that same stone falling down. The satellite orbiting in clockwise sense might as well have been orbiting in
counterclockwise sense. Looking at a video recording of such a movement, you cannot say whether the tape was played forward or backward: both directions of movement are equally plausible. This is called *reversibility*: all processes described by Newtonian mechanics can be inverted.

Of course we know that certain changes cannot be turned back. Watching a movie where shards of glass would jump up from the floor while mysteriously assembling themselves in the shape of a bottle, you would know that the tape was being played backwards. But such processes simply don’t fit into the Newtonian theory. In classical mechanics nothing ever gets lost. Nothing new is created either. There is no progress, no arrow of time, just the endless rearrangement of unchanging atoms. Although the speeds and positions of the atoms may change, everything else remains the same. Most fundamentally, all distinctions—between atoms, between positions, between velocities, between trajectories—are conserved. In other words, all information is conserved. With the information you got about a system’s state at this moment, you also have all the information about the system in the future and in the past.

In this world view, space is just a fixed arena in which the atoms move. The different positions or points in space are distinguished once and for all. They are determined for every atom and for every observer. Positions in space can be expressed in numbers. If you agree about the point of origin and the directions in which you measure (say, forward, left and up), every point can be characterized by three coordinates. These coordinates are unambiguous, and should be the same for every observer. Similarly, time is expressed unambiguously by numbers. If you agree about the moment you start counting (say, the birth of Christ) and the unit of measurement (say, a year), everybody will connect the same moment with the same number (say, 1997). The difference is that time only requires one coordinate. It is represented by a single dimension, a straight line that extends infinitely, without cycles or curves. The time dimension is independent of the three dimensions of space. If you add to the three coordinates of space the three coordinates of momentum, you have completely specified the state of a particle, and thus determined its trajectory. All change can therefore be reduced to numbers. There is no change in quality, only in quantity.

**The Ghost in the Machine**

This view of the universe as a clockwork, as an intricate but perfectly predictable machine, had one big problem. By revealing the laws of causality, it allowed people to
manipulate causes and effects, and thus to intervene in the processes of nature. However, it could only do so by assuming that everything is determined, and that there is no freedom to intervene. The hard-core mechanicists would argue that there is no contradiction. They would say that our freedom to act is only apparent. People too are completely determined in their thoughts and actions, and if you know enough about all the atoms that make up their body and brain, you would be able to precisely predict all their actions.

Although such a philosophy seems logically consistent, it does not fit in with the requirement that every world view should include a theory of values, of morals, and of actions. If all our actions are already determined, there is no sense in formulating rules of behaviors or setting goals or values to strive for. This would also mean the end of personal responsibility. The criminal who just killed a child could say that he had no choice, that he was predestined to commit this crime. It would not make sense to punish him, because whether we punish him or not, it would not stop him or anybody else of committing eventual further crimes. In conclusion, such a completely deterministic philosophy is difficult to apply to society.

It is hard to swallow for an individual too. People do have an experience of free will. They feel that they can make their own choices, and decide which actions to take. These actions may not always have the desired effect, but at least it seems possible that they would make a difference. How can you reconcile this idea of freedom with the determinism of mechanics?

Descartes proposed a way out of the dilemma. Although in his days the theory of mechanics was still rudimentary, Descartes knew enough about clockwork mechanisms to understand that they provide a very powerful way to explain natural processes. He believed that even animals were basically sophisticated, automatic machines. However, he did not accept that such mechanisms could ever explain or determine the human mind. Therefore, he separated it from the rest of the world, and distinguished two independent realms, mind and matter. Such a philosophy is called dualism. The deterministic laws of mechanics govern the material part of the universe, but not the mind. Thus, dualism allows the existence of free will without negating mechanics.

However, this solution to the free will problem created a new problem: if mind and matter are separate, then how do they interact? Clearly, our mind controls our body, which is made out of matter. On the other hand, we perceive material objects through our senses, and thus they affect our mind. The problem of perception is basically the problem of knowledge: how does our mind acquire knowledge about external, material phenomena?
The answer given by Descartes and the philosophers following him may be called the *reflection-correspondence theory*. The basic intuition is that if you perceive, or think about, an object, then what is present in your mind is not the object itself, but an image or reflection of it. This image is initially created by the senses and then stored in memory. The mechanism is similar to a photographic camera: the light coming from the object passes through a system of lenses (sense organs), and is projected on a screen (consciousness), where it leaves a permanent trace in the light-sensitive film (memory). The observer in this view is basically passive, absorbing information through the senses without interfering with the thing being observed. It is as if the observer is sitting in a box, insulated from the world, and looking at it through a thick glass window. Because of this separation, mind and matter can work independently, each obeying its own laws.

True knowledge, in this view, is merely an accurate image of the world. For every object in the world there should be a corresponding reflection in the image. A false theory, on the other hand, is like a distorting mirror, where the image does not correspond with the reality it is supposed to reflect. Some objects in the world will not be reflected in the image, while some reflections in the image will not correspond to existing objects in the world. The correspondence between world and image is objective and absolute: if people would observe precisely enough, there would be no room for interpretation or disagreement about what is true and what is false.

This philosophy assumes that there is no limit to the amount of knowledge you can gather. After all, the only thing you need to do to get more knowledge is to look around and accurately observe and memorize the things you see. As long as you make sure that the image is not distorted, there is nothing that stops you from adding ever more observations to your store of knowledge. In principle, you should be able to gather all information there is about any particular phenomenon. If you then apply the laws of mechanics, you should be able to perfectly predict the reactions of the phenomenon to whatever action you would like to apply to it. This means basically that whatever problem you have, you should be able to find the best possible solution. This is the assumption of perfect *rationality*, which is at the base of classical economics. Given a series of alternatives, you should always be able to determine which one will bring you closest to your goal.

But to implement your decision you should be able to act upon the world. This is the second strand of the mind-matter interaction. How does the mind affect the body, and through it the outside world? The dualist philosophers have never found a satisfying answer to this question. Today we would formulate the problem as the interaction between mind and brain, but in the days before Descartes, it was still believed that
cooling the blood was the main function of the brain. Descartes himself suggested that the organ through which the mind communicates with the body was the pineal gland.

The location of the interaction does not really matter, though. The fundamental problem is that the laws of mechanics completely determine the material world. If the mind would be able to affect matter, this means that it should somehow be able to overrule these laws. But that implies that mechanics would lose its universality, which is perhaps its greatest strength. Now, you might give up part of your theory if you receive something better instead. But the concept of mind is just one big mystery, about which Newtonian science has virtually nothing to say. The philosopher Gilbert Ryle has described the situation as “the ghost in the machine”. The body is basically a clockwork, a material, deterministic mechanism, which is mysteriously animated by something we cannot grasp, a “ghost”. Such a world view is not very satisfactory. The only alternative within the Newtonian approach is to take reductionism seriously, and argue that the mind is nothing more than a complicated arrangement of moving molecules. But that brings us back to denying the existence of free will, and therefore the possibility of any meaningful theory of ethics or action.

In practice, scientists assumed that people could choose their own actions, but that they would do that in the best possible way. Since there is in general only one “best possible” solution, this brought back some form of determinism to human decision-making. The mechanistic view of human action assumes that all possibilities and their consequences are perfectly understood, and that there is an unambiguous criterion that determines which possibility is the best. The decision then becomes trivial, and the result is beforehand determined. The only problem that remained was to specify the evaluation criterion. This criterion is usually called “utility”, implying that people would somehow choose the most useful alternative, but without specifying what it should be useful for. The utilitarian philosophers have tried to define utility more precisely as “the greatest happiness for the greatest number”. But that leaves open the question of what “happiness” means. Economists have a more down-to-earth view, defining utility as that what consumers try to maximize when buying goods.

The analysis of human behavior as trying to maximize some abstract quantity fits in well with mechanics. Indeed, the unique trajectory of an object in mechanics can be calculated as that trajectory which minimizes the “action”, a complex quantity related to energy. Minimizing the action is the same as maximizing the absence of action. All are examples of optimizing, of determining the one possibility which is somehow the “best” for a given criterion.
However, there is an essential difference between optimization when applied to people and when used in mechanics. “Utility”, whatever it may mean, is a quantity which tends to grow irreversibly. The benefit you get from the things you have bought tends to accumulate. “Action”, on the other hand, does not accumulate or diminish. It cannot even be defined for a particular state or moment, only for a trajectory. If it would accumulate, there would be an essential difference between past and future. But the reversibility of mechanical movement does not allow that. So we find another fundamental inconsistency in the mechanistic world view: on the one hand, it assumes that human action is driven by constant improvement, on the other hand, it assumes that on the level of matter no progress is possible.

Let us conclude this section by summarizing the components of the mechanistic world view. The world is a collection of interacting material particles or atoms. All other phenomena, such as society, people, life, chemistry, etc. can be reduced to combinations of these particles. All change can be explained by the laws of nature governing the movements of the particles. The future is essentially similar to the past, although quantitative progress is implicitly accepted for human action. Maximizing utility is the fundamental goal or value. The theory of action assumes that you can control nature by varying the causes so that their effects come as close as possible to the goal. To acquire knowledge, you simply need to observe precisely and assemble your observations into an image or reflection of reality.

**Industrial Society**

Compared to the agricultural world view, mechanism did away with the feeling of changelessness and passivity. Although the underlying framework was still static, Newtonian science was based on movement, on time as an infinitely extended line, rather than an endless cycle. People were no longer at the mercy of forces they did not understand. Through analysis and observation they could discern, predict and eventually control these forces, thus improving their lot. As a result, the emergence of the Newtonian world view was accompanied by a sudden speeding up of economic and technological progress. This eventually culminated in the industrial revolution, and the birth of modern, industrial society.

Although the idea of progress itself does not completely fit into the mechanistic frame, its appearance in society inherited many Newtonian features. First, progress was seen as basically material: the production of resources and goods. Progress was moreover quantitative: you can measure it with precise numbers, such as the amount of
goods produced, or their total worth in money. It was also objective: everybody agrees on what constitutes improvement, or “increase of utility”. Together, these features are captured by the economic concept of “scarcity”, which emerged at about the same time as Newtonian science. Economists assume that the supply of goods we need to lead to a happy, satisfied life, such as food, water, clothes, and houses, is necessarily limited. In general, we need more of them than is available. Therefore, progress is synonymous with alleviation of scarcity, that is, growth in the supply of goods.

To increase supplies, mechanistic principles were applied to make production more efficient. Where possible, production was automated according to the machine model. Even the people were organized in machine-like ways. This organization rested on three principles: standardization, compartmentalization and synchronization.

Standardization means that different objects and processes are made to behave uniformly. Like all particles obey the same laws of nature, and react to the same causes with the same effects, all products or components of the same type should have exactly the same properties. This makes their behavior predictable, and therefore controllable. Every object produced by a craftsman, whether it is a pot, chair or cupboard, is slightly different from the previous object. The same type of object produced in a factory will be identical to all other objects produced in the same line. Such mass production is much more efficient, since the same set-up can be used to manufacture an unlimited number of items. Its products are also more reliable. For example, you would not like to drive a car which does not fulfill all the same safety criteria as the other cars of the same category. Finally, standardization facilitates interoperability. For example, because tire types are standardized, you can safely combine not only different tires but even tires from different producers on the same car.

Compartmentalization is a direct application of the analytic method. By splitting up processes or objects in as many components as possible, you get a better understanding and control of the whole in all its fine details. Thus, manufactured objects will not be produced as a whole, but assembled out of standardized components, which generally will themselves have been assembled out of smaller components. This means that the production process needs to be divided into separate component processes, which are carried out by different people or machines. This entails extensive specialization.

It leads to the organization of production in assembly lines, where the components move automatically on conveyor belts through the subsequent locations. At each stage in the sequence a worker or machine carries out a particular operation, e.g. adding a strip of paint, tightening a screw or putting on a lid. This same operation is repeated continuously throughout the day, at the highest speed the worker can manage. Such assembly lines are
very efficient, but rather unpleasant for the people that are part of them. They effectively mechanize not only the production process, but the behavior of the people involved. Any form of personal variation, creativity, or freedom of action is banished. Charlie Chaplin’s movie “Modern Times” dramatizes the effects of such mechanization on human psychology.

The mechanical way of production also requires extensive synchronization. All workers on the assembly line should act in the same rhythm. If one works a little more slowly than his co-workers, the whole line is delayed. But synchronization is not limited to assembly lines. The Newtonian world view is based on the measurement of time. In the agricultural age, time was measured only in the most approximate ways. For a farmer on the fields, it did not matter much whether work started at 7 am or at 9 am. The only thing that mattered was that at the end of the day, the work was done. Therefore, there was no need for precise measurement. Rather than using hours, minutes and seconds, people would express time intervals in vague units such as a “cow milking time” or a “pateroster wyle” (the duration of a prayer). For accurate prediction and control, however, time must be determined precisely, and in the same way for everybody. Only if everybody continuously uses a clock, which is synchronized with all the others, can work be accurately planned. The industrial revolution brought about a strongly enhanced awareness of time. Children were taught that they should always show up on time, in class, at work or at an appointment. The day was divided in standard 9 to 5 intervals for work, with fixed periods for breakfast, lunch, dinner and sleep.

Space too was determined more accurately. In the Middle Ages, distances were still measured with subjective units such as an arm’s length, a stone’s throw, or a day’s ride. One of the first acts of the new government after the 1792 French revolution, just before industrialization, was the introduction of the metric system to standardize units. Precise measurement of space also led to the establishment of clear borders between regions. In the Middle Ages land had been divided in a patchwork of counties, duchies and bishoprics, with ambiguous and changing boundaries. The difficulty of movement and communication did not allow kings and emperors to maintain tight control over all the parts of their realm. Laws, regulations, even languages would vary from one county to the next. During the 19th century, norms and institutions were standardized over the whole country. In places like Germany and Italy, where there had not been any large scale, central government, the smaller units were integrated into a larger whole. These wholes were separated from the neighboring countries. This led to the creation of the nation state, which is still the fundamental political unit in the world.
Similar processes of integration and uniformization led to the creation of large companies and institutions, first nationally, then multinationally. Such huge organizations required complex administrative structures. Although bureaucracy existed in Egypt, Rome and China long before the Newtonian world view, mechanicism and industrialization perfected its function. The same principles of standardization, compartmentalization and synchronization were applied to make it more efficient. Functions were broken up into subfunctions, which were delegated to specialized departments or individuals. Each function was clearly defined and as much as possible separated from the others. In order to maintain coordination, the different people would report to a person at a higher administrative level, which would oversee and integrate their results. But since there were so many people involved, and each supervisor could supervise only a limited number of people, a next level of supervisors had to be created, and so on, producing towering pyramids of organizational levels.

We all know the effects of industrialization, positive as well as negative. On the material level, industrial society led to unprecedented growth and development. On the psychological, social and ecological levels, however, there were many negative side effects. The main reason is that the mechanistic world view does not know how to deal with non-material phenomena, such as culture, happiness, or interdependency. It fails to understand the qualitative aspects of development. Everything is reduced to a continuing increase in quantity. This leads to the rapid exhaustion of resources, which we discussed in the first part of this book. It is also responsible for ecological destruction caused by pollution and other side-effects of the maximization of production.

Psychologically, the mechanization of society leads to alienation, to the feeling that you don’t really belong to the society around you. Although industrialization increased the control of society as a whole over nature, paradoxically it may have diminished the control an individual has over his or her every-day activities. The farmer can decide autonomously when to milk the cows, where to store the wheat, or what to plant in the vegetable garden. The worker at the conveyor belt does not have any choice when, where or how to carry out the work. Everything is regulated beforehand in order to maximize production. There is no longer any real interaction between individual and environment. Moreover, for people living in the industrial cities, the environment is reduced mostly to brick or concrete buildings, straight, geometrical roads and canals, uniform housing. Gone is all the richness, variety and creativity of nature. The uniformity and rigidity of mechanism is mirrored in the structure of organizations. Laws and regulations are formulated once and for all. Everything that does not fit into the
predefined cases is ignored, or forced into a category where it does not belong. This leads to the many absurdities typical of complex bureaucracies.

Everybody complains about the rigidity, inefficiency and facelessness of bureaucracies. However, there are few alternatives to bureaucratic organization when it comes to the management of a large company or institution. The problem of bureaucracy is basically the problem of how best to cope with complexity. The mechanistic paradigm, with its analytic method, has helped us to conquer problems which were too complex for earlier approaches. However, when things get really complex and changeful, the mechanistic approach fails as well. The reduction to changeless particles or components is simply insufficient to explain the complex, interdependent developments which we discussed in the first part of this book. Yet, such reduction is the essence of the mechanistic world view. Therefore, we can only hope to manage the complexity of present society by starting from a radically different world view. Just like Newtonian science provided the foundation for the mechanistic world view, post-Newtonian science may help us to build its successor. Let us therefore review the major scientific insights which transcend mechanism.
PART III: The Emergence of a New World View

The Breakdown of Classical Mechanics

The Heat Death of the Universe

In spite of the impressive successes of classical mechanics, physicists soon discovered a number of phenomena that did not fit into its framework. When they found such an anomaly, they tried to reduce it to a classical description. However, in a few cases they were forced to develop an alternative theory. Each time, the divergence between the new theory and the old one was kept to a minimum. After all, the old one was so powerful, so well-understood, so logically coherent, that it would be silly to give up more of its advantages than strictly needed. So, they developed a number of theories that were like classical mechanics in all aspects, except one or two.

But those “non-classical” aspects were really different. Some of them were so weird from the classical point of view, that the people who developed these theories still don’t really understand them. The theories work, their predictions are accurate, but the intuition of what they really mean has gotten lost on the way. As a result, present-day physics is a strange conglomerate of three or four different conceptual frameworks, which each work very well in their own domain. Yet, they have little in common, apart from the features they all inherited from classical mechanics. Physics as a whole no longer provides a coherent picture of the universe. Yet, it is clear that the partial theories it has developed over the past century provide important building blocks for the new world view we are trying to assemble. We will discuss these extensions to classical mechanics in the historical order of their development.
The first of these theories is thermodynamics, and the associated theory of statistical mechanics. Thermodynamics studies the flow of heat. In the 19th century, scientists like Carnot and Clausius noted that the energy of movement can be converted to heat, for example, when a gas becomes hot during compression. Inversely, heat can be converted to movement, for example when hot gas is used to drive the piston of a steam engine. Yet, they observed that the second conversion is never perfect: there is always some heat that dissipates in the environment. Therefore, if you convert mechanical energy to heat and back to mechanical energy, the energy you end up with will be less than the one you started with. If you build a motor working on such a conversion of energy, the motor will slow down and eventually stop. The only way to keep it running is to provide it with additional energy, for example in the form of fuel or electricity. In other words, it is impossible to build a perpetuum mobile, a machine that runs forever without energy input.

The scientists summarized these observations in the two fundamental laws of thermodynamics. The first law says that energy is conserved. Whether you convert it to heat, to movement or to electrical energy, the total amount of energy in a closed system always remains the same. No energy is lost, and no energy is created. Therefore, if there is not enough energy in the system, a machine will not start moving on its own. The second law of thermodynamics, however, adds that the energy that is present tends to “degrade”. This is what happens when the motor slows down: although the energy in the system is still there, it has taken on a form (heat) which is more difficult to use for movement. Somehow, the quality of the energy has diminished. The two laws together explain why a perpetually moving machine cannot be built. The second law explains why the machine will slow down, while the first one explains why it cannot be speeded up again without fresh supply of energy.

The first law in fact follows directly from classical mechanics. Energy conservation is a special case of distinction conservation, which is the basis of Newtonian science. The second law, however, contradicts the principle of reversibility, which says that any movement or evolution could as well go in the opposite direction. The reverse of spontaneous slow-down is a spontaneous speed-up. But no one has ever witnessed a machine starting to move on its own, without energy supply. More generally, the second law contradicts distinction conservation. Indeed, imagine different motors working at different speeds, that is, having different states. Eventually, they will all slow down and come to a halt, that is, reach the same state. So, their trajectories through the space of states, which are initially separate, eventually come together in the same spot. The distinction between their states of movement vanishes.
Such behavior poses a serious challenge to classical mechanics. Indeed, Newtonian science sees the universe as a huge perpetual motion machine. The idea that that machine might ever crawl to a halt is alarming. That situation, where all movement would have stopped because its energy was converted to heat, has been called the *heat death* of the universe. If the universe is a clockwork, the heat death describes the situation where its spring has run down.

To see where the contradiction between mechanics and thermodynamics comes from, we must better understand the role of heat. Heat can only be used to generate movement when there is a difference in temperature. For example, the higher temperature in a cylinder creates a higher pressure, which will push forward a piston. However, temperature differences tend to disappear. Indeed, the thermodynamics researchers observed that heat always flows from a warmer body to a colder body, but never the other way around. Thus, the warmer parts cool down, while the colder parts warm up. Whether you add boiling water or an ice cube to your drink, after a few minutes warm and cold will have mixed and the liquid will have gone back to a uniform temperature. This dissipation of heat erases all temperature differences and therefore all possibility to generate movement. In thermodynamics, the resulting homogeneous distribution of heat is measured through the concept of *entropy*. The more uniform the distribution, the higher the entropy. When all temperature differences have gone, the entropy is at its maximum. The second law of thermodynamics can then be formulated more precisely by the statement that the entropy of a closed system tends to increase.

Such a reasoning in terms of temperature and heat was quite alien to mechanics, where everything is expressed in terms of particles and movement. To get to the roots of the paradox, it was necessary to reduce heat to movement. At first, physicists believed that heat was some kind of a special substance, “caloric”, which would flow like a liquid from one object to another. Later, they discovered that heat is just another form of energy of movement: not the movement of the object as a whole, but the independent movements of the different atoms and molecules out of which the object is made. Although the object as a whole may seem motionless, its microscopic components are continuously vibrating in different directions or even colliding with each other. Temperature measures the average energy of motion of each particle. Heat measures the total energy of all the particles together. If one part of an object is hotter than another part, that means that its molecules are moving faster. However, the faster molecules will collide with the slower ones, and thus pass on part of their energy. After a sufficient time, the energy will have spread evenly over all molecules. This means that the temperature has become uniform, and the entropy has reached its maximum.
In principle, you could describe the movement of each particle by classical mechanics. If you would then combine the trajectories for all the separate particles, you should be able to predict the behavior of the object as a whole. Therefore, the laws of thermodynamics, which describe the flow of heat in the object, should be derivable from the laws of mechanics. In practice, the particles are so small, and there are so many of them, that you would never be able to precisely measure all their initial states. Even if you had all that information, the number of particles is so astronomically large that even with the most powerful computer you would never be able to calculate the combined trajectory. Therefore, the laws of mechanics are useless.

The physicist Ludwig Boltzmann had the brilliant idea to use statistics to describe the behavior of particles. Instead of looking at the precise state of a specific particle, he would consider the “average” particle, and calculate the probability that such a particle would be in a particular state. This made the calculations much easier, while still providing reliable approximations. Thus was born the field of statistical mechanics. Instead of defining entropy in terms of temperature and heat, Boltzmann was now able to explain entropy in terms of probabilities. The reason that systems tend to go from low entropy to high entropy is simply because high entropy states are much more probable outcomes than low entropy states.
Fig. 1: a box containing air molecules, divided in two compartments with a hole in the dividing wall. The arrows represent the direction and speed of movement of the molecules. Since there are many more molecules in the left compartment, there will also be many more molecules crossing the divide from left to right.

Let me explain this idea with a traditional example: a closed box with two compartments (Fig. 1). The left compartment contains air, the right one is empty. Suppose that we make a hole in the wall separating the two compartments. An intuitive description of what happens is that air will be “sucked” out of the full compartment into the vacuum of the empty one, until both compartments contain equal amounts of air. This effect is easy to understand when you look at the individual air molecules in the two compartments. The molecules are continuously flying in all directions, colliding with each other and with the walls of the compartment. A molecule bumping against the central wall separating the full compartment from the empty one will be reflected, and move back towards the middle of the full one. However, if the molecule would land on the hole, rather than on the separating wall, it would not be reflected but continue its journey rightwards into the empty compartment. Since many molecules are thus continuously flying from the left to the right, the number of molecules in the right compartment will quickly increase. These molecules too will collide and be scattered in all directions. Some of them will move to the left, into the hole, and thus go back to the left compartment. However, as long as there are less molecules in the right compartment, there will also be less that move through the hole from the right, and thus rejoin their erstwhile companions. As long as the concentration of molecules on the left is larger, there will be more movement from left to right than from right to left. Only when the concentration of molecules in the two compartments is uniform will there be an equal flow through the hole in both directions. This configuration where the distribution is homogeneous is the one with maximum entropy.

Now, it is in principle possible that by some coincidence more molecules would move to the left than to the right. This creates a difference in concentration between the two boxes, and therefore a decrease in entropy. However, since the amount of molecules in either compartment is so astronomically large, a very large number of them must move “countercurrent” to produce a noticeable difference. The probability that such a vast number of molecules would all move together in the same direction is vanishingly small. It is so small, in fact, that we can assume that it will never happen. Because of the law of large numbers, the larger the number of coincidences needed for an effect to occur, the smaller the probability that the effect will occur. And with numbers as large as the
number of molecules in a gas, the resulting probability is as close to zero as you can practically get. Therefore, although entropy could decrease spontaneously, that is so improbable that the non-decrease of entropy has gotten the status of a law of thermodynamics. While we cannot predict the movement of the microscopic particles, the evolution towards homogeneity on the macroscopic scale is perfectly predictable.

The reasoning we made about molecules diffusing to evenly fill a box can be easily generalized to the homogeneous diffusion of heat. Just replace molecules moving to another region by molecules transmitting their energy to another region. In all cases the mechanism underlying the growth of entropy is the same: if particles move in different directions with the same probability, they will evenly spread over all directions, and erase any differences in local density.

This formulation of the second law highlights the paradox underlying statistical mechanics. The equal probability of movement follows from classical mechanics, where all directions are considered equivalent. A particle might move backwards or forwards: there is no preference. Therefore, the average number of particles moving in a particular direction will depend only on the number of particles present, not on the direction. On average, more particles will move out of a high density region than out of a low density one. Therefore, high density regions will become relatively less dense, and the differences in density will disappear. On the microscopic level of the particle, all directions of change have the same probability. On the macroscopic level of the gas, however, one direction of change—towards greater homogeneity—has a much larger probability.

This preferred direction creates an “arrow of time”. While classical mechanics does not really distinguish between past and future, statistical mechanics does. Time always flows in the direction of entropy increase, of greater homogeneity. This is progress of a kind, pushing evolution forward without ever allowing it to go back. But entropy increase is not what we usually call improvement. It is rather associated with the dissipation of energy, with wear and tear, with things running down and getting disorganized. Thus, thermodynamics has introduced irreversible change into the scientific world view. Yet, it still fails to explain the development of new organization which we normally associate with evolution.

**The Theory of Relativity**

The theory of relativity proposes a completely different departure from classical mechanics. While relativity theory has the same static, deterministic view of change as
classical mechanics, its interpretation of space and time is wholly different. Newtonian theory sees space and time as an absolute, unchanging theatre in which objects move. Thus, it assumes that all moving objects have a definite speed with respect to that background, which is considered to be at rest. The anomaly that triggered the development of the new theory was the famous Michelson-Morley experiment. The experiment tried to establish the absolute speed of the Earth when it moves through space. Since the Earth revolves around the sun, in some part of the year it will move towards a certain star, while six months later, when it has rotated 180 degrees, it will move away from it. Therefore, the light coming from that star should be moving faster relative to the Earth in the first case, and slower in the second case. By measuring the difference in speed, physicists hoped to determine the absolute speed of the Earth. To their great surprise, there was no difference. The speed of light seemed to be always the same, whatever the speed with which you would move relative to it.

In 1905, the 26 year old physicist Albert Einstein wrote a famous paper in which he proposed a solution to the problem. (in the same year, he published two more classic papers in which he solved two other outstanding problems in physics. A single inspired year was enough for him to revolutionize the whole of physics.) Einstein’s solution, which he called the theory of relativity, is based on two principles. First, all movement is relative. There is no absolute, static background against which movement can be measured. What seems at rest for one observer, will be moving for another observer. For example, for a person waiting in the station, a passing train seems to be moving. For a passenger sitting in the train, the train is at rest and the station seems to be moving. According to Einstein, both views are correct. The station-centered view is not in any way “better” or more correct than the train-centered view. Although the station is at rest with respect to the crust of the Earth, the Earth itself moves with respect to the sun, the sun moves with respect to the stars, and so on.

Although this relativity principle contradicted both common sense and Newton’s view of absolute space, it did not contradict the laws of classical mechanics. Indeed, these laws specify the effect of acceleration, that is, change in the speed of movement. Without acceleration, the laws only tell you that an object will remain in the same state of motion, whether that is rest, slow movement, or fast movement. Therefore, the introduction of the relativity principle did not require any real change in classical mechanics. It rather clarified the fact that the laws of nature are independent of the speed of the observer.

To explain the paradoxical results of the Michelson-Morley experiment, however, Einstein needed a second principle: the absoluteness of the speed of light. He postulated that the speed of light is always the same, for any observer, whatever his or her speed.
Both principles were in fact motivated by the same requirement. Einstein wanted to assure that the laws of nature would be the same for every observer. The laws of mechanics were already independent of speed. However, the laws of electromagnetism were not. The reason is that they are based on the speed of light, that is, the speed of electromagnetic radiation. If electromagnetism should look the same for every observer, then the speed of light should look the same as well. However, this brings us to the strange situation where all speeds are relative except the one of light. Somehow, the theory should explain why light has such a special place in the range of velocities.

Einstein proposed that the speed of light is in fact the limit speed, the highest speed any physical entity can travel. When a movement is accelerated so that its speed approaches the speed of light, strange things start to happen. It becomes more and more difficult to further increase the speed. The object is compressed in the direction of movement, its mass increases, and any clock traveling with it starts ticking more slowly. More and more energy is needed to accelerate further. In order to reach the speed of light itself, an infinite energy is needed. This means that massive objects can never travel with the limit speed. Only massless signals, such as electromagnetic radiation can travel with that speed. But they can only travel with the speed of light. You cannot slow down a light wave! The strange deformation of space and time taking place when an object travels at a speed close to the speed of light is often presented as the most spectacular aspect of relativity theory. However, it just derives from the need to have a smooth transition from the low speed domain, where speeds are relative, to the high speed domain, where there is an absolute limit.

The most important effect of Einstein’s two principles is that time and space can no longer be separated. For Newton, space is like an invisible, motionless substance, that is always there, independently of time or movement. The relativity of movement destroys that image, since two observers moving with respect to each could never agree about what motionless means. However, it seems that these two observers could still agree about time. It would suffice that when they meet they synchronize their clocks, and then, whenever it is 8 o’clock for the one, it will also be 8 o’clock for the other. But how do they know that the clocks remain synchronized? Suppose that both observers witness an event taking place far away—for example an explosion. Depending on their position and speed, the light from that event will need a certain time to reach them. For the one, the duration will in general be different from the duration experienced by the other. So the one may conclude that the event took place at 8.23 while the other would believe it took place at 8.34. Who is right? Einstein would say that both are right, but each within their own “frame of reference”. 

- 111 -
Suppose now that they would observe two explosions in two different places. For the one, the two explosions may seem simultaneous, since the light signals arrive at the same moment. However, for the other observer, explosion A seems to happen well before explosion B. A third observer might even claim that B in fact happened before A! Again, it is impossible to say who is right. Simultaneity of separate events is relative. The only way to establish absolutely that an event A occurred before another event B, is when a signal emitted from A reaches B. That signal tells an observer in B that A has already taken place. However, this means that the distance and the duration separating A from B should be such that a signal can cross that distance for the given duration without going faster than the speed of light. For all events separated by distances and durations which do not fulfill that requirement, it is impossible to say which event came before the other. All the events that can be reached from A by such a signal constitute the future “light cone” of A. (see Fig. 2).

![Figure 2: the light cone formed by an event A. The event B is in the future light cone of A, B' is outside the light cone and B'' is in the past light cone](image)

All the events that can similarly influence A by such a signal form its past light cone. The remaining events are situated outside the light cone, which means that they cannot communicate with A. They are neither in the future nor in the past of A. Yet, you cannot say that they are simultaneous with A either. In relativity theory, the three independent
dimensions of space and the one dimension of time are replaced by a four-dimensional “space-time” continuum, where you cannot determine the absolute time of any event. The only thing you can establish absolutely is whether an event is inside, on, or outside the light cone of another event, that is whether it can interact with it via a signal slower than light, as fast as light, or no signal at all. Particles or objects can only move inside the light cone, or on the light cone if they are massless. No movement, information transfer or cause-effect connection can occur outside the light cone, that is, with a speed faster than light. In fact, physicists tend to equate causal connections with connections inside or on the light cone. Chains of such connections, representing the movement of particles or signals, are called “world lines”. They are the four-dimensional equivalent of trajectories in classical mechanics.

Let us summarize the difference between classical mechanics and relativity theory. Static objects and their trajectories have been replaced by world lines connecting events. For two events happening in different places, you can no longer distinguish whether they are simultaneous or not. Therefore, you cannot attribute an absolute time to the events. Movement, speed, time, space have all become relative, dependent on the speed and position of the observer. The only distinctions that are absolute are the ones distinguishing connections inside the light cone from connections outside or on the light cone.

Ten years after Einstein proposed this theory of “special relativity”, he expanded it to the more universal theory of “general relativity”. General relativity extends the relativity of motion to the relativity of acceleration. Einstein noted that the effects of acceleration are indistinguishable from the effects of gravitation. This led to a new theory of gravitation which is more accurate than the one of Newton. However, it is of an extreme mathematical complexity. This is due to the insight that gravitation can be seen as a deformation of the geometry of space and time. Such “deformed”, non-Euclidean geometries are very difficult to grasp. They are even more difficult to use for calculations. Therefore, I will not go into any further details about general relativity, but just note that it keeps the same general outlook as special relativity, with its events, light cones and world lines.

Let me just mention one of the most spectacular applications of general relativity: black holes. The stronger a gravitational field, the higher the speed you need to escape from it (this is called the “escape velocity”). This is illustrated by the high speed a rocket needs to leave the Earth, and the comparatively lower speed it would need to leave the Moon. The more matter you compress together, the stronger the gravitational attraction it produces. At a certain point, the attraction becomes so large that the escape velocity
would be higher than the speed of light. Since nothing can go faster than light, this means that nothing could ever escape from such a dense gravitational field. Although objects that come near to the field would be attracted by it, and tend to fall into it, they would never be able to get out. Such a field is like a hole in space-time. Since not even light can come out of it, the hole is black—the deepest black you could imagine.

For a long time, black holes remained purely theoretical concepts, deriving from the mathematics of general relativity. Nowadays, however, astronomers have started to accumulate evidence that black holes exist in our galaxy. Although you cannot see a black hole directly, you can see its effect on nearby matter. On the one hand, the gravitational attraction of a black hole deflects the trajectories of neighboring stars. On the other hand, the matter being sucked into a black hole undergoes such a strong acceleration that it starts to emit radiation. Astronomers can see this radiation, at least for as long as the matter hasn’t crossed the black hole’s boundary—the “event horizon”.

The most important thing we need to remember about black holes is that they provide another example of one-way movement, of changes that cannot be undone. Thus, general relativity confirms the existence of irreversible movement that was first proposed in thermodynamics.

The Uncertainty Principle

Einstein’s theory was the first to attract the attention to the role of the observer. Different observers see different things when looking at the same phenomenon. However, in the theory of relativity you can still determine what one observer sees on the basis of the observations of another one. If you know the relative speed, position and orientation of the two observers, then you can use a mathematical transformation to convert the observations of the one into the observations of the other. For example, if an observer in the station sees a train moving to the north with 50 miles per hour, then you know that a passenger in the train will see the station moving to the south with a speed of 50 miles per hour. This is no longer the case in our next non-classical theory: quantum mechanics. Different observers looking at the same quantum phenomenon will not only see different things, they will in general not even be able to determine what the other one saw.

The name of “quantum mechanics” refers to the “quantum”, the amount of energy exchanged during interactions between atoms or particles. In order to explain the distribution of energy radiating from a hot object (a so-called “black body”), the physicist Max Planck had to assume that such electromagnetic radiation is exchanged in separate bits, which he called “quanta”. According to the classical theory, however,
Electromagnetism is carried by continuous waves, not by separate units. Einstein explained this phenomenon in another revolutionary paper written in 1905. He suggested that light consists of photons, particle-like “pieces” of electromagnetism.

A few years later, the opposite phenomenon was observed: electrons, which everybody thought to be particles, appeared to undergo interference, which is typical of waves. This double phenomenon, where waves behaved like particles and particles behaved like waves, was called the wave-particle duality. The statement that something is at the same time a particle, hence discontinuous, and a wave, hence continuous, is a logical contradiction. This confronted the physicists with a paradox. It required them to look at things in a radically different way.

The Danish physicist Niels Bohr, the first to propose a model of the atom, was also the first to propose an explanation for this conundrum. This explanation, which is known as the “Copenhagen interpretation”, is still the most widely accepted, although it has been joined by dozens of alternative theories. According to Bohr, we cannot know physical reality as it is, independently of ourselves. We can only make theories or representations of how we interact with physical systems. These representations must be formulated in the language of classical mechanics, because that is the only language which is intuitively clear for everybody. In the case of microscopic systems, such as atoms and elementary particles, however, there is no complete classical representation. There are only partial representations, such as the wave view and the particle view. These representations each capture a particular aspect of the system, but ignore the other aspects. Bohr called such representations complementary. This means that they are mutually exclusive, but jointly necessary for a complete description of the phenomenon. In other words, although you cannot represent a quantum system as a wave and as a particle at the same time, you somehow need to use both representations if you want to understand all the possible behaviors of that system.

To show why there must be complementarity, Bohr analysed the process of observation. He showed that the observation apparatus needed for measuring a particle-like property of a system (like position) cannot be used at the same time as the apparatus needed for measuring a wave-like property (like momentum). Hence, it is impossible to measure both properties at once. On the other hand, every measurement perturbs the system, since it requires the exchange of at least one “quantum” between apparatus and system. The effect of that perturbation is unpredictable. Therefore, the two measurements cannot be done one after the other either, since they will distort each other’s results. Therefore, you cannot determine both position and momentum of a quantum system.
The German physicist Werner Heisenberg expressed this conclusion as a mathematical law. His famous “uncertainty principle” shows that a high precision in the measurement of a property like position always goes together with a low precision in the measurement of a complementary property like momentum. The more you know about the one, the less you know about the other. If the one is perfectly certain, the other must be wholly uncertain. This means that quantum mechanics no longer allows you to make perfect predictions, like classical mechanics. If you know the state of a quantum system, you may be able to predict the result of certain observations (for example its speed), but not of others (for example its position). In general, you can only determine the probability of a certain result, not whether it will actually occur or not.

For physicists trained in the Newtonian world view, this was a very unsatisfactory situation. Although he helped start up the approach, Einstein himself was deeply unhappy with quantum mechanics. He believed that the world is intrinsically deterministic. “God does not play dice” is how he expressed this conviction. Therefore, he concluded, if quantum mechanics cannot predict what will happen in the world, the theory must be incomplete. If we would know the information lacking in quantum mechanics, then we would be able to make perfect predictions. Nowadays, such missing information is called a “hidden variable”, a property of the system which determines its future behavior, but which for some reason we cannot observe. Einstein’s objections were the start of a famous debate between him and Niels Bohr. Year after year, Einstein tried to prove that quantum mechanics is incomplete, but each time Bohr managed to find a flaw in his reasoning. At one point, when Einstein came up with a particularly sophisticated argument to show how two complementary properties could be measured together, Bohr even needed to use Einstein’s own theory of general relativity to prove him wrong.

Only one of Einstein’s arguments, the so-called “Einstein-Podolsky-Rosen paradox”, has survived until today. In it, Einstein and his collaborators propose a situation in which an observation of a particle would immediately change the state of another particle far away—at least if you believe quantum mechanics. According to relativity theory, no influence can move faster than the speed of light. Therefore such an immediate effect seems impossible. Einstein concluded that there isn’t such a change of state, and that the second particle must already have had the property which, according to quantum mechanics, should only appear after the observation. Therefore, quantum mechanics is incomplete.

This was not really a proof that quantum mechanics was wrong, only that it does not fit in with a certain interpretation of relativity theory. The only way to establish who was right, Einstein or Bohr, was to perform an experiment. However, in 1935, when the
argument was proposed, no instrument was precise enough to do the necessary measurements. It is only in 1982, long after Einstein and Bohr had died, that physicists were able to perform a reliable experiment. The results were indisputable: the predictions of quantum mechanics are correct, those of the hidden variable theories are wrong. A spate of similar experiments since then, testing various paradoxical predictions of quantum mechanics, have all turned up the same result. Of all theories known to humanity, quantum mechanics is probably the one whose predictions have been confirmed most accurately until now.

These results have profound implications for the physicists’ world view. On the one hand, the two particles in the Einstein-Podolsky-Rosen situation, however far apart, behave like a single entity. Quantum effects don’t seem to care about distances. A quantum wave can spread out over the whole of space. Any observation of part of the wave will immediately affect all other parts of the wave, even if they are lightyears apart. This is called “non-locality”: the wave and its interactions are in general not localized in a particular point or region of space. However, quantum non-locality does not contradict relativity theory, as Einstein believed. The reason is that the non-local effects cannot be used to transfer signals. Although “something” seems to be transmitted from one region to another, that something is neither matter nor information. Therefore it is not limited by the speed of light.

Another implication is that we cannot explain the unpredictability of a quantum system by assuming that we simply lack information about its “hidden variables”. However, this leaves open a host of options as to what causes quantum indeterminacy. There are about as many interpretations as there are theorists who have thought about the issue, from the most far-fetched to the most down-to-earth. One way to distinguish them is by the amount of attention they pay to the observer. In the quantum theory, the indeterminacy appears solely when an observation is carried out; otherwise the system evolves in a perfectly predictable way. However, in how far does it make sense to speak about a predictable evolution, if that evolution cannot be observed? Some theorists have gone so far as to suggest that the mysterious observation process and its strange effects on quantum systems only occurs when an observer becomes conscious of a system. Therefore, quantum mechanics would really be a theory of consciousness.

At the opposite end of the spectrum, some people have proposed that indeterminacy is part of the fabric of the universe itself, independently of the presence or absence of human observers. According to quantum mechanics, you cannot predict when an unstable system, such as a radioactive atom, will disintegrate. But such disintegrations take place all the time, whether you observe them or not. One
interpretation, the “many worlds hypothesis”, proposes that each time such an unpredictable event occurs, in fact all possible outcomes are realized, but in different universes. It is as if each quantum event forces the universe to split up into a number of parallel universes, each realizing one of the possible outcomes of the event. Although this theory may seem pure science fiction, it does not contradict any known facts, since it assumes that we can any way only follow one branch out of the billions of parallel universes which are created every moment. The other universes simply remain unobservable.

My Brussels colleague and friend Dirk Aerts has proposed a more down-to-earth interpretation which revives the hidden variable idea, but now involving the observer. He suggests that the non-classical probabilities of quantum mechanics could be explained by assuming that we lack information, not about the system being observed, but about the observation process. Such a theory might be generalized to situations outside of physics. Social scientists are well-acquainted with the “observer effect”, the influence of the researcher on the person or group she investigates. The answers people give to a questionnaire often depend on how the questions are formulated and who asks the questions. For example, people are more likely to agree with a friendly and attractive investigator. This observer effect is difficult to determine. By generalizing the calculation of probabilities in quantum mechanics, Aerts and his collaborators hope to develop an “interactive statistics”, which would measure in how far the results are influenced by the way the observation was done.

Let us conclude this section by summarizing in what way quantum mechanics departs from the classical world view. Most obviously, quantum mechanics does away with the view of the passive observer, who just registers what is going on without interfering. Moreover, it shows that no theory or model can ever hope to capture the whole of reality. Quantum mechanics also gives up the idea of determinism or predictability. More fundamentally, it violates the principle of distinction conservation, since the same quantum system can produce different observation results, while different systems can produce the same results. Finally, it suggests that at the lowest level space does not really separate things, since objects that are arbitrarily far apart can still act together as a single entity.

**Chaos**

The indeterminacy of quantum mechanics may seem to have little relevance for everyday life. After all, quantum effects are only detectable for microscopic particles. They are
negligible for large, macroscopic objects (the so-called “classical limit”). Yet, in the last few decades, physicists have become aware that even “classical” systems can behave in an intrinsically unpredictable manner. Although such a system may be perfectly deterministic in principle, its behavior is completely unpredictable in practice. This phenomenon was called deterministic chaos.

To explain its origin, we must go back to the concept of linearity. We noted in our discussion of classical mechanics that classical systems are in general not linear, but that they are often treated as if they are. Linearity means basically that effects are proportional to causes. If you hit a ball twice as hard, it will fly away twice as quickly. Another way of expressing this is summation: the total effect is the sum of the effects of the individual causes. For example, if you are pushing a car that ran out of fuel, and want it to move twice as fast, you might either push twice as hard, or find someone else to help you push. The effect would be the same. In the example of the car, the system is not perfectly linear: when you push twice as hard, the car will not move exactly twice as fast, but only approximately. You would not make a big mistake, though, if you would assume that the effect is proportional to your effort. Many practical situations are like that: they are not exactly linear, but you can approximate them quite well with a linear function. Linear equations are solved easily, but non-linear ones are in general very hard or impossible to solve. Therefore, until the beginning of this century most non-linear problems in classical mechanics were approximated by linear ones. However, cases started to accumulate where linear functions were clearly not good approximations.

One of the most famous is the three-body problem. Newton’s theory of gravitation provides a simple solution to the problem of two mutually attracting bodies, for example the sun and one of its planets. However, as soon as a third body comes into play, for example another planet, the problem becomes mathematically unsolvable. In practice, astronomers work with approximations, where the attraction to the most important body, in this case the sun, is taken as the basis, while the effect of a third body is brought in as a perturbation. As we have seen, predictions based on this approximation are in practice very reliable. The reason this works is because the gravitation exerted by the planets is tiny compared to the gravitation exerted by the sun. However, nobody can prove that they are absolutely reliable. It is very well possible that the solar system is unstable, and that the gravitational attractions between the different planets may lead one of the planets to suddenly escape into outer space. Let us hope that this won’t be the fate of the Earth!

We cannot predict whether such catastrophic effects will occur because they depend on undetectable changes in the initial conditions. In the two body problem, if one
of the conditions is changed a little, the effect will not be very different. For example, if the moon would be brought a little closer to the Earth, its trajectory would remain basically the same. This is no longer true in the three-body problem. A tiny change in one of the variables, for example the speed of the planet Venus, might result in a totally different outcome, for example the planet Mars crashing into the sun. This is called “sensitive dependence on initial conditions”. The effects are extremely sensitive to changes in the conditions that cause them. This is the essence of non-linearity: effects are no longer proportional to causes. Small causes may have large effects. In a way, “sensitive dependence” is nothing more than the rediscovery by scientists of the old wisdom which is captured by the phrase “for want of a horseshoe the kingdom was lost”. Processes which are very sensitive to small fluctuations are called chaotic. This is because their trajectories are in general very irregular, so that they give the impression of being random, even though they are driven by deterministic forces.

The meteorologist Lorentz has invented yet another expression, the “butterfly effect”. While studying the equations that determine the weather, he noticed that their outcomes are strongly dependent on the initial conditions. The weather is a chaotic system. The tiniest fluctuations in air pressure in one part of the globe may have the most spectacular effects in another part. Thus, a butterfly flapping its wings somewhere in Chicago may cause a tornado in Tokyo. This explains why scientists find it so difficult to predict the weather. To predict future situations, they need to know the present situation in its finest details. But obviously they will never be able to know all the details: they cannot monitor every butterfly flapping its wings! The fewer details they know, the less accurate their long term predictions. That is why reliable weather predictions seldom extend more than a few days in the future.

Such chaotic processes basically work as amplifiers: they turn small causes into large effects. That means that small, unobservable fluctuations will affect the outcome of the process. Although the process is deterministic in principle, equal causes having equal effects, it is unpredictable in practice. Indeed, causes that seem equal to the best of our knowledge can still have unobservable differences and therefore lead to very different effects. In particular, such a chaotic process can amplify the effects of quantum fluctuations, and thus smuggle the indeterminacy of quantum mechanics into the macroscopic world.

The interaction between quantum mechanics and chaos has not been extensively studied yet. However, the combination of chaos and thermodynamics has turned out to be very fruitful. The main innovation of the thermodynamic view is the idea that different causes lead to the same effect, the state with maximum entropy. This equilibrium state to
which all other states converge can be seen as an attractor: it is as if it attracts different possible states of the system, so that all trajectories come together in the same point. One example of an attractor is, of course, a black hole. For a less sensational example, consider a pendulum. If you let it move freely, it will slow down after a few oscillations and come to rest in its lowest position, whatever the position it started from. The initial differences in the starting positions (causes) are gradually reduced until they completely disappear.

![Fig. 3: a point attractor: the arrows represent trajectories starting from different points but all converging in the same equilibrium state](image)

Let us now make the system non-linear. This means that we will add a force which amplifies differences. The combination of the two influences, one dampening differences, one amplifying them, can produce the most complicated behaviors. The simplest one is just a regular oscillation, which never stops, like the pendulum of a clock which is powered by a spring. This pattern of movement is again an attractor, since different initial positions of the pendulum will all converge to the same periodic trajectory. But such an attractor no longer consists of a single point, but of a closed, one-dimensional line of points. This is called a limit cycle, since it represents the closed trajectory in which the system will settle in the limit. (It is ironic that the traditional mechanical clock, whose regular movement became a paradigm for classical mechanics, in fact achieves this regularity through the balance of two opposing, non-classical forces.)
When the non-linearity of the system becomes more important, more complicated attractors may appear. The final trajectory in which the system settles may have a very irregular shape, without any apparent periodicity. Yet, this eventual trajectory is still an attractor, because neighbouring trajectories are “sucked” into it, losing their freedom to get out again. In general, an attractor is a region in the space of possible states which the system can enter but not leave. In that sense, an attractor is like a “black hole” in space, constantly sucking in material, but never letting anything go. That region can have different dimensions, of which the zero-dimensional point attractor, and the one-dimensional limit cycle are just the simplest cases. Attractors with non-integer, “fractal” dimensions are called strange attractors. Trajectories moving inside a strange attractor are totally chaotic.

Another characteristic of highly non-linear systems is that they have in general several attractors. A simple thermodynamical system, on the other hand, has only one attractor, the maximum entropy state. When there is more than one attractor, the main question is in which of those attractors the system will end up. Imagine that each attractor corresponds to a lake or sea, and that the trajectories leading into an attractor correspond to the rivers and streams flowing into those lakes. Depending on where it falls, rainwater will follow either one river or another, ending up in either one lake or another. The complete area drained by a river is called its basin. Similarly, each attractor has a basin, which is the surrounding region in state space such that all trajectories starting in that region end up in the attractor. The basins belonging to different attractors are separated by a narrow boundary. This “watershed” often has a very irregular shape. For initial
positions close to the divide, it is very difficult to determine to which attractor they lead. Small fluctuations can push the system either into the one or into the other basin, and therefore either into the one or into the other attractor. (see Figure 5). The appearance of such a border separating two attractors is called a bifurcation. Close to the border, the system behaves chaotically, inside the basin it moves predictably towards the attractor.

Fig. 5: three attractors with some of the trajectories leading into them. Their respective basins are separated by a dotted line.

When a dynamical system is put under stress, for example by increasing the energy or flow through it, the number of attractors tends to increase. One way to understand this is by noting that more energy allows more amplification of small differences, and therefore more varied types of behavior. Suppose you start with one attractor. When you increase the stress, you get a bifurcation: the attractor splits in two. The system now has two stable patterns of behavior. For example, imagine that you let water run through a tap. When the water runs very slowly, it comes out in regular, periodically falling drops. When you open the tap a little bit more, it may happen that there are two patterns of running: either big drops succeeding each other quickly, or a
thin, continuous stream. Sometimes you have the one pattern, sometimes a slight fluctuation in the pressure makes the water switch to the other pattern. When the stress is increased further, more bifurcations take place, and the attractors split up further. First, you have 4 possible regimes, then 8, then 16, then 32, and so on, ever more quickly. At a certain point, the number of attractors becomes infinite and the system is erratically jumping from the one to the other all the time. This is true chaos. The behavior of the system has become totally unpredictable. Coming back to the water tap, this is what happens when the tap is opened fully and the water is running out turbulently, in one big, irregular waterfall, with droplets spraying in all directions.

In summary, the study of non-linear systems has made the principle of distinction conservation, and therefore determinism, meaningless in practice. In spite of the underlying classical mechanisms, chaotic systems are unpredictable. Similar causes can lead to very dissimilar effects, as when two neighbouring states end up in two far-away attractors; and dissimilar causes can lead to the same effect, as when two far-away regions in an attractor basin converge into the same attractor.
Wholes, Parts and Systems

The Wholeness of Nature

The new developments in physics we have sketched all concern the trajectories of physical systems. These trajectories appear much more uncertain and observer-dependent than classical mechanics assumed. However, physics does not really question the analytic method used to arrive at these trajectories. It still assumes that physical systems can be reduced to elementary, independent pieces of matter. Yet, the more deeply we investigate elementary particles, the less they look like the “billiard balls” of classical mechanics. The wave view of quantum mechanics already shows how a particle can be “smeared out” to cover the whole of space, and how the properties of seemingly independent particles can be correlated.

The inclusion of relativity theory into quantum mechanics led to the development of quantum field theories. Instead of viewing particles as indestructible fragments of matter, field theories see them as temporary materializations of an encompassing energy field. The field represents a fundamental force, such as electromagnetism or gravitation, spreading out over the whole of space. Particles can be created or destroyed, or even exist in a kind of “limbo” state, as virtual particles, which are neither there nor not there. The only thing that is permanent is the field, which holds all these potential appearances together. In that view, all particles are in fact part of a larger whole, which pervades the universe. This theory may seem more mystical than scientific, but its significance lies in the accuracy of its predictions. Insofar that it makes concrete predictions, they turn out to be very precise. However, because of the complexity of the underlying mathematical model, predictions can only be made for the most trivial cases. As soon as more than a few interacting particles are considered, calculations become simply impossible.

These ideas have reminded many people of the holistic world views predating Newton, such as the philosophies of Buddhism and Taoism. Both visions seem to imply that everything in the universe is connected to everything else. That would mean that the analytic method, which is based on dividing systems into their components, leaves out the essential part: the interconnectedness. However, there does not seem to be a clear alternative to separating out parts and aspects when trying to understand a phenomenon.
Models that try to encompass everything, like the field theories, simply become unusably complex. Remember that a person can only keep about seven items at once in short term memory. Therefore, you simply cannot think about problems that involve more than seven aspects simultaneously. Mathematical theories or computer simulations can work with a larger number of entities, but they too quickly break down in their capacity to handle large numbers.

The only possible alternative might be the intuitive awareness of the whole, the “oceanic feeling” experienced by mystics after deep meditation. But such a holistic awareness is not the same as scientific understanding. The mystic may feel a deep connectedness or understanding of Nature as a whole, but does not therefore become capable to predict specific events, or to solve concrete problems. In conclusion, in spite of its shortcomings, the analytic method must remain part of scientific modelling. However, its limitations can be overcome by a complementary approach, which we will call the synthetic method. Where the analytic method breaks a phenomenon apart, to better understand its components, the synthetic method puts the pieces back together, to better understand their function in the whole. The synthetic method is based on the assumption that the whole is more than the sum of its parts. Putting different components together creates something new. An automobile is something that can drive. A collection of wheels, tyres, motor, steering wheel, etc. cannot drive. A person is a living entity. Take the person apart into the different organs and tissues, and you get a collection of dead components. Such properties of the whole which get lost when you separate the parts are called “emergent”.

Originally, such emergent properties seemed quite mysterious, as they could not be explained in the mechanistic world view. Different thinkers felt that something was missing in that view. The vitalists thought that life could only be explained by assuming that organisms were animated by some mysterious “life force”. Others thought that emergent properties had to belong to a new type of entity, which they called a “holon” or “integron”. Still others revived Descartes’ old dualism, and said that you need to add mind or consciousness to the theories of matter.

However, these approaches remain stuck in the old view that you can explain phenomena by reducing them to a collection of particles, forces or substances. Just adding new entities to that collection does not solve the problem. Whether you call the particles “electron” or “holon”, or call the forces “gravitational”, “spiritual” or “vital”, does not really make much difference for the kind of theory that you are building. The difference between living and dead matter cannot be reduced to the mere presence or absence of a “life force”. What the whole has more than its parts is not some mysterious
fluid or force, but organization: the complex pattern of interconnections, which makes the parts, fulfills specific functions for the whole. The first approach to study this organization was systems theory.

The Systems Approach

The systems approach integrates the analytic and the synthetic method, encompassing both holism and reductionism. It was first proposed under the name of “General System Theory” by the biologist Ludwig von Bertalanffy. von Bertalanffy noted that all systems studied by physicists are closed: they do not interact with the outside world. When a physicist makes a model of the solar system, of an atom, or of a pendulum, he or she assumes that all masses, particles, forces that affect the system are included in the model. It is as if the rest of the universe does not exist. This makes it possible to calculate future states with perfect accuracy, since all necessary information is known.

However, as a biologist von Bertalanffy knew that such an assumption is simply impossible for most practical phenomena. Separate a living organism from its surroundings and it will die shortly because of lack of oxygen, water and food. Organisms are open systems: they cannot survive without continuously exchanging matter and energy with their environment. The peculiarity of open systems is that they interact with other systems outside of themselves. This interaction has two components: input, that what enters the system from the outside, and output, that what leaves the system for the environment. In order to speak about the inside and the outside of a system, we need to be able to distinguish between the system itself and its environment. System and environment are in general separated by a boundary. For example, for living systems the skin plays the role of the boundary. The output of a system is in general a direct or indirect result from the input. What comes out, needs to have gotten in first. However, the output is in general quite different from the input: the system is not just a passive tube, but an active processor. For example, the food, drink and oxygen we take in, leave our body as urine, excrements and carbon dioxide. The transformation of input into output by the system is usually called throughput. This has given us all the basic components of a system as it is understood in systems theory (see Fig. )
When we look more closely at the environment of a system, we see that it too consists of systems interacting with their environments. For example, the environment of a person is full of other persons. If we now consider a collection of such systems which interact with each other, that collection could again be seen as a system. For example, a group of interacting people may form a family, a firm, or a city. The mutual interactions of the component systems in a way “glue” these components together into a whole. If these parts did not interact, the whole would not be more than the sum of its components. But because they interact, something more is added. With respect to the whole the parts are seen as subsystems. With respect to the parts, the whole is seen as a supersystem.

If we look at the supersystem as a whole, we don’t need to be aware of all its parts. We can again just look at its total input and total output without worrying which part of the input goes to which subsystem. For example, if we consider a city, we can measure the total amount of fuel consumed in that city (input), and the total amount of pollution generated (output), without knowing which person was responsible for which part of the pollution. This point of view considers the system as a “black box”, something that takes in input, and produces output, without us being able to see what happens in between. (in contrast, if we can see the system’s internal processes, we might call it a “white box”). Although the black box view may not be completely satisfying, in many cases this is the best we can get. For example, for many processes in the body we simply
do not know how they happen. Doctors may observe that if they give a patient a particular medicine (input), the patient will react in a certain way (output), e.g. by producing more urine. However, in most cases they have little idea about the particular mechanisms that lead from the cause to the effect. Obviously, the medicine triggers a complex chain of interconnected reactions, involving different organs and parts of the body, but the only thing that can be clearly established is the final result.

Fig. a system as a “white box”, containing a collection of interacting subsystems, and as a “black box”, without observable components.

The black box view is not restricted to situations where we don’t know what happens inside the system. In many cases, we can easily see what happens in the system, yet we prefer to ignore these internal details. For example, when we model a city as a pollution producing system, it does not matter which particular chimney produced a particular plume of smoke. It is sufficient to know the total amount of fuel that enters the city to estimate the total amount of carbon dioxide and other gases produced. The “black box” view of the city will be much simpler and easier to use for the calculation of overall
pollution levels than the more detailed “white box” view, where we trace the movement of every fuel tank to every particular building in the city.

These two complementary views, “black” and “white”, of the same system illustrate a general principle: systems are structured hierarchically. They consist of different levels. At the higher level, you get a more abstract, encompassing view of the whole, without attention to the details of the components or parts. At the lower level, you see a multitude of interacting parts but without understanding how they are organized to form a whole. According to the analytic approach, that low level view is all you need. If you know the precise state of all the organs and cells in the body, you should be able to understand how that body functions. Classical medicine is based on this reductionist view. Different alternative approaches to medicine have argued that such a view misses out the most important thing: the body is a whole. The state of your mind affects the state of your stomach which in turn affects the state of your mind. These interactions are not simple, linear cause-and-effect relations, but complex networks of interdependencies, which can only be understood by their common purpose: maintaining the organism in good health. This “common purpose” functions at the level of the whole. It is meaningless at the level of an individual organ or cell.

One way to understand this is the idea of “downward causation”. This concept was proposed by the great methodologist Donald T. Campbell, with whom I had the privilege to collaborate in the last few years before his death. According to reductionism, the laws governing the parts determine or cause the behavior of the whole. This is “upward causation”: from the lowest level to the higher ones. Campbell suggested that in emergent systems, the laws governing the whole also constrain or “cause” the behavior of the parts. This he called “downward causation”.

Let us illustrate this by looking at the structure of our jaws. A reductionist would say that jaws are chunks of a bony substance, consisting of atoms of calcium and a few other elements, which move relative to each other according to the law of levers. However, this does not explain the particular shape of the jaws, the fact that they contain teeth, the space left between the jaws for the passage of food and air, and so on. However, when you look at the jaws as part of the system “body”, it becomes clear that their main function is to facilitate the chewing and swallowing of food. The fact that particular calcium atoms fill particular positions and not others cannot be explained by the laws of physics and chemistry. A jaw that does not leave space for the tongue satisfies the laws of mechanics just as well as a jaw that does. Only the higher level laws of biology can explain why a jaw must have a shape so that food can be easily swallowed. If you compare human jaws with the jaws of other species, you will moreover notice that
the cavities between the bones are such that the mouth has an optimal shape for producing the sounds needed for speaking. This can only be explained by laws at a yet higher level, the one of culture or sociology. This level sees a human being as part of a larger social system, that can only function if its components are capable of spoken communication. Thus, the structure of our jaws is governed at least by the laws of physics governing atoms, the laws of chemistry governing molecules, the laws of biology governing organisms, and the laws of sociology governing social systems. None of these laws can be reduced to the laws of the level below.

The same reasoning can be made for most of the things that surround us. Although the behavior of a transistor in a computer chip is governed by the laws of quantum mechanics, the particular arrangement of the transistors in the chip can only be understood through the principles of computer science. The structure of the DNA molecule, which codes our genetic information, is determined by the laws of chemistry. Yet, the coding rules themselves, specifying which DNA “triplet” stands for which amino acid, don’t derive from chemistry. They constitute a law of biology. Each level in the hierarchy of systems and subsystems has its own laws, which cannot be derived from the laws of the lower level. Each law specifies a particular type of organization at its level, which “downwardly” determines the arrangement of the subsystems or components at the level below. When we say that the whole is more than the sum of its parts, the “more” refers to the higher level laws, which make the parts function in a way that does not follow from the lower level laws.

Although each level in a hierarchy has its own laws, these laws are often similar. The same type of organization can be found in systems belonging to different levels. For example, all open systems necessarily have a boundary, an input, an output and a throughput function. The cells in our body need food and energy in the same way that the body as a whole needs food and energy, even though the cells receive these substances in a different form. The material is different, but the function is the same: to allow the cell or organism to grow, repair itself, and react to adverse effects. Similar functions can be seen at the level of society, which also needs an input of “food” (including ores, raw materials, agricultural produce) and energy, which it uses for self-repair and growth. Closed systems at different levels have many features in common as well. The binding forces which hold together the planets in the solar system, the atoms in a molecule, or the electrons in an atom, although physically different, have a very similar function. The embeddedness of systems in supersystem holds for all types of systems: societies consist of people which consist of organs, which consist of cells, which consist of organelles,
which consist of macromolecules, which consist of molecules, which consist of atoms, which consist of nucleons, which consist of quarks.

Thus we find similar structures and functions for different systems, independent of the particular domain in which the system exists. General Systems Theory is based on the assumption that there are universal principles of organization, which hold for all systems, be they physical, chemical, biological, mental or social. The mechanistic world view seeks universality by reducing everything to its material constituents. The systemic world view, on the contrary, seeks universality by ignoring the concrete material out of which systems are made, so that their abstract organization comes into focus.

Mind out of Matter

Measuring Information

We saw that the separation between mind and matter, between the observer and the observed, forms one of the basic unsolved problems of the classical world view. By showing how you can scientifically understand a phenomenon without reducing it to its material components, the systems approach made a first step towards a solution of the problem. Further steps were taken by a number of disciplines closely associated with systems theory: the theory of information, cybernetics, artificial intelligence and cognitive science. We will discuss these developments roughly in their historical order.

What distinguishes a scientific from a more intuitive approach is often the procedure of “measurement”: scientific modelling requires an accurate registration of the properties of the phenomenon being studied. The problem with the mind, however, is that you cannot directly observe it. The only way we get in touch with minds is by personal experience, but that is hardly an objective method of registration. Behaviorist psychology has tried to solve the problem by only studying observable behavior. Using the black box approach, behaviorists have studied the mind as a pattern of correlations between stimuli (input) and responses (output), without wondering what happens in between. However, it is clear that the mind is much too complex to be understood as a mere correlation. But if you want to model its internal organization, you will still need some objective features
that you can measure, so that you can test the theory. The first “mental” feature to be expressed in such a quantitative manner was information.

Nowadays everybody is used to the idea that information is measured in bits. However, in the beginning of this century, the only things that could be measured where purely physical features, such as length, duration, temperature or weight. There was no way to measure, for example, the efficiency with which a communication channel, such as a telephone line, transmits information. The communications engineer Claude Shannon set out to do just that. To succeed he had to understand how information is communicated. He noted that when you want to tell something to someone else, you will first convert or code that idea to a physical form, such as a spoken sound or a scribble on a page, which is then decoded by the receiver. When you use an indirect communication medium, such as the telephone, further conversions are needed, from sound to electrical impulses, and perhaps from electrical impulses to radio waves, and from there back to sound. In each conversion, the physical form of the message is altered. Yet, to achieve communication, the content must remain the same. However, the different communication channels are not perfect, and part of the message may get lost because of conversion or transmission errors on the way. For example, your correspondent may not be able to read all of your handwriting, or the person you phone to may not understand everything you say because of hissing and crackling on the line. Communication engineers use the generic term “noise” for such uncontrollable errors or distortions. Given these intrinsic limitations of the channel, how can you make sure that a maximum of information is transmitted with a minimum of errors? This is the problem that Shannon tried to solve. There are many ways to code a message. The best coding will be the one which puts the most information in a message, but is least prone to be distorted by noise.

But how do you measure the amount of information? Let us consider the most simple type of communication. Suppose someone asks you a question, such as “Is it raining?”, and you answer either “yes” or “no”. In that case, you have communicated one bit of information. Before you answered the question, your interlocutor did not know what the weather was: she was uncertain which of two alternatives, rain or no rain, applied. Your answer informed her, it made her uncertainty disappear. Therefore, information can be seen as the opposite of uncertainty. In this case the uncertainty was limited to two possibilities: “yes, it rains” or “no, it does not rain”. However, your partner might also have wanted to know whether it is warm or cold. Your answer to that question will give her another bit of information. In fact, she now knows which of four possible situations is actual: warm and dry, cold and dry, warm and raining, or cold and raining. The two answers provide two bits of information, which together select one out of four
alternatives. Every question or set of questions will raise a number of possible answers. The larger the number of possibilities, the larger the uncertainty. The larger the uncertainty, the more information is needed to resolve it. Therefore, the information content of an answer or a message increases with the number of alternatives that it eliminates.

However, the number of alternatives itself is not a good measure of information. You would rather have it increase with the number of answers needed. In the example of the weather, the 4 possibilities are resolved by 2 “yes-no” answers, that is, 2 bit of information. Adding another yes-no question would increase the number of alternatives to 8, but the amount of information to 3 bits. Mathematically, the number of bits corresponds to the logarithm (to the base 2) of the number of alternatives. The logarithm is a function which transforms products (determining the number of alternatives) to sums (determining the number of bits). This makes the measure of information additive: the information contained in a sequence of answers or messages is the sum of the information contained in each separate message.

Until now, our examples considered binary questions, having exactly two possible answers, because this is the simplest possible case. However, the measure of information as the logarithm of the number of alternatives applies to any possible number, whether it can be expressed as a power of two or not. For example, you might classify the state of the weather as hot, warm, mild, cool or cold. In that case, a selection out of 5 possible states would give you \( \log 5 \) bits of information, that is, about 2.3 bit.

To get Shannon’s full formula for information, we need to make one more generalization. We have assumed until now that the different alternatives have the same probability. However, that is in general not true. For example, if you live in the desert, and somebody asks you whether it rains, it is very likely that you will answer “no”. A negative answer would merely confirm expectations, and therefore provide practically no information. The unexpected “yes” answer would be informative, but would occur so rarely that on average the information would still be low. In fact, normally the answer would be so uninformative that people would not bother to ask you the question, unless they had a special reason to assume it might be raining, that is, unless the probability of rain would be much higher. The more expected a message is, the less information it contains. Therefore, information can also be seen as a measure of surprise.

Shannon’s measure of information takes into account expectations by using the probability of each alternative rather than the number of alternatives. If all possibilities have the same probability—as we implicitly assumed in the first example—that probability is the inverse of the number of possibilities. Therefore, the information
measure we presented initially, namely the logarithm of the number of possibilities, is a special case of Shannon’s more general formula, for the situation where all probabilities are equal.

We don’t need to go into more details about the formula here. What is important, though, is that Shannon’s measure for information turns out to be the exact negative of Boltzmann’s measure for entropy. In other words, information seems equivalent to the lack of entropy. Remember that entropy was introduced in thermodynamics to explain why energy dissipates into heat. Entropy measures the degree of diffusion or homogeneity of a collection of particles. A homogeneous distribution of particles means basically that a particle might be anywhere, with equal probability. With such a maximum entropy distribution, you simply have no clue where a particle could be situated. In other words, maximum entropy is the same as maximum uncertainty, and therefore minimum information.

The second law of thermodynamics, which says that entropy in a closed system tends to increase, therefore also means that the information you have about such a system tends to decrease. The longer you let the system evolve on its own, the less you will know about where its components are situated. When the system has reached its equilibrium state, where entropy is maximal, you will have lost all information you had about its configuration. Remember the box with the two compartments. Before you open a hole in the internal wall, you know that all the particles are in the left compartment. After you have let the particles diffuse, you no longer know where any given particle is: it might be in the left one or in the right one, with equal probability. In the first situation, you had information about the compartment where the particles are. In the second case, you have lost that information. Thus, Shannon’s measure of information establishes a fundamental relationship between energy as studied in thermodynamics, and knowledge as studied in the information sciences.

The information measure has many other applications. The most obvious one is that it can be used to calculate the amount of information that is stored, for example, in a message, a book, a database or a computer file. Everybody who uses a computer knows how many bits, bytes (1 byte = 8 bits), kilobytes (1000 bytes), or megabytes (1 million bytes) can be kept in the computer’s memory or hard disk. More subtly, it helps you to find more efficient ways of storage. For example, the English language is not very efficient: it needs long strings of symbols (letters) to express relatively little information. The reason is that the different letters and combinations of letters have very different probabilities. For example the letters “e” and “n” are much more frequent than “x” or “q”, and the combinations “ou” or “th” are more frequent than “fs” or “hw”. If you
look at all the possible combinations of three or more letters, you will notice that most of them (for example, “hxa”, “btk”) are never even used. If you could recode the English alphabet so that the new symbols and their combinations would be more or less evenly used, you would be able to express the same phrases with a much shorter string of symbols. If you would represent these symbols as binary digits (1 or 0), you would need less of these bits than if you would start from the normal alphabet. Thus, you could store the same information in a much smaller memory space on your computer. This is the principle underlying the many schemes for memory compression. They are especially useful if you have to transmit a lot of information (say, all the works of Shakespeare) over a low capacity channel (say, a telephone line). The better the compression, the less time will be needed to transmit the message.

Compression comes at a price, though. A message that contains more characters than needed to express the information contained in it is said to be redundant. The redundancy is the difference between the maximum amount of information that could have been encoded in a message of that length and the actual amount it contains. For example, a message consisting of the same phrase repeated ten times is very redundant: it contains only one tenth of the information that it could have contained. Redundancy has one big advantage: it makes it easy to correct errors. If the ten repeats of the sentence would each contain a few transmission errors, you would probably be able to reconstruct the correct sentence by comparing the different versions. If the sentence was transmitted only once, you would have no way to determine which parts are correct. Still, if the sentence would be in English, you might be able to guess where the mistakes are. For example, if you get the string “tke inftrmation”, you would probably guess that what is meant is “the information”. This is because you know that the combination “tk” is very improbable, while “th” is much more probable. The redundancy of the English language makes it much easier to communicate over poor channels (such as noisy telephone lines). Using Shannon’s theory you should be able to find the ideal coding, which contains enough redundancy to allow for easy error-correction, depending on the quality of the line, yet manages to compress a maximum amount of information in a given transmission.

In spite of these many applications, Shannon’s theory of information has a fundamental shortcoming. Although it has a lot to say about the structure and the amount of information in a coded message, it is totally ignorant about the meaning of that message. Whether you transmit the works of Shakespeare, the New York telephone book, a series of political speeches, a random collection of meaningless words, or even the same letter repeated a billion times, if these messages contain the same number of
characters they will have the same information content, as measured by Shannon’s formula. The meaning is in the mind of the message’s sender or receiver, not in its coding. Although many people have tried to understand the meaning of a message in terms of what it stands for, the most important meaning is in what you can do with it. If you live in Tokyo and someone sends you the New York telephone book, that message will most likely be as useful to you as a volume filled with random characters. However, if you are looking for the phone number of a correspondent who lives in New York you will find the message most informative. In order to understand such pragmatic meaning, we need a theory about how information can help us to reach our goals: cybernetics.

**Cybernetics: the science of purpose**

In the mechanistic world view, there is no place for purpose or goal-directedness. All mechanical processes are determined by their cause, which lies in the past. A goal, on the other hand, is something that determines a process, yet lies in the future. To a Newtonian scientist, the idea that an as yet non-existent, future state could influence the present, seems wholly unscientific, not to say mystical.

The thesis that natural processes are determined by their future purpose is called teleology. It is closely associated with vitalism, the belief that life is animated by a vital force outside the material realm. Our mind is not an aimless mechanism; it is constantly planning ahead, solving problems, trying to achieve goals. How can we understand such goal-directedness without recourse to the doctrine of teleology? Systems theory made it possible to understand life as a form of organization, without postulating non-physical forces. Cybernetics similarly solved the problem of goal-directed action.

The word “cybernetics” was coined by the mathematician Norbert Wiener. It is derived from the Greek word “kubernetes”, which means “helmsman”, that is, the navigator responsible for steering a ship towards its chosen destination. The English verb “to govern” has the same root. The insight at the basis of cybernetics is that of circularity. If the effect of an action leads back to its cause, you have a closed loop or causal circle. For example, an increase in the number of births will lead to a larger population which will in turn lead to more people having children, and therefore an increase in the number of births. The increase in birth rate is cause and effect at the same time. Such a circular mechanism is called a feedback loop. The effect is fed back into the cause, so that the two become inseparable. Although this is not strictly true, we could say that both the past (cause) influences the future (effect) and the future influences the past.
In this example, however, the future effect which “influences” the past or present state is identical to that present: increasing births produce increasing births. To have goal-directedness we need a future state which is different from the present state, yet somehow “pulls” that state towards itself. We can achieve this through negative feedback. In positive feedback, an increase in some quantity (e.g. birth rate) causes more of the same. In negative feedback, an increase produces less of the same. For example, if the supply of some good, such as paper, increases, its price will go down. A lower price will in turn motivate the paper producers to lower their production and therefore the paper supply. Thus, increase in supply leads to decrease in supply (and the other way around). The net result is that supply will tend to fluctuate around an equilibrium value. Too strong changes will lead to an opposite effect, counteracting the deviation.

We need one more component to build a goal-directed system. The equilibrium value around which the paper supply fluctuates will itself fluctuate. It will be affected by a multitude of factors, such as consumer demand, speculations on the stock market, and the availability of wood pulp, chemicals and other resources. However, imagine now that some organization wishes to keep the equilibrium at a fixed value. For example, the government may want to maintain the value of the currency at a fixed rate, in spite of speculations and other disturbing factors. If the supply of currency becomes too large, the central bank will buy up currency, thus decreasing supply. If the supply is too small, the bank will sell currency, thus pushing up its rate. This is again a negative feedback loop, but this time with a fixed equilibrium or reference value, to which the supply will always return. This reference level functions as a goal. All actions of the bank are directed at reaching or maintaining the goal state. If they are succesful, they give the bank control over the currency rate.

This example may seem to already presume too much. After all, banking officials can be assumed to be intelligent, goal-directed individuals. However, the mechanism we sketched does not require any human intervention. The buy-sell order might as well have been given by a computer program that monitors the supply of currency. To show that goal-directed action requires only the most basic physical mechanisms, we will discuss the example of the thermostat. The thermostats used to keep the temperature constant in a tropical fish tank are extremely simple. They consist of a fixed plate and a moving strip. The strip consist of two metallic layers, which expand at a different rate when heated, thus deforming the strip. The strip is made such that it bends towards the plate when the temperature goes down, and away from the plate when the temperature goes up. If the temperature reaches a certain low level, the strip comes into contact with the plate, thus creating an electric current which activates a heating element inside the fish tank. The
heating increases the temperature. This makes the bimetallic strip bend away from the plate, and thus break the contact. This stops the heating, so that the aquarium starts to cool down, until it reaches the reference temperature at which the cycle starts again. Thus, the temperature of the fish tank is kept at a constant value, in spite of the fluctuations in outside temperature. The reference temperature, which is determined by the distance between the plate and the bending strip, functions as a goal, which the thermostat at all times tries to reach or maintain.

Keeping some variable like temperature or currency rate at a fixed value may seem like a rather static type of goal-directedness. But the negative feedback mechanism works equally well when you aim at a moving target. For example, a heat-seeking missile pursuing an enemy plane will constantly change its goal or reference point, depending on where the plane appears on its radar screen. But once the target’s position has been established, it will direct its own trajectory by constantly correcting or counteracting deviations from the target course. In fact, it were such military applications, like automatic anti-aircraft guns, which gave Wiener the inspiration to model goal-directedness through negative feedback. When he discussed these first ideas with his friend Arturo Rosenblueth, a medical doctor, he was delighted to hear that the same negative feedback loops underly much of the functioning of our organism. Our body temperature is maintained at a constant rate by a principle very similar to that of the thermostat. Thousands of other variables, such as blood pressure, sugar content of the blood, or amount of oxygen, are similarly kept at their reference levels by counteracting various perturbations. This capability of the organism to maintain a constant inner milieu in spite of ever changing circumstances has been called “homeostasis” by Rosenblueth’s teacher Walter Cannon.

In 1943, Rosenblueth and Wiener wrote down their discovery in a classic paper entitled “Behavior, Purpose, and Teleology”. In it, they showed how the apparently teleological behavior of purposive, goal-seeking systems can be explained by feedback. Although the circularity of feedback reconciles goal-directedness with causality, it does transgress the Newtonian principle of distinction conservation. Indeed, goal-directed systems exhibit “equifinality”: different initial states all lead to the same final goal state. Thus, the discovery of feedback was another nail in the coffin of the classical world view.

In 1948, Wiener expanded his ideas into a book with the carefully crafted title “Cybernetics, or communication and control in the animal and the machine”. Present-day cyberneticists no longer restrict themselves to the study of animals and machines, but include subjects such as people and organizations. Yet, communication and control are still the cornerstones of their theories. As we noted, the effect of goal-directed negative
feedback is control over the factors that affect the goal. When you suddenly get in a draft, your body starts to shiver, thus producing more heat which keeps your body temperature at the required level. When you are exposed to strong sun, on the other hand, your body starts to sweat, thus bringing the body temperature down. The net result is that these outside disturbances do not affect your body the way they would if there was no feedback mechanism. Your body temperature is kept within strict limits: it is under control. This control functions by counteracting or compensating perturbations before they have become too large. If you would start to sweat only after your body temperature had increased with 10 degrees, you would die from overheating. In order to counteract perturbations quickly and accurately, your brain needs to know precisely the state of your body. Therefore, your skin is covered with sensors for heat and cold which send signals to the brain as soon as there is a drop or increase in temperature. The brain reacts by sending signal that activate sweating or shivering reflexes. These signals constitute information, and their transmission to and from the brain constitutes communication. Therefore, goal-directedness, control and communication of information are intimately connected. Thus, Wiener’s cybernetics builds directly on Shannon’s information theory.

The examples of control systems we discussed are all quite simple, with a single goal and a single variable measuring deviations from that goal state. As the example of the body suggests, real control system can be immensely more complicated, with hundreds of intricately entwined feedback loops regulating thousands of variables to approach dozens of interacting goals and subgoals. More specifically, goals can be subordinated the one to the other. For example, the thermostat of your house may have the simple goal to maintain a certain temperature. But the more important goal is to keep the house comfortable for its inhabitants. This may mean that the thermostat would seek different temperatures depending on whether there are people inside the house or not. Sophisticated sensors may warn the thermostat when there is no one in, and thus trigger a lower reference temperature.

In complex control systems, such as organisms or organizations, goals are typically arranged in a hierarchy, where the higher order goals control the settings for the lower order goals. For example, your goal of avoiding thirst, that is, keeping up the necessary level of liquid in your body, may activate the lower order goal of drinking a glass of water. This will in turn activate the goal of reaching out and bringing the glass to your lips. At the lowest level, this entails the goal of keeping your hand steady around the reference position while moving up the glass. Although for most of us this happens so naturally that we are not even aware of it, people with Parkinson’s disease or cerebral palsy find this exceedingly difficult. As Rosenblueth and Wiener understood, such
diseases are characterized by a misfunctioning of the negative feedback loop which controls the position of the hand.

**Artificial Intelligence**

In spite of the complex and adaptive behavior exhibited by multilevel control systems, the cybernetic model still does not explain real intelligence. Traditional control assumes that deviations from a goal can be measured by variables such as temperature or position, and that the necessary counteractions follow directly from the size and direction of the deviation. For example, if your hand moves a little too far to the left, your brain will correct it automatically by ordering the muscles to tense toward the right. However, in many cases it is not at all obvious how to reach a goal from the situation you are in. For example, if your car breaks down, and your goal is to make it drive again, you would not know in which “direction” your goal lies, or how much “distance” you need to cover in order to get there. You’ll have to analyse the problem, using whatever knowledge you have about how cars work, try out different actions, and hope that you will eventually get the desired effect. Such problems demand knowledge and intelligence, rather than automatic, preprogrammed reactions. The field of artificial intelligence (AI) was the first to develop and apply concrete models of such problem-solving activity.

The goal of AI is to design computer programs that behave in an intelligent, human-like way. When the first computers were developed in the 1940’s, they were basically meant as calculating machines. To calculate the value of a mathematical function, say a square root, you need a list of rules which tell you what to do in what order. Such a “recipe” for calculating a function is called an algorithm. It is purely deterministic: everything you should do is determined beforehand; there is no choice or uncertainty. If you apply the rules correctly, you are guaranteed to get the correct result (though this may take a long time...). If all problems could be solved by applying algorithms, the world would be a very simple place indeed. Unfortunately, most problems require guessing or trial-and-error. Still, guesses are usually informed. When you try to guess what’s wrong with your car, you usually already have some ideas of likely causes, such as an empty fuel tank, or a wet spark plug. If you would give advice to someone whose car does not start, you would suggest a number of rules-of-thumb, such as “first check the level of fuel”. Unlike algorithms, such rules are not guaranteed to solve the problem. Moreover, it is not obvious which of the many rules should be tried out first. Such a system of rules which only helps you to make informed guesses, without
guarantee of success, is called a “heuristic”. The origin of AI is in the study of heuristics for problem solving.

To understand how you can teach a computer to solve problems, we will analyse the typical components of a problem. A problem is characterized first of all by the fact that your present situation (for example, a car that doesn’t start) is different from your goal (a moving car). In order to get from your present situation to your desired situation you will need to make a number of moves, which will in general bring you through a sequence of intermediate situations. For example, a first move could be to ascertain that the fuel tank is empty, the second move to find a container, the third to find a fuel station, the fourth to fill the container in the fuel station, the fifth to bring the container back to the car, and the sixth to fill the car’s fuel tank. If you are lucky, as in our example with the fuel, that sequence will lead you more or less straightforwardly to your goal. However, you could make any number of false starts or detours. For example, after walking 3 miles along the highway with your container, you might find out that there is no fuel station in that direction. This will in general lead you to backtrack, that is, come back to one of the previously explored situations, and start exploring a different avenue. All possible situations you can reach from your initial problem situation constitute a “space” of situations, the problem space. Your goal state, where the problem is solved, constitutes one of the points in that space.

Perhaps the most general heuristic is called “hill-climbing”. Imagine that your goal is to reach the highest point in the problem space. In that case, the obvious method to reach the goal is to always choose that path which goes most steeply upwards, assuming that it will lead to the top of the hill. Of course, you can always define your problem space so that “height” corresponds to “closeness to the goal”. If you can in some way measure the distance to the goal, you can use that number to specify the “height” of all points in the problem space. Thus, every type of problem can be reduced to “hill-climbing”. The simplest method to do hill-climbing is called “generate-and-test”. First, you generate all the situations or points in the problem space that you can reach by applying one of the possible moves from your present situation. Then, you test how close each of those is to the goal, or, in other words, how “high” each one is. Finally, you move to the one that scores best, and start the procedure again. In that way, you will move up in the quickest, most direct way.

Although hill-climbing is a universal method which can be applied to any type of problem, it has intrinsic limitations. By always ascending the steepest slope, you risk ending up on top of a small hill, while the highest peak which you really want to reach remains far out of sight. Any top or peak, however high or low, is an end point for the
generate-and-test rule. Indeed, once you are on the top, all roads lead downwards, and there is no longer any rule that can tell you which one will lead to the really highest point. At best, you can try to choose the road that goes down most slowly, but that does not guarantee that it will lead to the real summit. The fact that there tend to be several peaks or hill-tops in a problem space is a general difficulty besetting problem-solving. It may look as if you are moving closer to your goal, but in fact you may be heading for a dead end, from which you will have to backtrack.

In the example of the fuel station, we were discussing moves in a real physical space. Problem spaces in general, however, are abstract, conceptual spaces. When we want to design an intelligent, problem-solving computer, we are not thinking about a robot that would actually go out and fetch fuel. We are rather thinking about a program that can reason abstractly, and solve the problem theoretically. In order to reason about a practical situation, you need to have a conceptual representation of that situation in your head. You need to know the possible moves and their effects, so that you can try them out in your mind in order to discover the one that will solve the problem most efficiently. This knowledge is used as a representation of the problem.

The conceptual problem space is like a map of the real problem space. For example, instead of walking 3 miles in order to find that there is no fuel station, it would be much more efficient if you could plan the shortest route to a fuel station by examining a map. AI researchers similarly assume that people reason about problems by using a conceptual map. To build an intelligent computer you then simply need to store a map of the problem domain in the computer’s memory, and provide it with a heuristic program for efficiently exploring that map.

The creation of such a map is called “knowledge representation”. The knowledge about the problem domain needs to be given a particular format so that it can be easily stored and used. For example, in real maps you need to decide which symbols you will use to represent towns, villages, churches, or hospitals, and which color and thickness of lines will be used to denote rivers, roads, railways and highways. Depending on the format, a map or representation can be easy or difficult to use. Imagine using a map in which horse trails are represented by thick blue lines, rivers by brown shading, and highways by thin dotted lines! More important even than the particular symbols and formats used, is the question which parts of the domain to include. Too much detail will clutter the map, and make orientation difficult. Too little detail may leave out the specific thing you are looking for. What is included moreover depends on the purpose of the user. If your purpose is to plan a journey, you will basically be interested in roads, railways and the names of cities and villages. But if you are a geologist, you may be more
interested in elevation above sea level, hydrological basins and composition of the soil. As a sociologist, on the other hand, you would rather be interested in the geographical distribution of population densities, average wealth, and type of employment. Including all these types of information on the same map would merely create a hopeless mess.

The same problems confront AI researchers. Hundreds of different formats and structures have been proposed to represent knowledge. Some of the more popular ones include predicate logic, production rules, frames and semantic networks. Each has its own advantages and disadvantages. None is obviously superior. They are all limited in the amount of detail they can include without making problem solving too complicated. The result is that AI has evolved from general methods for problem solving to specific programs covering a very narrow domain, such as trouble-shooting for a particular brand of cars, or diagnosing pulmonary diseases. These programs are called “expert systems”: they try to answer specialised questions the way a human expert would. Unlike the human, however, they are totally helpless if a problem falls outside their restricted domain, or is not covered in the rules that have been programmed into them.

An obvious way to tackle this rigid specialization is to let the AI program learn, that is, extend its own knowledge each time it encounters a problem that it cannot solve. Thus it would not need to rely solely on the knowledge programmed into it. But how can a system autonomously create new knowledge? We will now discuss the main issues in the generation of knowledge.

**Construction and Deconstruction**

When developing new knowledge, the classical view of knowledge as a reflection of outside reality is surprisingly unhelpful. As we discussed earlier, the Newtonian world view sees knowledge as a snapshot, taken through the camera of our senses and stored on the film of our memory. This would suggest that an AI system would simply collect data about its environment, either through its sensors or through the input it receives from the people using it, and store those different bits and pieces into its memory until a coherent picture emerges. Unfortunately, if you just more or less randomly collect pieces of data, it is extremely unlikely that something like a coherent map would appear.

The “maps” or “representations” used in AI are cleverly organized systems of knowledge which try to explain a maximum of situations through a minimum of concepts and rules. The relative success of such cognitive systems inspired psychologists to study the organization of human thought and memory in a similar way. This produced the new discipline of cognitive psychology. The older, behaviorist psychology saw the mind as a
black box, with stimuli as input, responses as output, and nothing that could be scientifically analysed in between. Cognitive psychologists tried to see the mind as a white box with information entering as input, being processed by a complex cognitive system made out of many interacting components, and exiting in the form of behavior. This, however, did not yet explain where that cognitive system came from.

The problem of how cognitive systems develop in our mind is the same as the problem of scientific discovery. Indeed, a scientific theory is a complex system of knowledge, based on observations. In the beginning of the 20th century, people thought that theories are generated through induction: from a number of recurrent observations, the researchers induce a general rule. For example, if you see the sun coming up in the morning for many days in a row, you induce the rule that the sun always comes up in the morning. However, countless attempts to find a logic of induction only resulted in failure. Philosophers of science, such as Karl Popper, concluded that theories cannot be derived from observations. At best, observations can either confirm or refute theories, but they can never prove them. Proposing a theory is a creative act. It cannot be reduced to mere accumulation of observations.

One way to see this is by noting that a good theory should be able to predict events that have not happened as yet. If it could only describe what we already observed, it would be pretty useless. However, the number of events that could happen is infinite, while the number that have actually been observed is finite. This means that in principle you could generate an infinite number of theories, which all explain the same observed data, but which make different predictions for the events that have not yet happened. New observations may help you to eliminate all those theories that made incorrect predictions. But the number of remaining theories will still be infinite, however large the number of observations. Therefore, we need more than observations to find a good theory.

In psychology, researchers such as Jean Piaget have come to a similar conclusion. The often unconscious and intuitive theories that we use when we reason about the world are not derived from the things we have observed. Theories or cognitive systems are constructed. They are individual creations, which assemble a jumble of facts, beliefs and intuitions into a coherent structure. This system necessarily goes beyond the things we have observed. It is based on assumptions and rules which can never be proven by observation. Every such construction is individual. Different people will interpret the same facts in different ways, because their cognitive systems have been constructed differently. Therefore, all knowledge is intrinsically subjective. Knowledge is not a mirror which objectively reflects the outside world. It is a personal construction which
tries to make sense out of the chaos of impressions with which we are constantly bombarded.

Even our memory of personal experiences is a construction. This is illustrated most dramatically by the “false memory syndrome”. By smartly designed experiments, psychologists have shown how you can make people believe that they have witnessed something which never actually happened. In one of such experiments, subjects were given a list of events that supposedly happened when they were children, and asked which ones they remember. Most of the stories were collected from their parents or siblings and checked for their accuracy. However, some of the stories were invented by the experimenters. These stories describe events that might plausibly have happened to the subjects, such as getting lost in a shopping mall, or accidentally spilling a bowl of drinks on someone’s clothes during a party. The parents were asked to confirm that these events never actually occurred. Yet, when these stories were presented to the subjects as if they were true, a sizable percentage of the people claimed to effectively remember the events. When prodded by the experimenters to remember more accurately, they came up with concrete details, such as the people present or the weather that day, which were not part of the original story. The more often they were interviewed about these childhood anecdotes, the more their “memories” became intense and detailed.

This false memory syndrome may explain some dramatic court cases, in which therapists were convicted for making their patients falsely believe that they were abused as children. If a psychotherapist is convinced that his or her client’s emotional problems stem from some repressed childhood experiences, he or she can by suggestive questioning make the client construct a number of detailed “remembrances”, which are practically indistinguishable from reality. Without independent evidence, it is virtually impossible to distinguish true experiences from falsely remembered ones. In one case, after a series of therapy sessions a young woman became convinced of having been raped several times by her parents, and forced to undergo two abortions using a coat hanger. Her father, a clergyman, had to resign from his post when the allegations became public. However, a medical examination revealed that she not only never had been pregnant, but in fact still was a virgin. The patient sued the therapist for implanting false memories and received a $1 million settlement.

Other psychological experiments demonstrate the “creativity” of memory in a less dramatic way. In one classic experiment, subjects watch a movie about an incident, for example a traffic accident or a man being mugged in a street. Afterwards, the psychologists ask them detailed questions about what they saw. Depending on either their own biases or the suggestions of the experimenter, most subjects tend to recall distorted
versions of what happened. For example, after watching the mugging incident, in which a person was attacked by a white man and afterwards helped by a black man, several viewers reported the attacker to be black.

The tendency to remember more than was actually there can be demonstrated even more simply. Suppose that you listen to a long list of closely related words, such as “hot”, “cool”, “winter”, “freezing”, “chilly”, “warm”, “shiver”, “ice”, etc. When afterwards you try to remember as many words as possible, it is likely that you will recall a word like “cold”, which was not in the list. In fact, if the list is well-chosen so that all words are strongly associated with “cold”, you are more likely to remember “cold” than any of the words that were actually there.

These experiments give us a hint of the mechanisms responsible for creative memory. Just like the scientist who tries to develop a theory that fills in the many gaps between the observed data, our memory tries to fit observations into a harmonious pattern or “story”. The facts that are lacking because they were forgotten or simply never observed tend to be replaced by plausible conjectures, that make the whole into a coherent system. If you don’t remember how the attacker looked, but you remember seeing a negro, and you believe that black people tend to be muggers, then you will unconsciously infer that the black man was in fact the attacker. Similarly, if you associate “cold” with “chilly”, “warm”, “ice” and all the other words you heard, your memory will automatically tend to assume that “cold” must be part of the list, even if you never actually heard it.

Memory of events is in fact a type of knowledge where creativity is minimized, since we are trained to be as accurate and objective as possible. The observation that even here “facts” are unconsciously manufactured shows that the idea of knowledge as the mechanical registration of perceptions is inadequate. It seems even more ludicrous when considering more abstract types of knowledge, such as ideas, beliefs, theories or philosophies. Such knowledge is constructed in order to solve our particular problems. Therefore it must be as simple and coherent as possible, and help us to reach our own goals. Outside reality does not provide the material for this construction. It merely gives us some hints of how things might be, and leaves the rest to our imagination.

The role of imagination in making up plausible constructions can be illustrated by a party game, which I have played a few times. The game is great fun, but you need an unsuspecting victim who has never played the game before. A volunteer is sent away to another room while the rest of the group discuss their plan of action. When the volunteer comes back, the others tell him that they have invented a “dream” for him. This is a story in which the volunteer takes part, and which supposedly takes into account his
experiences and personality. Being a “dream”, however, it does not need to be based on any real or even physically possible events. The game then consists in the volunteer trying to guess the story that the others have made up for him. He can ask questions, but the others are only allowed to answer “yes” or “no”. For example, he might ask: “Does it happen here?” or “Do I take part as my real self?” The volunteer is prodded to let his imagination take over, and not try to be too analytic, because otherwise he would never be able to guess something as fanciful as a dream. In the beginning, his questions will be very vague and general, but as he gets more answers, his picture of the dream sequence will get more concrete, and so the questions become more specific. For example, “Does the car arrive after the lady has left?” If he cannot think of any more questions to ask, the game is over, and the volunteer is asked to tell the dream as he has reconstructed it from the answers he got. Depending on the inventiveness of the volunteer, the resulting story can be very involved and detailed indeed. However, when he then asks whether his reconstruction of the original story is accurate, the group tells him that there never was a story!

The only thing the others discussed was a rule for answering his questions. The basic rule is very simple: every question that ends with a certain letter, say “s”, is answered with “yes”; all others are answered with “no”. The only exception to the rule is that answers should be consistent. Thus, if the volunteer asks the same question with different words, the answers should remain the same, even if the last letters of the questions are different. Since all groups members give the same answers (though they sometimes may get slightly confused because they heard or understood the question differently), and since the different answers confirm each other, the victim gets the impression that they effectively have agreed on a coherent story. Each further answer gives him one more bit of information, which he uses to further develop the story he has in mind. Yet, the end result, however coherent, detailed and involved, is purely a creation of his imagination, with no counterpart in any reality except his own mind.

Although this example is rather extreme, it helps us to understand the process of mental construction of ideas, beliefs and theories. Observation of the outside world does provide us with information, in the form of “bits”, which helps us to decide whether we should accept or reject a certain hypothesis. But we must find the meaning of these bits ourselves, by fitting them into our own conceptual designs. There is no way in which we can determine whether these constructions “correspond to” or “reflect” outside reality. This philosophy of knowledge is called “constructivism”.

Although many natural scientists haven’t grasped its importance yet, this insight has spread through most of the sciences which study mind and society. In artificial
intelligence, it has made scientists understand that really intelligent systems must be *autonomous*: they must be able to develop their own knowledge instead of getting it preprogrammed from the system designer. In cybernetics, it has led to the “second order” approach, which studies how observers construct models of systems, rather than studying the systems directly. In education, it tells researchers that children will only learn new ideas if they actively construct their own mental models, instead of trying to absorb the model of the teacher. Therefore, knowledge cannot be taught by *instruction*, but only by *construction*. In psychotherapy, it has led to the understanding that people have their own personal “constructs” through which they make sense of the world around them. Emotional problems are often due to inadequate constructs, which make people interpret their experiences in a needlessly negative way. They can be treated by helping patients to create more positive constructs. In the humanities, the awareness that all ideas, arguments and ideologies are merely subjective constructions has led to the method of “deconstruction”. This critical approach takes different conceptual systems apart in order to show how they could have been assembled differently—thus exposing their biases and hidden assumptions. In sociology and philosophy of science, constructivism has produced the concept of “the social construction of reality”, which emphasizes the role society plays in our shared mental constructions. Since outside reality remains forever out of reach, the only reality that can be said to exist in some objective way is the one about which everybody agrees. If different physicists agree that the pattern of dots they observe on their photographic emulsion constitutes the trace of a particle, say an electron, then we might say that the electron is “real” by social consensus. But that does not imply that it is “real” as an objective, physical entity.

Even the physicists themselves have been pushed towards a more constructivist philosophy. Indeed, quantum mechanics has taught them that we will never be able to know more about the electron than the patterns it makes on different detectors. If we try to grasp what these patterns mean, whether they “really” were made by a particle or by a wave, we come to various paradoxes. These can be avoided by giving up the idea of an object, either particle or wave, that exists independently of us. Instead, we should see the electron as merely a mathematical abstraction, represented by its “wave function”, which we construct in order to find coherence between the different observations.

Constructivism rejects one of the cornerstones of the Newtonian world view: that observation can give us an objective picture of the world “out there”. It moreover tells us that the mind is an active, creative agent, rather than a passive channel through which information is registered. However, it does not yet tell us how such creative processes really work. This will be the subject of the next chapter.
From Being to Becoming

Self-Organizing Systems

In the mechanistic world view there is neither creation nor destruction. Everything that exists now has already existed in some different arrangement in the past, and will continue to exist so in the future. All change can be reduced to movement of particles in space. In such a philosophy, there is no place for either the creativity of mind, or the creativity of matter.

Yet, during the last decades physicists have discovered that matter is creative indeed. According to quantum field theory, particles can be created, destroyed, or recombined to form new particles. According to general relativity theory, the geometry of space and time can change. Both theories together provide us with a glimpse of how the universe itself may have come into being. The astronomer Hubble observed around 1930 that the different stars and galaxies are moving away from each other. This means that the universe is expanding, and that it is larger now than it was in the past. If we extrapolate this movement backwards, we come to a point in the distant past when the universe was infinitely small. This point in time where the expansion started is known as the “Big Bang”. It can be seen as the effective creation of the universe: the origin of space, time and matter.

The precise mechanisms through which the different structures of matter and space came into being are as yet only dimly understood. Indeed, they require the application of theories to the description of extraordinary circumstances, which are totally different from anything we can observe in nature or in the laboratory. Physicists will never be able to reproduce the extreme temperature and density of matter during the Big Bang. They can only extrapolate their theories to situations in which they have never been tested, and hope that the theories still hold. Even then, all these theories in some way assume the existence of space and time; therefore, they become meaningless if you try to extrapolate them to a point where there wasn’t any space, time or matter yet. Thus,
they can only give us hints of what might have happened, but not reconstruct the whole process of creation.

The creativity of matter is not limited to the beginning of Time, though. The emergence of novel structures is easy to study in the laboratory. Perhaps the simplest example is crystallization, the appearance of a beautifully symmetric pattern of dense matter in a solution of randomly moving molecules. You can easily observe it yourself by dissolving plenty of salt in a glass of water, and then letting the water slowly evaporate. If you hang a thread in the solution, small crystals of salt will start appearing on the thread. The formation of a snow crystal is a similar phenomenon, with vapor molecules condensing out of the air to form a complex shape with a six-fold symmetry. A different example is the Bénard phenomenon, the appearance of a honeycomb pattern of cells in a liquid heated from below. Unlike crystals, the cells are not rigid structures of molecules, but patterns of perpetual movement, where hot liquid comes up inside the cells, while liquid cooled by the air sinks down in between the cells. More complicated examples are certain chemical reactions, where it suffices to add a number of ingredients to a solution in order to see dazzling spirals of pulsating color appear.

What all these examples have in common is self-organization: the spontaneous appearance of structure or pattern, without an external agent or force imposing it. It is as if the system of molecules arranges itself into a more ordered pattern. We have already argued that such a phenomenon contradicts the mechanistic world view. But it doesn’t fit into our intuitive picture of the world either. If a system, such as a flower, a building or a watch, shows signs of organization we tend to assume that someone or something must have arranged the components in that particular order. If we cannot find a person responsible for the design, we are tempted to attribute it to God, or some unknown intelligent force. This intuition is confirmed by the second law of thermodynamics, which says that in a system left to itself order can only diminish, not increase. So the first step to explain self-organization must be to reconcile it with thermodynamics.

In the example of crystallization, the solution is relatively simple. The randomly moving molecules which become fixed within the crystalline structure pass on the energy of their movement to the liquid in which they were dissolved. Thus, the decrease in entropy (disorder) of the crystal is compensated by an increase in the entropy of the liquid. The entropy of the whole, liquid and crystal together, effectively increases.

In the case where the self-organizing system is in continuous movement, the solution is less obvious. The Belgian scientist Ilya Prigogine received a Nobel Prize for his investigation of that problem. Together with his colleagues of the Brussels School of thermodynamics, he has been studying what he called “dissipative structures”. These are
patterns such as the Bénard cells or the Brusselator reaction, which exhibit dynamic self-organization. Such structures are necessarily open systems: energy and/or matter are flowing through them. The system itself is continuously generating entropy, but this entropy is exported out of the system. The system actively “dissipates” entropy to the environment. Thus, it manages to increase its own organization, at the expense of the organization of the environment. The system circumvents the second law of thermodynamics, which says that entropy must increase, simply by getting rid of excess entropy, by pushing it outside its boundary. The most obvious example of such dissipative systems are living organisms. Plants and animals take in energy and matter in a low entropy form as light or food. They export it back in a high entropy form, as waste products. This allows them to reduce their own internal entropy, thus counteracting the degradation caused by the second law.

The export of entropy in a dissipative system clarifies why self-organization does not contradict the second law. However, it does not yet explain how or why self-organization takes place. Prigogine noted that such self-organization typically takes place in non-linear systems, which are far from their thermodynamical equilibrium state. But that is more an observation than an explanation of how self-organization precisely works. Not only thermodynamicists studied self-organization, though. The cyberneticians, who studied how systems can control themselves, were obviously also interested in how systems can organize themselves.

One of the founding fathers of cybernetics, the Englishman W. Ross Ashby, proposed what he called the principle of self-organization. Rather than looking at concrete physical or chemical systems, he considered the issue in the most general, most abstract terms. He noted that whatever happens, a complex system will always tend to evolve towards a state of equilibrium, or what we would now call an attractor. As we saw earlier, when a system reaches an attractor, it can no longer go back to the states outside of the attractor. Therefore, its freedom to move through the state space has diminished. This also means that the uncertainty about the state the system is in has diminished: we know that it can only be in one of the attractor states, all the other states have been excluded. Therefore, our information about the system’s state has increased. Since information is the negative of entropy, this means that the entropy of the system has decreased. This is not in contradiction with Prigogine’s observation, since Ashby did not say what would happen with the entropy outside the system. The essence of self-organization is that the internal entropy decreases. Whatever happens with the entropy outside is irrelevant for the system itself.
Another founding father of cybernetics, Heinz von Foerster, formulated a second principle of self-organization, which he called “order from noise”. He noted that paradoxically the more disordered the perturbations (“noise”) that affect the system, the more quickly it will self-organize (become ordered). The idea is very simply: the more a system is made to move through its state space, the more quickly it will end up in an attractor. If it would just stay in place, no attractor would be reached and no self-organization could take place. Prigogine formulated the analogous principle of “order through fluctuations”. As we discussed in the section on chaos, non-linear systems have in general several attractors. When a system resides in between two attractor basins, it will be in general a chance variation, called “fluctuation” in thermodynamics, which will push it either into the one or the other of the attractors. Without fluctuations, it might remain forever in the no man’s land in between. We will discuss these principles of self-organization in more detail later.

Since the 1950’s and 1960’s, when self-organizing systems were first studied by the thermodynamicists and cyberneticists, many further examples and applications of self-organization have been discovered. Prigogine generalized his observations to argue for a new scientific world view. Instead of the Newtonian reduction to a static framework (“Being”), he sees the universe as an irreversible “Becoming”, which endlessly generates novelty. The cyberneticians went on to apply self-organization to the mechanisms of mind. This gave them a basis for understanding how the mind constructs mental models without relying on outside instruction. A practical application that grew out of their investigations are the so-called “neural networks”. These are computer models of how the neurons in our brain interact. Neural networks are very different from the traditional AI systems. Unlike such problem solving systems, there is no centralized control in a neural network. All the “neurons” are connected directly or indirectly with each other, but none is in control. Yet, together they manage to make sense out of complex patterns of input. Not even the system’s programmers can tell you how the neural network comes to its conclusions. The only thing they know is that after a repeated “training” process, the network self-organizes into a system of attractors, which can then be used to efficiently classify different inputs.

The same phenomenon, a multitude of initially independent components, such as molecules or neurons, which end up working together in a coherent manner, turns up in the most diverse domains. The production of laser light is a key example of such collective behavior. Atoms or molecules that are “excited” by an input of energy, tend to emit the surplus energy in the form of photons (“light particles”). Normally, the different photons are emitted at random moments in random directions. The result is ordinary,
diffuse light. However, under particular circumstances the molecules can become synchronized, all emitting the same photons at the same time in the same direction. The result is an exceptionally coherent and focused beam of light. Such a laser beam can even be used to illuminate the Moon from here on Earth, without the light diffusing along the way. The physicist Hermann Haken, who analysed lasers and similar collective phenomena, was struck by the apparent cooperation or synergy between the components. Therefore, he proposed the new discipline of “synergetics” to study such phenomena.

Another example of spontaneous collective behavior comes from the animal world. Flocks of birds, shoals of fish, swarms of bees or herds of sheep all react in similar ways. When avoiding danger, or changing course, they move together in an incredibly synchronized manner. Sometimes, the swarm or shoal behaves as if it were a single giant animal. Yet, there is no “head fish” or “bird leader”, which coordinates the others and tells them how to move. Each animal in the swarm simply watches its neighbours to decide how it should move. Computer simulations have shown how you can reproduce the behavior of swarms by letting each individual follow a few simple rules, such as keeping a minimum distance from others, and following the average direction of the neighbours’ moves. Yet, out of these local interactions a global, coherent pattern emerges.

Swarms are but one of the many self-organizing systems that are now studied through computer simulation. This method is at the basis of the new domain of “complex adaptive systems”, which was pioneered by a number of researchers connected to the Santa Fe Institute in New Mexico. These complexity theorists study systems consisting of many interacting components, which undergo constant change. The behavior of such complex systems is typically unpredictable, yet exhibits various forms of adaptation or self-organization. A typical example is an ecosystem, consisting of organisms belonging to many different species, which compete or cooperate while interacting with their shared physical environment. Another example is the market, where different producers compete and exchange money and goods with consumers. Although the market is a highly chaotic, non-linear system, it usually reaches an equilibrium in which the many changing and conflicting demands of consumers are all satisfied. In fact, the collapse of the communist states has shown that the market is much more effective at organizing the economy than any centrally controlled system. It is as if a mysterious power ensures that goods are produced in the right amounts and distributed to the right places. Two centuries ago, Adam Smith, the father of modern economics, called this organizing principle “the invisible hand”. Nowadays, we can simply call it self-organization.
When analysing how complex systems adapt to changing circumstances, complexity theorists look at more than self-organization, though. The biologist Stuart Kauffman, who studied the development of organisms and of ecosystems, distinguishes two separate mechanisms: self-organization and selection. Let us now look at the second part of this couple.

**Natural Selection**

Long before self-organization was discovered, there already existed a theory which explained the appearance of organization and adaptation: Darwin’s theory of natural selection. Around the beginning of the 19th century, different scientists were wondering where the incredible diversity of living species could have come from. The recent discoveries of fossils seemed to imply that the different animals and plants had not always been the way they are now. This led some scientists to conclude that there must have been a process of evolution, leading to the development of new types of organisms.

The French biologist Lamarck was the first to propose a mechanism to explain such evolution. Plants and animals can obviously adapt to different circumstances. A tree growing on an exposed hill-top will grow less tall than a tree of the same species growing in a forest. It seemed logical that these adaptations would be passed on to the organism’s offspring. If each generation would adapt further in the same direction, the changes would accumulate and eventually lead to a difference so large that the latest generation would no longer belong to the same species as its ancestor. For example, it is likely that the forefather of today’s giraffe was an antelope-like animal with a short neck, that was feeding on tree leaves. According to Lamarck, by stretching their neck in order to reach higher branches, subsequent generations of giraffes would have gotten longer and longer necks, eventually culminating in the animal we know now.

Although Lamarck’s theory seems intuitively acceptable, it has two major shortcomings. First, it became clear that acquired adaptations, such as stretched necks, are not passed on to offspring. Thus, adaptations cannot accumulate. Therefore, they do not lead to long term changes. Second, the theory did not really explain how individual organisms adapt. The example of the giraffe stretching its neck to reach higher branches seems rather obvious. But what about more complex adaptations, such as the giraffe’s intricate pattern of brown spots? You can hardly imagine a giraffe straining to turn its fur brown in particular patches, but not in others. Very few features of plants and animals can be explained by a controlled or planned effort to adapt.
Charles Darwin was aware of this problem, and searched for a theory that would solve them. His inspiration came from the way breeders produce new races of dogs, cats or cattle. Suppose you want to create a race of very big dogs, which would be more frightening to burglars. Then, you would start with the biggest dogs you can find and let them breed. Not all the offspring of this first generation would have the same size. There is always some variation, even between siblings. Some dogs would be smaller than their parents, some would be bigger. You would then select the biggest ones, and let only those breed to produce the second generation. Again, some of the second generation would be bigger than their parents, some would be smaller. You would select the bigger ones to breed the third generation, and so on. Unlike the effects of stretching, the increases in size with each new generation would accumulate. Thus, in principle you could produce ever bigger dogs. The incredible variety in dog races, from the pocket-size Chihuahua to the giant Irish wolfhound, shows that selective breeding is very effective for creating new types and shapes.

The only thing Darwin needed to add was a mechanism of selection that did not involve a breeder trying to improve the race. This mechanism he found in nature itself. From the various offspring of a given parent, some will be better adapted to their environment than others. For example, in a savannah, where many animals are trying to eat the leaves of a few tall trees, taller giraffes are likely to get more food than their shorter siblings. Therefore, the tall ones are more likely to survive and to have a lot of offspring themselves. From that second generation, it will again be the taller ones that are most advantaged, and therefore they will have most offspring. Thus, different varieties of plants and animals are naturally selected by the environment: those that are most adapted or fit, are most likely to survive and reproduce. Herbert Spencer summarized this principle with the memorable—though somewhat exaggerated—phrase of “the survival of the fittest”. Thus, the environment plays a role similar to the breeder.

This mechanism of natural selection not only solves the problem of accumulation, but also the one of involuntary adaptation. The giraffe does not need to know that a longer neck will make it more fit to eat high leaves. Neither does it need to know that a pattern of brown patches will give it a better camouflage in the predominantly yellow and brown savannah landscape. It is just that if some giraffes happen to be taller, or to have more pronounced patches, then they will be selected, and pass on their genes for tallness or brownness to their offspring. This principle is summarized by the phrase that variation is in general “blind”. Evolution by natural selection does not need foresight of which variations will turn out to be most adaptive. Variations take place in all directions,
positive as well as negative, but only the positive ones are retained. That allows evolution to produce ever fitter organisms.

Darwin introduced natural selection to explain the origin of species. Since there are many different environments, and within each environment, many different ways to make a living (“niches”), different species will evolve, which are each adapted to one specific niche. Thus, a single species can differentiate into a number of differently specialized breeds, such as the different dog races, which were selected for different purposes.

The theory of evolution was highly controversial when it was first proposed. Indeed, it seemed to imply that humans and animals descend from a common ancestor, and that the complexity of life can be explained by random mechanisms, without any design by a Creator. This idea was (and still seems to be) hard to swallow for many people raised in the Christian tradition. Moreover, the biological mechanisms responsible for variation and selection were not yet clearly understood. The origin of species also seemed irrelevant to physics, the central discipline of the new Newtonian world view. Finally, the emergence of novelty in evolution did not fit in with the Newtonian principles. Therefore, during the 19th century, Darwin’s theory of evolution was a matter of debate, rather than an integral part of the scientific world view.

In the 20th century, however, the new insights into genetics clarified the mechanisms of biological variation and heredity. This led to a more accurate, even mathematical theory of biological evolution: Neo-Darwinism. In spite of recurrent claims to the contrary, the present theory of biological evolution is hugely successful. The evidence gathered to support the theory is immense, and as yet there is no rival theory which can explain even a fraction of that evidence.

**The New Darwinians**

Biology was only the beginning of Darwinism’s ascent. Around the beginning of the century, thinkers such as Herbert Spencer and William James already noted that natural selection applies to other fields as well. More recently, the philosopher Daniel Dennett, in his book “Darwin’s Dangerous Idea”, has called natural selection “the single best idea” anybody has ever had. Its power lies in the fact that it is extremely simple, yet can be applied to the most diverse problems. Let us review some of its manifold applications.

One of the more obvious applications is scientific creativity. When a scientist tries to solve a problem, or find a new theory, he or she will typically consider alternative combinations of possibly relevant ideas. Most of these alternatives will turn out to be
wrong. Yet, one of them may happen to fit the data, and thus solve the problem. That combination will then be retained, while the others are forgotten. Thus, a scientist generates variation by recombining ideas, and applies selection by eliminating the poor combinations. This mechanism of variation and selective retention explains how science can efficiently accumulate good ideas. The philosopher of science Karl Popper extended this approach by noting that observations can never prove a theory, they can only perform selection by eliminating (“falsifying”) all theories that do not fit the data. Donald T. Campbell further generalized this approach by arguing that you can explain all forms of knowledge, scientific or other, as the products of variation and selection. Thus, he founded the domain of “evolutionary epistemology”.

During the last few decades, the application of Darwinian principles to new domains has boomed. For example, selection mechanisms have been used to explain the evolution of culture in general, including its different irrational aspects, such as fads, fashions, and cults. Selection is not only based on survival but also on reproduction. An idea “reproduces” if is is taken over by another person. In analogy with genes, the biologist Richard Dawkins called such replicating ideas “memes”. Successful cults or fashions are those which spread quickly, that is, which are good at reproducing themselves. Such principles may explain the emergence of cultural groups and institutions. Together with the results of sociobiology, which looks at the genetic evolution of social behavior, they may provide an evolutionary foundation for sociology. An as yet more popular approach is evolutionary economics, which sees business competition as an instance of natural selection.

Evolutionary psychologists consider the human mind as the product of natural selection. This allows them to explain a number of universal human characteristics, such as the psychological differences between men and women, what makes people sexually attractive, or the fear of being excluded. For example, our preference for gardens with large lawns and a few trees or bushes may have been shaped by the savannah landscape in which humanity first evolved. The neurophysiologists Gerald Edelman and Jean-Pierre Changeux have argued that our brain, perhaps the most complex system of all, develops through the variation and selection of assemblies of neurons.

Such “neural Darwinism” was in part inspired by the new understanding of our immune system. When our body is invaded by foreign agents, such as bacteria or viruses, we produce antibodies that are specifically targeted at the invaders. But how does our organism know which antibodies to produce? At first, researchers thought that the immune system somehow observes the structure of the invaders, and then produces antibodies to match that structure. This is the same thinking as the one that views
knowledge as a reflection of reality: just register your observations, and you will have knowledge about the phenomenon you were observing. Constructivism has shown that our mind does not work in that way. Research into the immune system has shown that our body does not work like that either. What happens instead is that the immune systems produces an incredible variety of different antibodies. A few of those antibodies match the structure of the invaders sufficiently well to be able to attach themselves to the foreign cells, and thus neutralize them. The most successful of the antibodies are reproduced in great quantities, at the expense of those that did not match. Thus, the mechanism is one of blind variation, followed by selection of the most fit antibodies.

The immune reaction is just one of the growing number of phenomena that fit under the header of “Darwinian medicine”. Another important application is the spread of infectious diseases. According to the theory of natural selection, successful parasites should either survive long or reproduce fast. If an infection is so virulent that it kills its host quickly, thus killing the remaining parasites as well, it must be successful at reproducing, that is, infecting new hosts. Parasites, such as the AIDS virus, that are slow at infecting new hosts must keep their hosts alive for a long period, otherwise they would be eliminated by selection. Therefore, by reducing the possibilities for infection, we should be able to make parasites evolve to a less deadly form.

Another application of Darwinian medicine is the evolution of the human body. By analysing the environment in which humans originally evolved, we can understand better which factors are healthy or unhealthy. This sometimes leads to surprises. For example, dioxins, a byproduct of many industrial processes, are notorious poisons. In tests with laboratory animals, the tiniest concentrations of dioxins turn out to be effective killers. Yet, in spite of the large worldwide production, no person has ever died from dioxin poisoning. Years after the Seveso disaster, in which a factory fire spread large amounts of dioxins over the Italian countryside, no one seems to have suffered permanent damage. People turn out to be much more resistant to dioxins than animals. Unlike animals, during thousands of generations, humans have been using fire, for heating and for cooking food. Since the smoke of burning wood contains quite some dioxins, natural selection appears to have promoted human genes that provide resistance to dioxin poisoning. Genes that could not cope have simply been eliminated.

A wholly different application of Darwinian reasoning to medicine is the manufacture of drugs. Just like the immune system produces the right antibodies by selecting those that match the invaders, researchers are now discovering new medicines by generating a large variety of different molecules and filtering out those that work. If you want to intervene in a physiological process, say a raise in temperature, you need to
identify the chemicals the body uses to control that process. If you could find a molecule that reacts with those chemical messengers, you could stimulate or block that process. Thus, you could develop a drug to combat fever, for example. Drug molecules typically function by binding either to the chemical signals themselves, or to the sites that produce them. To discover such molecules, you can use the target chemical or binding site as a template, and bring it into contact with billions of different molecules. The few molecules that fit the template will remain attached to it. They can thus be filtered out, and reproduced in great quantities. In that way, you can produce drugs to measure. Such Darwinian chemistry is obviously not limited to drug manufacture. Chemicals for any kind of function, say, hardening plastic or protecting metal, could be discovered by generating a large variety of candidate molecules and filtering out those with the right properties.

The same principles lie at the base of evolutionary computation. John Holland, one of the complexity theorists associated with the Santa Fe Institute, founded the domain of “genetic algorithms”. The idea is that if you wish to solve a complex problem on your computer, such as determining the shortest route connecting a host of different places (the “travelling salesman problem”), you generate a variety of candidate solutions and select the ones that score best. Until here, there is nothing new compared to the filtering of molecules. What distinguishes genetic algorithms is the next step, which is directly inspired by biological evolution. You allow the best candidates to “reproduce”, while exchanging parts of their components in the process. This is analogous to sexual reproduction, where parts of the chromosomes of father and mother are recombined to produce new chromosomes for the child. Together with this recombination, different “mutations”, small random changes, take place. This results in a variety of new candidates, which inherit most of the good qualities from their parents. You can guess what happens next: from this second generation, the best ones are selected and allowed to reproduce, bringing forth a third generation. After many rounds of reproduction, the best surviving solutions are usually much better than the ones from the first generation. Thus, the process of evolution can be run at very high speed on the computer, producing solutions to the most complex and subtle problems in a relatively simple way. The effectiveness of this method is perhaps the best evidence, apart from biology, for the power of Darwinian mechanisms.

From this review of evolutionary thinking, it appears that physics is about the only science untouched by Darwinian ideas. The reasons seem to be twofold. On the one hand, physics is the discipline that is most strongly rooted in the Newtonian world view. Physical theories, which are mostly still based on deterministic principles, leave little
room for the unpredictability and irreversibility of evolution. On the other hand, self-
organization seems to provide physicists with an alternative model for the creativity of
nature. To fully understand the mechanisms of creativity, we obviously must integrate
these different approaches.
The Need for Integration

Failed Unifications

The new scientific developments we have reviewed provide us with many interesting ideas, but no overall theory. They propose alternatives for many of the Newtonian assumptions. Yet, these alternatives have little in common, and it is not clear how they fit together. The growth of science has been accompanied by an increased specialization, and an explosion in the number of disciplines and subdisciplines. Each discipline has developed its basic concepts, but these are largely independent from the concepts of the other disciplines. With the erosion of the mechanistic world view, the global picture has been lost. Since the beginning of the 20th century, scientists and philosophers have therefore felt the need to integrate the different disciplines. The call for such “transdisciplinarity” is repeated by every new generation. Yet, until now, the results have been rather disappointing.

The first systematic attempt to synthesize the methods of the different disciplines was the movement for the unity of science. This movement was associated with the Vienna Circle, a group of brilliant philosophers working in Austria. Their aim was to create a universal language, in which all scientific ideas could be expressed. The philosophy they developed is called “logical positivism”. In a way, this philosophy takes the mechanistic world view to its extreme conclusion. Newtonian theory still assumes the existence of invisible phenomena, such as force or energy. Positivism says that all such unobservable entities should be banned from science. Only direct observations can guarantee accuracy and objectivity. Logical positivism adds that in order to build theories, these observations should be arranged in a system of logic. The observations themselves are the logical “atoms” of the theory, the smallest units of meaning.

Positivism fell apart when it became clear that observations will never be sufficient to induce a theory. It is ironic that two of the original members of the Vienna Circle, Ludwig Wittgenstein and Karl Popper, later became the most influential critics of its positivist philosophy. The failure of logical positivism did not immediately kill the attempts to bring the different sciences back together, though. Otto Neurath, another member of the Vienna Circle, remained active as editor of the “Encyclopedia of Unified Science”. The most famous book in that series was published in 1964: Thomas Kuhn’s
“The Structure of Scientific Revolutions”. It is a double irony that this is the book which showed that there is no universal language of science. According to Kuhn, different scientific theories or “paradigms” each have their own language of terms and concepts. These languages essentially stand on their own: they cannot be translated the one into the other.

Recently, another attempt at unification has gotten a lot of publicity: the so-called Grand Unified Theory (GUT). This (as yet non-existent) theory is the Holy Grail of physicists. It would unify the quantum field theories, which describe elementary particles, and the theory of general relativity, which describes the geometry of space-time. At the same time, it would bring together the different known forces of nature in a single mathematical model. As we noted earlier, the mathematical complexity of this enterprise is bewildering. Einstein spent the last decades of his life trying to develop such a theory, but failed. Yet, many physicists believe that recent advances, such as the theory of “superstrings”, may finally bring the Grand Unified Theory within reach. Some of them have been touting that eventual unification as the “Theory Of Everything” (TOE), the magical formula which would answer all questions about the universe in which we live.

Even if such a GUT would be found, it should be clear from our review of recent developments that such a theory would solve few of the outstanding problems. Though it might unify quantum mechanics and relativity theory, it would have nothing to say about thermodynamics, chaos or self-organization. Thus, it would not even bring back unity to physics. To assume that it would solve the problems of biology, psychology or sociology only shows poor understanding of those disciplines. To suggest that it would answer the questions raised by artificial intelligence, constructivism or the theory of evolution is simply preposterous. At best, the “Theory of Everything” would show us the common origin of electromagnetism, gravitation and the nuclear forces, while unravelling the microscopic structure of space, time and matter. The belief that it could explain phenomena outside of physics, such as life, mind or society, is merely a vestige of old-style reductionism.

The last attempt at transdisciplinary unification we will discuss is the General Systems Theory (GST). As we noted earlier, systems theory searches for universal principles of organization, which apply across all domains, from physics to chemistry, biology, psychology and sociology. The economist Kenneth Boulding, one of the founders of GST, has argued that the phenomena studied by these disciplines can be arranged in a hierarchy, ordered by complexity. While the traditional disciplines each remain within their level of the hierarchy, GST tries to explain how the different levels
are connected to each other. Thus, Boulding argued, systems theory can play the role of “the skeleton of science”.

From the different attempts, GST seems to have been the most fruitful. Systems theorists and their close collaborators, the cyberneticists, have indeed shown that concepts such as input-output, hierarchy, self-organization, entropy, information and feedback, can help us to understand systems at all the disciplinary levels. Moreover, they may explain how one level can emerge from another level. Thus, the concepts of systems theory and cybernetics provide the beginning of a universal scientific language. This language avoids the shortcomings of logical positivism because it doesn’t reduce phenomena to logical atoms. Instead, it focuses on connections, interactions and processes, thus weaving the different parts or aspects of a phenomenon into a coherent whole.

Yet, the picture offered by systems theory is everything but complete. Plenty of gaps remain. One reason for this failure is that it is very difficult to get rid of all Newtonian limitations. Systems theory is in many respects still too reductionist and static. In particular, it has failed to fully absorb the idea of evolution. It looks at systems as given entities, which are neither created nor destroyed. It also has not completely assimilated the new philosophy of constructivism and its views of knowledge. A second reason is that integration of all the different scientific theories is an extremely complex task. The number of concepts and principles that must be examined, reformulated and integrated in an encompassing framework is perplexing. Even with good ideas and plenty of motivation, this problem demands many people with expertise in the different disciplines working together efficiently for many years.

This brings us to the third reason for the failure of systems theory: lack of organization. Systems theory and cybernetics have never really been accepted in the academic world. There are few departments researching systems theory in the different universities. There are even fewer courses. There are almost no universities where you can get a degree in systems theory. Yet, the many well-attended congresses show that quite a lot of researchers are interested in the systems approach. The problem is that these people generally don’t work in a systems department. They officially belong to various specializations, from mathematics to management. They do their systems work almost as a side-activity. Few scientists can afford to work full-time on systems theory. Even those who do are generally based in a department focusing on a more specialized domain. These systems researchers have themselves been educated in one of the traditional domains. They may be physicists, psychologists or biologists, but they practically never
have a degree in systems theory. Thus, their contributions tend to be biased towards their own speciality, while ignoring the contributions from other disciplines.

This, together with the intrinsic complexity of the problem, leads to a fragmented approach. Different researchers work in parallel on the same problem, without being aware of each other’s results. Each researcher tends to develop his or her own terminology or conceptual framework, which does not fit in with the one of the others. The most basic reason for systems theory’s failure to unify science is that it is not unified itself. In fact, there is not really something like “the theory of systems”. There are dozens of theories, approaches and conceptualizations. Although they share a general systems-inspired philosophy, they differ in the concrete details. Therefore, people working in the systems domain nowadays call themselves “systems researchers” or “system scientists” rather than “systems theorists”.

This fragmentation has led me to get involved in an organization that aims to remedy the problem: the Principia Cybernetica Project.

The Principia Cybernetica Project

The Principia Cybernetica Project was conceived by Valentin Turchin, a Russian cybernetician. Turchin began his research career in physics, but turned to the new field of computer science in the 1960’s. He created REFAL, the Russian answer to the programming language LISP which is used for artificial intelligence. Computing and cybernetics gave him the inspiration for the concept of metasystem transition (MST), the emergence of a higher level of control. He used this concept to analyse evolution as a sequence of transitions which each produce systems at a higher level of complexity. These levels are similar to the hierarchy of levels proposed in systems theory. The difference is that Turchin focused on the evolutionary process that creates these levels. Thus he escaped the basically static framework of systems theory.

Val Turchin quickly discovered that he could apply his concept of metasystem transition to a host of different issues. Some difficult problems in programming, such as the generation of compilers for programming languages, could be tackled through MSTs. Turchin set out to analyse the evolution of life as a sequence of MSTs. This brought him to the level of human culture, which he again characterized as a sequence of MSTs. He even applied his cybernetic philosophy to society, arguing for a political system very different from the totalitarian Soviet state in which he lived.

This merely added to the troubles he already had with the authorities. Working with the famous dissident and Nobel prize winner Andrei Sacharov, he had founded the
Moscow chapter of Amnesty International. Because of these activities he lost first his research laboratory, then his job. While being harassed by the KGB, the infamous secret police, he had to hide his book manuscripts, in the constant fear of being discovered. At this point he was facing long-term imprisonment. Because of his high profile activities, the authorities proposed him a less threatening alternative: permanent exile from the Soviet Union. Though he did not like the idea, he had little choice. Thanks to an offer to teach at the City University of New York, he managed to immigrate to the USA with his family, in 1977. There he continued his research, while his two book manuscripts, “The Phenomenon of Science” and “The Inertia of Fear and the Scientific Worldview”, were translated into English and published by the Columbia University Press.

In 1987, he met Cliff Joslyn, a young researcher who had been reviewing his books, and been very much impressed by them. Joslyn was one of the rare people who had an official degree in systems science, although his background included software engineering and cognitive science as well. Turchin told him about his plans to further develop his ideas into an integrated philosophical system, while involving different contributors. Instead of the hierarchical structure which Turchin imagined, Joslyn proposed to develop the system as a network of linked concepts. The new technology of hypertext would allow this network to be efficiently implemented on a computer. The different contributors could then use electronic mail to communicate and discuss the developments. These new electronic technologies should make the ambitious project more manageable, by speeding up communication and publication, while keeping the growing system of concepts both flexible and well-organized.

Thus, Turchin and Joslyn founded the Principia Cybernetica Project, and became the first members of its editorial board. The name “Principia Cybernetica” (Latin for “cybernetic principles”) was inspired by “Principia Mathematica”, the famous series of books in which the philosophers Whitehead and Russell developed the foundations for 20th century mathematics, thus unifying its different branches. Characteristic for Russell and Whitehead was that they used mathematical methods to analyse mathematical reasoning itself. Similarly, the Principia Cybernetica Project applies cybernetic concepts and technologies to the development of cybernetics itself. It is probably not a coincidence that “Principia Mathematica” is also the title of the book in which Newton proposed his theory of mechanics, which would form the basis of the scientific world view.

In 1989, Joslyn and Turchin wrote down their ideas in the form of a proposal, which they circulated on CYBSYS-L, an electronic mailing list devoted to cybernetics and systems theory. This proposal garnered many reactions, some enthusiastic, some very critical. When I read the proposal, I was struck by the similarity with my own ideas to
create a network of people that would develop a theory of complexity and evolution. Their philosophical ideas also resonated strongly with my research. It quickly became clear that we basically wanted the same things, and so I joined Val Turchin and Cliff Joslyn to become the third member of the project’s board.

Since then, the Principia Cybernetica Project (PCP) has been doing quite well. Many people have contributed in one way or another to its developing cybernetic worldview, by writing texts for PCP publications, giving lectures at PCP organized conferences, or simply by discussing the ideas in public or private debates. Cliff Joslyn’s vision of an electronic network to support the collaborative development of the project became reality with the appearance of the World-Wide Web. PCP started its own web server in July 1993, at a time when there were still only some 200 web sites around. Unlike other web sites, Principia Cybernetica Web was organized from the start according to a well-thought-out philosophy. Thus, the network of documents could grow and develop, without turning into a labyrinth where you would never find the things you are looking for. At the moment of writing, it contains some 1400 cross-linked documents, describing the different concepts and principles underlying Principia Cybernetica’s world view. From the beginning the web was also designed to be interactive, so that in any place users could add comments (“annotations”) to the text they read. Five years after the start of the project, our progress report noted that over two thirds of Cliff’s ambitious list of objectives had been realized.

In spite of its innovative methods for collaborative development, the central aim of PCP is still the creation of an integrated cybernetic philosophy. It is more difficult to determine to what degree PCP has achieved this last objective. First, it will always be debatable in how far PCP’s philosophy is as “complete and consistent” as the initial proposal required. Obviously, a project as ambitious as the creation of an evolutionary, cybernetic world view will never be finished. There will always remain issues that need further details and clarifications. Second, most of the components of the Principia Cybernetica philosophy already existed in some form before the start of the project, in the research of Turchin, Joslyn or myself, in the work of other contributors, such as Donald Campbell and Joel de Rosnay, or even in the legacy of the original visionaries, such as von Bertalanffy, Shannon, Wiener and Ashby, whose innovations inspired the project. There is room for debate in how far the project’s results are genuinely new, or merely the collection and organization of existing ideas.

Still, I believe that the philosophy developed within the project has reached the stage where it deserves to be brought to the attention of a wide public. This is already happening through the project’s web server, which is being consulted every day by some
thousand people. Yet, few of those people will grasp the larger picture, or get a sense of the coherence and extension of the philosophical system. Therefore, this book is intended to convey my own understanding of the Principia Cybernetica world view. Although my PCP colleagues are likely to agree with most of what I write, my account is personal, emphasizing the ideas that I find most important. PCP allows room for both consensus statements and individual contributions, with which the other participants may not fully agree. The present book should be seen as such an individual contribution.

In this personal view, I want to start from the process of evolution. First, I will analyse the underlying mechanism of evolution, arguing that self-organization and natural selection are merely two sides of the same coin. Then, I will show how these mechanisms spontaneously produce ever more complex systems. Thus, I hope to convince you that something can emerge out of nothing, that evolution is fundamentally creative. There is no need to postulate the laws of nature, the hand of God, or a hidden blueprint for evolution, if you want to explain how the world came into being. The presence of variation and selection, a self-evident assumption, is all you need to kick-start a self-organizing universe. Once the principles are clear, I will reconstruct the main stages in the evolution of the universe, showing how the major levels, studied by the different disciplines, arose: matter, life, mind and society. Thus, I hope to reveal how the results of the scientific disciplines fit together in this encompassing picture of the world.
PART IV: First Principles

The Difference that Makes a Difference

Ockham's Razor

Our discussion of the Newtonian world view concluded that it is based on two principles: analysis and causality. The first one says that to understand a phenomenon, you must take it apart, distinguishing as finely as possible between its different components and properties. The second principle says that after you have accurately made those distinctions, you will be rewarded by the observation that these distinctions are conserved. Conservation of distinctions is the same as causality. It means that given the present situation you can predict all future effects, and retrodict all previous causes. Thus, you can reconstruct the complete trajectory of the system, from the distant past to the distant future. The reason you can determine the trajectory unambiguously is because trajectories never intersect: they always remain separate. If the phenomena around you seem unpredictable, it is only because you have not distinguished their trajectories precisely enough.

Our review of more recent scientific developments paints a very different picture. New theories such as quantum mechanics and chaos have shown that you will never be able to make distinctions that are conserved. For example, if you distinguish precisely the position of a particle, the uncertainty principle says that you will necessarily lose all information about its speed. Therefore, you cannot predict how the particle will react. The same causes can lead to different effects, while different causes can lead to the same effects. Thus, the principle of causality has lost its universality.

Systems theory has made it clear that the principle of analysis is limited as well. Because the whole is more than the sum of its parts, analysing a system into its components will never allow you to completely understand that system. Systems theory
tells you that you should not only consider the smallest, atomic distinctions between the parts, but also the larger distinctions, which separate a whole from its surroundings. You cannot understand the human body by focusing on its cells, molecules or even atoms. You must look at the body as a whole to understand the function of the molecules in the cells, of the cells in the organs, and of the organs in the organism.

The theories of self-organization and evolution have taught us that distinctions between wholes or parts are dynamic: they evolve, appear and disappear. The same applies to the mental distinctions we use to categorize the world. Constructivism has shown that we continuously build new theories, including new distinctions, to make sense of our observations. The observations do not create the distinctions. It is rather the other way around: our mental distinctions, constructed by our ever active mind, determine which phenomena we will pay attention to.

These conclusions are basically negative. They tell us that the principles of distinction conservation are no longer valid. They don’t tell us which new principles should replace them. My purpose here is to try and formulate such new principles. On the one hand, these principles should explain why the Newtonian approach often works. They should remind us that analysis is a useful method for understanding complex phenomena, and that many, though not most, processes are causal, and therefore predictable. On the other hand, they should also explain the non-Newtonian phenomena, which are characterized by emergence, unpredictability and self-organization.

To discover such principles, we have very few guidelines to help us. Since these principles must be absolutely general, applying to any kind of system or process, they cannot be based on observations. An observation can only tell us about the specific phenomenon we perceive. It cannot tell us about other phenomena. Therefore, as we saw when discussing constructivism, scientists need to make conjectures that fill in the gaps. There are an infinite number of conjectures that all can explain the same observations. One method that can help us to choose the best one is called “Ockham’s razor”.

William of Ockham was a medieval philosopher, who emphasized the need to think clearly. He provided the inspiration for the monk in Umberto Eco’s famous mystery novel, “The Name of the Rose”, who, like a medieval Sherlock Holmes, used logical reasoning to solve the grisly murders. Ockham noted that if you wish to explain something you should not use more assumptions than needed. In other words, you should keep your explanation as simple as possible. For example, if you find a man dead at the foot of a stair, without further evidence, you would rather assume that he has fallen and fractured his skull than that he was beaten to death. The first explanation only requires gravity. The second explanation requires at least another person, a murder weapon, an
occasion, and a motive for killing. Of course, the charm of detective novels is that the simplest explanation is rarely the right one. In real life, however, the simplest explanation is more likely to be true than any of the more complicated ones.

Ockham’s principle of simplicity reveals its power in scientific theorizing. When you are confronted with a multitude of possible theories that all explain the same facts, simplicity is often the only criterion available to choose the best one. For example, suppose you have made two observations, represented by two data points on a diagram, and want to find a general rule describing all possible further observations. One way is to draw a straight line through the two points, and assume that all further observations will lie on that line. However, you could also draw an infinite variety of the most complicated curves passing through those same two points, and these curves would fit the data just as well. Only simplicity can guide you in choosing the “straight” relation as best candidate theory.

In practice, you will rarely have to choose between a large number of rival explanations. Often, it is already difficult enough to come up with just one plausible hypothesis. This hypothesis might be suggested by an analogy with another case or by theoretical reasoning, or it might just be a hunch. However, your conjecture may in fact say more than really needed, and thus unduly limit your explanation. For example, suppose you want to explain the following observation. Lying under an apple tree, you witness several times how an apple gets disattached from the branch and falls to the ground. Your first hypothesis might be that if something is an apple, and it is no longer supported by a branch, then it will fall. All further observations confirm your hypothesis, so it looks like a good candidate theory. However, your theory makes some very specific assumptions, which are of doubtful utility. Does it really need to be an apple, or even a fruit, in order to fall? Does it need to get disattached specifically from a tree branch? The simpler hypothesis is that any object which is not supported will fall to the ground. This simplified theory can in fact predict a much larger number of cases, because it is more general.

It is in cases like this that Ockham’s method is most useful. What Ockham tells us is that from our hypotheses, we should remove all assumptions that are not strictly needed. Thus, we can reach a maximally simple and general theory. That is why the method is called “Ockham’s razor”. It helps you to “shave off” everything superfluous, making your theory sleek and shiny. One of the reasons for the effectiveness of science is precisely that scientific theories have done away with the host of gods, spirits and miracles which the earlier world views used to explain nature. Science strives to achieve
the most simple and general theories, by ruthlessly throwing out ideas—such as “caloric”, “ether” or “life force”—that don’t really explain anything.

However, that does not mean that science no longer needs Ockham’s advice. Science too has the tendency to accumulate “ballast”: concepts and rules that make a theory needlessly complicated. The reason is simple: when your otherwise reliable theory is challenged by observations it cannot explain, you would rather add some assumptions than throw away the theory and start from scratch. Perhaps, those additions are really needed, because your theory was too simple to catch the intricacy of nature. But it is possible that another, simpler theory might explain both the previous and the new observations. The rare cases where a theory is thrown overboard and replaced by a new, more general one are rightly called “scientific revolutions”.

But like we saw with classical mechanics and the subsequent non-classical theories in physics, even these revolutionary theories inherit much of the concepts from their predecessors. Thus, theories tend to grow and become more complex, absorbing additional concepts, rules and even complete new layers of explanation. For example, quantum field theories have developed out of classical mechanics, which subsequently absorbed the layers of quantum mechanics, special relativity, and field theory. As a result, the theory becomes so top-heavy that it can in practice only describe the simplest cases. This will only get worse in the hoped-for Grand Unified Theory, which would add a layer for space-time geometry, and possibly one for superstrings. Such theories become more and more difficult to understand, to apply and simply to work with. It is no surprise then that progress in fundamental physics has very much slowed down since the development of quantum field theory around 1940. This virtual standstill has even led the science writer John Horgan to announce the “end of science”. (Although the “end of physics” might have been a more appropriate moniker.)

Apparently, the only way out is to take Ockham seriously, and start anew. By using the simplest possible concepts and principles as a foundation, we may hope to build a broad framework, without drowning in complexity. When we look back at all the classical and non-classical theories we discussed, one concept seems to emerge again and again: distinction. When we observe the world, the first thing we must do is distinguish between different phenomena. A distinction can be seen as the unit of knowledge. Without distinction, there would be nothing to perceive or to describe. Distinction is the basis of classification. Knowledge is the recognition of regularities, of recurrent relations between different classes or categories.

Let us consider a very simple piece of knowledge: the rule which says that heavy things fall. We can formulate this elementary “theory” more precisely as follows: if
something is heavy (H), then it will fall (F). Reduced to its essence, it would read: H → F. What do H and F stand for? H denotes the category of heavy things, F the category of things that fall. What does that mean in practice? To apply the rule H → F in a concrete situation, we must determine which things belong to the category H, and which belong to the category F. In other words, we must distinguish between the things that belong to H or F and those that don’t. One way to distinguish between H and non-H is to use a balance. If you put the thing you are testing on a balance and the balance gives you a non-zero weight, you can conclude that the thing belongs to H. This is a quite accurate method to observe. However, you might as well use the intuitive method of taking the thing in your hand and feel whether it seems heavy or not. The distinction you would make might be less precise than the one using the balance, but it would still be a distinction. Similarly, you can determine whether a phenomenon belongs to the category F or not, for example by dropping it and seeing whether it reaches the ground. If it is a balloon filled with gas, it will not reach the ground; if it is a stone, it will.

There are in general many ways to make a distinction. The important thing, however, is that a distinction can be made. If I would tell you that all things that are “wonderful” (W) are also “marvelous” (M), you would be hard-pressed to suggest a method to distinguish wonderful from non-wonderful things. Perhaps, you might still have an intuitive sense of what is wonderful, though. But if I now tell you that if something is “pretertorious” it is also “oblacquious”, it would just be non-sense to you, since you cannot in any way imagine how an oblacquious phenomenon would differ from a non-oblacquious one. All knowledge can be expressed as a system of “if...then...” rules, which connect one distinction to another one. These distinctions are meaningful just because they are connected to other distinctions. In that sense, a distinction is a difference that makes a difference. The difference between oblacquious things and non-oblacquious things does not lead to any other differences. You cannot perceive or determine which things are oblacquious things. The statement that something is oblacquious would not in any way help you to predict anything else than oblacquiousness itself. Since oblacquiousness does not explain anything, we can use Ockham’s razor to cut it out of our theory.

Thus, we may conclude that distinctions which don’t lead to further distinctions are void. A similar principle was formulated by the philosopher Leibniz, Newton’s rival in the development of calculus and its implications for the scientific world view. Leibniz’s principle is called “the identity of the indistinguishables”. It says that if we cannot distinguish two entities in any way, then we must conclude that they are in fact
one and the same thing. Thus, there was no reason to consider two different things to start
with.

This principle can be applied to the most diverse cases. For example, when the
vitalists proposed that it is the presence of a life force that distinguishes living systems
from non-living ones, it was up to them to show how this difference made a difference.
When others showed how the properties of living systems, like reproduction or
metabolism, could be explained by their chemical and physical organization, there was no
longer any need for a life force. Indeed, if systems that reproduced without a life force
could no longer be distinguished from those that reproduced with a life force, then the
whole concept of life force became useless. Ockham’s razor made scientists throw out the
concept.

The demise of vitalism is now generally accepted. However, intelligent people
still sometimes fall into the trap of proposing ideas that are by definition incapable of
explaining anything. A recent example comes from the study of consciousness. Although
most scientists study consciousness by investigating its functioning and effects, such as
awareness of outside events, activity in the brain, or mechanisms of thought, the theorist
David Chalmers has called attention to what he calls the “hard problem” of
consciousness. To explain this, Chalmers proposed the following illustration. Imagine a
normal person like you and me, who has conscious experience of what is going on. Imagine
now a wholly different creature, a “zombie”, who looks, talks and behaves like a
normal person, but who merely functions like an automaton, without any conscious
experience. Scientists using their traditional methods would not notice any difference
between the zombie and the normal person. Therefore, Chalmers concludes, they are
missing out on the hard problem of what consciousness really is.

This may seem like a deep philosophical reflection on the limits of science and
the essence of consciousness. However, if we reason like Leibniz, we must conclude that
the hard problem does not exist: it is merely a creation of Chalmers’ imagination. Indeed,
if the zombie is in all respects indistinguishable from a normal person, then Leibniz
would tell us that there is no difference. How else would you know that the people
around you aren’t zombies already? You assume they have conscious experience similar
to yours because they behave in all other respects similar to you. But if you would take
Chalmers’ reasoning seriously, then you might start to get nightmarish fantasies in which
you are the only real, conscious person in the world, and all the others are merely
sophisticated automatons that pretend to be like you. In conclusion, Ockham’s razor
proposes a very easy solution to Chalmers’ “hard problem”: just forget about it!
Leibniz’s principle is a very powerful aid for simplifying our theories. However, if properly thought through, it can give us much more. Indeed, the “difference that makes a difference” is just another way of describing a process that conserves distinctions. And distinction conservation is the same as causality. Thus, Ockham’s razor together with the concept of distinction has allowed us to reconstruct the basis of the classical world view.

We have managed to derive the basic principles of Newtonian science without needing all its additional assumptions about matter, space, time, atoms, forces and energy. What rests us now, is to show how this reasoning can be extended to cover the non-Newtonian phenomena as well. But first, we must investigate how the concept of distinction can help us to understand order, disorder and complexity.

**Order and Disorder**

Newtonian science sees the world as ordered, regular and harmonious. Post-Newtonian science has attracted our attention to the essential role of disorder, chaos and complexity. To understand the relation between order and disorder, we must first define these concepts more precisely. Order is characterized by invariance, repetition and predictability. In an ordered system there are no surprises. Everything follows the rules. Nothing departs from what is expected.

This intuitive understanding can be made more precise by introducing the concept of *symmetry*. A symmetric pattern is characterized by some kind of repetition. For example, patterns on wallpaper constantly repeat themselves after even distances, both in the vertical and in the horizontal directions. Similarly, a mirror symmetric figure, such as the human face, can be created by simply taking the left half, and adding its reflection on the right. In mathematics, such repetitions ‘with a twist’ are represented by geometrical *transformations*. The simplest transformations will shift a pattern horizontally or vertically, like on wallpaper. More complicated transformations will rotate the pattern around an axis, reflect it along an edge, magnify it, or reduce it. A pattern is said to be symmetric if its appearance is not changed by a group of such transformations. For example, if you shift wallpaper up or down over the right distance, the picture you see will be the same as before the shift. If you take a uniformly colored ball, and rotate it, the result will look exactly the same as before the rotation. Whether you look at somebody’s face directly or in a mirror, the image you see will be the same. On the other hand, if you would take a non-symmetric pattern, such as the text on a page, and look at it in the mirror, the result would be dramatically different.
Such transformations do not affect a symmetric pattern, because the parts of the patterns simply repeat themselves. The transformation merely moves a part of the pattern to a place where that same motif already exists. Therefore, it is as if nothing has happened. Assuming that your left and right eyes are exactly the same, you would not see any difference if the two had switched places (like they do in the mirror). There is no distinction. That is what makes symmetric patterns predictable: you only need to know a part of the pattern (say, your left eye), and still you can tell how the other part (say, your right eye) will look like. The same pattern appears again and again. That is why symmetry is the essence of order.

There are different types of symmetries, depending on the specific transformations that leave the pattern invariant. Your face is symmetric under reflections, but not under rotations. Turn your head 90 degrees to the right, and you’ll see a very different image. The symmetry of the human face is quite limited. The more transformations you can perform on a pattern without changing its appearance, the more symmetric it is. A brick wall, where the same brick shape is repeated every few inch, both vertically and horizontally, is more repetitive than a human face. Crystals, where the same molecular structures are repeated in all three dimensions, are invariant under a whole range of transformations, including reflections, shifts over specific distances, and rotations over specific angles.

The more transformations leave the pattern invariant, the more symmetric, predictable or ordered the pattern is. This also means the less information you need to reconstruct the pattern. Indeed, the more transformations you can apply, the easier it is to reconstruct the complete pattern from a small piece. You just need to apply all the transformations to the fragment in order to generate all the other parts by repetition. For example, with a brick wall you just need to know the shape of a single brick in order to reconstruct the complete wall. The more transformations there are, the smaller the part that can be used as “seed” to produce all the other parts.

What is the most ordered pattern of all? The maximum of symmetry is empty space. Indeed, no transformation, whether it is a reflection, rotation, magnification or shift, will change the appearance of empty space. Perfect order is completely predictable. You don’t need any information at all to know how it looks. The uncertainty, and therefore the entropy, is zero. Imagine an infinitely extended, perfectly homogeneous sheet of paper. Whether you move the paper sideways, look at it in the mirror or under a microscope, or rotate it, you won’t see any differences. Perfect order is characterized by complete lack of distinctions. No part is in any way distinguishable from any other part. If we apply the principle of the identity of the indistinguishables, we must conclude that
there aren’t any parts. There simply isn’t anything at all. Perfect order is the same as emptiness.

This means that for something to exist, order can never be complete. Let us then try to understand what lack of order, or “disorder”, precisely means. We defined order as invariance of a pattern under transformations. Disorder then means that a pattern will vary under transformations. The parts of a disordered system are all different. Knowing one part doesn’t give you a clue about what the other parts would be like. Disorder means unpredictability or lack of repetition. The uncertainty, and thus the entropy, is large. For example, the grains of sand on a beach have all different shapes and positions. While a crystal is the prototype of an ordered system, a gas is the prototype of a disordered system. A gas is a cloud of molecules that move chaotically. The position of one gas molecule does not give you any information about the position of the other molecules. The different molecules are completely independent. Yet, a gas is not perfectly disordered either. Indeed, most of the space in a molecule cloud is still empty, and empty space, as we saw, is ordered. Moreover, in between collisions, the individual molecules move in straight, predictable trajectories.

So, what would be the limit of complete disorder? We should try to imagine a cloud where every single point behaves differently from every other point, and this in a completely unrepetitive way. Perhaps a good metaphor would be foam, consisting of tiny bubbles of different shapes that are constantly forming, expanding, shrinking or bursting, while filling the whole of space. However bizarre this notion may seem, theoretical physicists would probably recognize it. Indeed, quantum theory predicts this kind of structure for empty space. The quantum picture of the vacuum is one in which “virtual” particles are constantly being created and destroyed in every corner. The particles are called “virtual” because they are so short-lived that you cannot detect or manipulate them. They might as well not be there. In that respect, the quantum picture of space is the same as the Newtonian one: one big emptiness. The only difference is that in exceptional circumstances, such as in the neighbourhood of a black hole, virtual particles may become real. Some theorists believe that the universe originated out of such a “quantum fluctuation of the vacuum”.

In conclusion, the limits of order and of disorder appear remarkably similar: emptiness. To understand this better, we can go back to the concept of symmetry. Not only perfect order but perfect disorder is symmetric, albeit in a different way. In disorder, where every part is different from every other part, there is no repetition. However, because the parts are independent, the probabilities of finding different parts in different states are the same. The parts are not influenced by each other, and therefore any single
part is as likely to be in a particular state as any other part. Remember that a gas, the prototype of a disordered system, is statistically homogeneous: any position in the cloud is as likely to contain a gas molecule as any other position. Actually, the molecules themselves are not evenly spread. But if we consider average positions, the law of large numbers tell us that the spread will be symmetric or homogeneous. The differences in behavior between the individual molecules disappear at the macroscopic level, where the gas is homogeneous. These are differences which don’t make a difference!
This principle is illustrated in Fig. 1. On the left side you see a square filled with a regular grid, where every black dot is followed by a white dot, and vice versa. The seed pattern of one white next to one black dot is identically repeated throughout the square. Therefore, knowing the color of a single dot is sufficient to determine the colors of all other dots in the square. On the right, you see the same square filled with a random distribution of black and white dots. The effect is similar to the “snow” which you can see on a TV screen when it is not tuned to a station. The color of any dot is independent.
of the color of any other dot. Therefore, in order to reconstruct the pattern in the square you need to have complete information about the color of all the dots. Whereas the first picture might represent molecules arranged in a crystal, the second might depict a dense gas or the “space-time foam” that fills the quantum vacuum. Yet, if you look at the two squares from a sufficient distance they will look uniformly grey, without any structure or distinction.

In conclusion, the extremes of order and disorder touch. Whether you make all parts of the system identical, or all parts different, the net result is that you lose all structure. To make meaningful distinctions, you must have a system which is neither completely ordered nor completely disordered, but somewhere in between. This position in between order and disorder we will call “complexity”.

**Complexity**

Complexity is a fashionable concept. Everybody agrees that our present society is getting ever more complex. Post-Newtonian scientists have become aware that the universe is much more complex than we used to think. The “sciences of complexity” are all the rage at the moment. Yet, it seems that no one can really explain what complexity is. Most complexity scientists would agree with the suggestion that complexity is situated in the no man’s land between order and disorder, or—to use a more fashionable expression—”on the edge of chaos”. However, this tells us more about what complexity is not—neither order nor chaos—than about its positive features. The different definitions proposed by complexity theorists have only one thing in common: their limitations. All definitions fall short in one respect or another. They either classify something as complex which we intuitively would see as simple, or call an obviously complex phenomenon “simple”.

- 180 -
One way to produce a pattern in between order and disorder is simply to combine the randomness characterizing disorder with the symmetry characterizing order. A simple illustration is the kaleidoscope, where you shake a collection of colored beads to produce a random arrangement, and then look at the symmetric pattern formed by its reflections through a set of mirrors. Fig. 1 shows a similar computer-drawn pattern, where randomly drawn lines have been reflected along four axes. Patterns like that have a weird kind of beauty, but are they complex? In really complex systems, such as a society, an organism or the ecology of the rain forest, you cannot separate the random and the symmetric features so neatly. There certainly is both randomness and symmetry involved, but they rarely exist in a pure form. Even the human body is not really symmetric: if you look at the inside organs, they tend to be either on the left (the heart) or on the right (the stomach).

To get a better grasp of what complexity means we must go back to its Latin root, *complexus*, which means “twisted together”. This means that in order to have a complex you need two or more distinct strands, which are joined in such a way that it is difficult to separate them. Similarly, the Oxford Dictionary defines something as complex if it is “made of (usually several) closely connected parts”. In essence, then, complexity requires
both *distinction* and *connection*. Without distinction, there are no parts, only a simple, uniform whole. Without connection, the parts are independent, like the molecules in a gas. This means pure disorder, which, as we have seen, is basically simple.

A system will be more complex if more parts can be distinguished, and if more connections between them exist. For example, a hammer consists of a head connected to a handle. Therefore, it is minimally complex: two parts—that is, one distinction—and one connection. A clock, on the other hand, consists of hundreds of distinct parts connected in lots of different ways. It is definitely more complex than a hammer. The human body consists of billions of cells, tissues, molecules, organs, connected by the most intricate chemical bonds, interactions and circuits. It is immensely more complex than either a clock or a hammer.

The more parts a system has, the more items we will have to enumerate if we want to describe that system. Therefore, more complex systems will require longer descriptions. Such descriptions could be simplified if the parts would be similar, like the molecules in a crystal or the links in a chain. Then you could describe the system simply as “take one link (or molecule), attach another one to it, and repeat this operation a thousand times”. This would characterize an ordered, repetitive system. You could similarly shorten the description of a disordered system. To describe a gas, your rule might be “take a molecule and put it in a random position. Repeat a billion billion times.” This description would not exactly match the gas you want to describe since the random positions would not be the same in the description and in the real gas. Moreover, each application of that description would produce a different distribution of positions. But statistically, the law of large numbers guarantees that the different descriptions and the real gas would be indistinguishable in practice, since many random choices in the end produce a homogeneous distribution.

What characterizes complex systems is that their descriptions cannot be simplified in this way. The molecules in a complex system, such as a cell, depend on each other in a variety of subtle ways. There is neither repetition nor statistical independence. That makes it difficult to model such a complex system. To be reliable, the model should be almost as complex as the system it represents. Therefore, models or theories describing complex systems are either unreliable or so complicated that they are difficult to use. The number of components that must be included in the description is so large that it exceeds the capacity of our brain, or even the capacity of the most powerful computers. Answering questions or solving problems about complex systems is intrinsically hard. That is why “complex” is often used as a synonym for “difficult”. 
In fact, the most popular measure of complexity for a pattern is the length of its shortest complete description. In other words, simple things are those that can be described with few words or symbols. Although this definition captures our intuition that complex phenomena are difficult to represent, it overlooks several issues. First, in practice you can never find complete descriptions of a realistically complex system, such as the human body. Second, even if you had a complete description, you could never be certain that it was the shortest possible. There may always be further methods to compress or simplify the description. Most importantly, the length of the description will depend on the language or code you use to describe the phenomenon. A complicated geometric shape, such as the Mandelbrot set, is virtually impossible to describe in natural language, but may be expressed very concisely by a mathematical formula. On the other hand, the description of an every-day situation, such as going to the restaurant, is trivial in natural language, but unmanageably long-winded if you would use mathematics. Since, as we argued, there is no universal scientific language, there cannot be a universal description, and therefore no universal description length.

We started our discussion of basic principles by introducing the concept of distinction. Let us now look at the second concept needed to understand complexity: connection. Distinctions are connected if information about the one also gives you information about the other. Knowing the one, you can infer likely features of the other. But that information is never complete. Otherwise, connections would be the same as the repetitions or “transformations” that map one part of a symmetric system onto another part. These mappings conserve all information, which is why one part of an ordered system is sufficient to reconstruct the other parts.

A simple way to think about connections is our basic principle of the difference that makes a difference. The verb “make” implies a creative process. The new difference, although made out of the old one, is not the same. If it would be identical to the old one, there would not be two differences, there would be only one. On the other hand, the fact that it is made out of the old one, implies that something is carried over from the old to the new: there is a continuity or transfer of information. In that sense, the difference that makes a difference is perhaps the simplest description of an intrinsically complex phenomenon, characterized by both distinction (novelty) and connection (continuity).

For example, knowing that an English word starts with the letter “t”, you can infer that the next letter may well be an “h” or an “o”, but almost certainly is not a “k” or a “p”. However, you can’t predict which letter precisely will follow. The ordering of letters in English is neither repetitive nor random. The underlying arrangement is very complex. As we discussed in the section on information theory, there is a lot of redundancy in
English, which can be used to compress messages by recoding them to a more efficient system. Yet, although it is easy to compress English a little, it gets more and more difficult to eliminate all the redundancies. This illustrates my point that is impossible to find the shortest possible description. Typical compressions will reduce the length of a message with one half, which is not very much given how much redundancy there really is. The reason for this difficulty is that the dependencies in the English language exist on many levels, not only between single letters but between letter pairs, triplets, syllables, word parts, words, sentences, and even between paragraphs, sections, chapters and books.

This presence of structure at different levels or scales is another fundamental property of complexity. One of the basic ideas of systems theory is that you can consider either the individual parts of a system, or the whole they belong to. This is similar to a change of scale. A scale transformation will magnify or reduce the phenomenon you are looking at. It is as if you either zoom in and see all tiny details under a microscope, or zoom out and see merely the rough outline from afar. Like other transformations, such as reflections or rotations, a scale transformation determines a particular kind of symmetry. A phenomenon is scale-invariant or scale-symmetric if it looks the same after a magnification or reduction. For example, when you study a fern leaf from close, you will see that the sections of the leaf have exactly the same structure as the whole leaf (see fig.). In turn, the subsections of these leaf sections look again the same as the section, and so on, down to the smallest visible leaf parts. Similarly, a branch of a tree will often look similar to the tree itself, and the side branches look similar to the main branches. Another example is a rugged mountain landscape, where the large hills and peaks appear essentially similar to the smaller hillocks and rocky outcrops that are part of them.
Such “self-similar” patterns, where the parts have the same structure as the whole, are called fractals. According to our definition, they are ordered along the scale dimension. That means that they present a series of structures at different levels, from the trunk to the largest branches, subbranches, subsubbranches, etc., but that you only need to know one of these structures in order to reconstruct the whole. Therefore, a fractal shape is not very complex. Yet, it is more complex than a shape which has only a single level of structure, such as a circle or a square. You might say that such shapes have a “narrow scale”: they extend only over one scale of magnification. Zoom out and the circle becomes a dot and disappears from view. Zoom in, and you see only the empty space in the middle of the circle. A fractal, on the other hand, has a “wide scale”: however you zoom in or zoom out, you will always see the same recurrent pattern.

To find real complexity along the scale dimension, you need to imagine a system that has wide scale, but where the patterns that emerge while you magnify it change constantly. The human body may serve as an example. When you zoom in with your microscope you will first see organs and muscles, then tissues, cells, organelles, polymers, monomers, atoms, nuclei and finally elementary particles. All these levels have their characteristic structures, but they are all different. Still, they are not independent. The structure of a molecule, for example, is largely determined by the properties of the atoms that make it up. Yet, as we saw when we discussed the concept of “downward causation”, the whole is not completely determined by the parts. The same collection of atoms can be arranged in different ways to form different molecules.

Thus, we see that the whole and the parts, that is, the structures at different scales, are to some degree dependent on each other, to some degree independent. This is precisely how we defined complexity: connection, but not complete determination or repetition. It is because of this relative independence of wholes and parts that the analytic method, which separates parts out from wholes, often works. Similarly, it explains why the synthetic method, which ignores the parts in order to look at the wholes, is useful. More generally, it shows why the systems approach, which considers a system at a given level while taking into account both the supersystem at the level above and the subsystems at the level below, is better than either. The essence of the systems approach is that you first distinguish a system, separating it from its surroundings, then connect it back to those surroundings through its input and output. The multilevel approach we discussed is merely an application of this method along the scale dimension: first distinguish the system at the level you are interested in, then connect it back to the level below (subsystems) and the level above (supersystem) to better understand how it functions.
Complexity is the basis of existence. As we saw, distinction without connection (disorder) and connection without distinction (order) are in the limit just emptiness. Unlike order and disorder, however, complexity does not know any limits. Although there is only one way in which you can make a system completely ordered or completely disordered, there is an infinite number of ways in which you can make it more complex. You simply need to add distinctions and connections in different places and at different levels. Depending on where you add them, you will get different types of complexity. For example, you might make a car more complex by connecting it to an electrical grid, thus turning it into a trolley. Or, you might add an automatic braking system, or a second pair of doors. These types are mutually incomparable: in general, you cannot say that the one is intrinsically more complex than the other. That is the reason why it is impossible to find a universal measure for complexity. Complexity can increase without bound in different directions. There is no limit of maximal complexity. That is one reason why a complex universe is intrinsically more interesting than an ordered or disordered universe.

Although different systems can be ordered according to their degree of complexity, that order will not be complete, but partial. That means that for two arbitrary systems A and B, you might find that A is more complex than B, or that B is more complex than A. In general, though, you would not be able to compare A and B’s complexity. For example, you could argue that a modern propeller plane is more complex than the first airplane built by the brothers Wright a century ago, since it contains many additional parts and connections between parts. However, you cannot really determine whether a propeller plane is more complex than a jet plane, a helicopter or a ship, since these are organized in a totally different way, with different components performing different functions. You cannot just start to count the number of elementary components, such as screws and bolts, and decide that the machine with the largest number of bolts is the most complex one. The problem is that the whole is more than the sum of its parts. To determine overall complexity, you need to take into account interlocking distinctions and connections at many different levels and scales. Adding those together would be like adding apples and pears: you don’t get a number that is independent from the kind of things you add. It is as if you would have to answer the question: Are five apples and one pear more or less than three mice and two corkscrews?

In conclusion, our definition of complexity as distinction and connection implies that you cannot measure the overall complexity of a system. However, it does give you a method to determine whether a given system has increased or decreased in complexity. You increase complexity by adding either distinctions or connections (or both). The first process, which creates more differences between parts or aspects, may be called
differentiation. The second process, which creates more connections or dependencies between parts, thus making them more coherent, may be called integration. Complexification, the process that increases complexity, is then simply a combination of differentiation and integration.

Our reasoning has shown why the universe must be complex if it is to exist at all. However, it does not tell us how much complexity there is, or whether there is a tendency for complexity to increase or decrease. We could easily imagine a world that is much less complex than the one in which we live. In the next chapters we will start to explore where all the complexity we see around us may have come from.

Bootstrapping the Universe

An important implication of our analysis of complexity is that it leads to a relational philosophy. In a complex world, everything is connected to something else. Nothing exists on its own, separated from the rest. If a system S was disconnected from the rest of the universe, then that rest of the universe, including us, would never be able to observe it. S would not have any effect on anything in our part of the universe, and we would not be able to influence S in any way. Therefore, it would be the same as if S did not exist. Perhaps, we might say that S exists in another, “parallel” universe. We could imagine an infinite number of such parallel universes, containing the most bizarre phenomena. The only feature they would have in common would be that they are totally unreachable from our universe. This is merely another formulation of the defining principle of a “universe”, namely that it contains everything we could ever hope to come in contact with.

This connection with the rest of the universe is what distinguishes something that exists, that has a place in our universe, from something that does not exist. Thus, existence is a relation. Existence is not an intrinsic property of an entity. It is the relation an entity has with all other existing entities. Separate an entity from the rest of the universe (assuming that this is possible), and it ceases to exist. More generally, all properties are relational. Indeed, a property can always be expressed as a distinction between the phenomena that have the property and those that don’t. And the principle of the difference that makes a difference tells us that distinctions can only exist in relation to other distinctions. No distinction can exist on its own. Therefore, existence, properties, entities, complexity, even the universe itself, only make sense as relations.

This principle was at the base of the philosophy of Leibniz, Newton’s rival. It explains why he was so critical of Newton’s idea of an absolute space and time, which
exist independently of any things they contain. Leibniz developed a relational picture, in which the universe consists of “monads”, rather than atoms. Monads exist only by virtue of the way they are connected to the other monads. However, unlike Newton’s theory of mechanics, Leibniz’s theory of monads did not produce any concrete predictions. It has therefore been mostly forgotten. My task here will be to show how 20th century scientific ideas can help us to turn Leibniz’s relational philosophy into a more practical framework.

One reason why relational philosophies have not been of much use in science until now is that they tend to lead to a naive holism. The idea that everything is connected to everything else seems to imply that it is pointless to distinguish or isolate objects of study. But in order to cope with complexity, you need to take things apart and temporarily forget about everything except the few components you wish to study. One lesson of systems theory is that you can isolate a system by drawing a boundary around it, thus separating it from its environment. The reason you can do that is because, in complexity, distinction is as important as connection. (Remember, without distinction you get the emptiness of complete order.) However, you should always remember that the system is part of a larger whole, with which it interacts through its input and output channels. This defines the system as a relation, connecting input to output. You can choose to either focus on this relation (like in the black box view), or on the system’s components (like in the white box view). The two views are complementary, and support each other. Thus, the focus on distinctions saves our relational philosophy from simple-minded holism.

Another reason for the failure of relational ideas is mathematics. One advantage of Newton’s world view is that it can be easily expressed in a mathematical form. Atoms and their properties fit naturally into the mathematics of elements and sets. Absolute space, time and movement are described elegantly by geometry and calculus. Leibniz’s theory of monads did not have such a neat mathematical expression. The foundation of modern mathematics, set theory, is inherently non-relational. It is based on independent elements, and sets of such elements. Relations are added only as an afterthought: while relations can only be defined in terms of elements, the elements exist without prior need for relations.

This does not mean that a relational world view must lack the rigor and precision of mathematics. Some recent mathematical developments, such as category theory or the theory of hypersets, have a definitely more relational flavor. Although this is merely a first sketch, I have shown in some of my scientific papers how a relational view can be expressed in a mathematical form. Such a relational mathematics can be based on what I have called the bootstrapping axiom. An axiom is a primitive assumption underlying a
mathematical theory. Axioms determine the fundamental properties of the elements of the theory. Together with the definitions, which introduce new types of elements, they allow a mathematician to deduce theorems. Theorems express different regularities that govern the elements. For example, I have shown how the few simple axioms and definitions of my relational framework allow you to deduce the basic space-time structure of relativity theory, with its light cone and causal connections.

I will here not go into details about the axioms, definitions and theorems of my relational theory. I’d rather try to convey the underlying ideas. The concept of bootstrapping derives from a story by the legendary Baron von Munchhausen, who supposedly had the most incredible adventures. In one of these adventures, the baron, while walking through a swamp, got stuck with his feet in the mud. He wanted to pull himself loose, but found nothing to grasp or to support him, neither branch nor stone. Thus he sank deeper and deeper. He then had the brilliant idea to pull on his own bootstraps, and managed to lift himself out of the mud. A quick try will convince you that the idea of lifting yourself up by your bootstraps is as preposterous as any of the baron’s tales. Newton’s laws of mechanics tell us that any action, such as your bootstraps being pulled up, must be accompanied by an equal reaction, such as your shoulders being pulled down by the weight of your boots. The net effect is zero: you remain in the same place. To raise yourself, you need something to push against: a support or foundation. Similarly, if you want to raise a building, you need a foundation to build upon.

The same principle seems to apply if you build a theory. The theory must be grounded firmly. On that foundation, you can build complicated descriptions and explanations. The foundation itself, however, is not built on something else: it just is there. The concepts of absolute space, time, matter, and force, form the foundation of Newton’s mechanics. You can explain complicated objects, movements and interactions in terms of these fundamental components. However, within the Newtonian world view you cannot explain where these components have come from. They are simply postulated. You have to accept them at face value. In spite of all the advances since Newton, this problem still besets physicists. Although quantum field theories have shown that matter and force can both be reduced to elementary particles, they still take space and time for granted. Therefore, they cannot explain the origin of space and time. The Big Bang, the creation of the universe, must remain a mystery.

The same limitation applies to elementary particles. Physicists explain how the matter we see around us get its properties from the specific combinations of particles that compose it. However, they cannot explain where those particles get their properties from. Since the particles are by definition elementary, you cannot account for their features by
reducing them to their components. Each particle’s features are uniquely its own. Therefore, the more elementary particles you discover, the more features you need to explain. In the 1950’s, this problem became more and more pressing as the list of known particles grew to several dozen.

In the 1960’s, the physicist Geoffrey Chew proposed a way out of this predicament. He suggested that particles do not consist of still more elementary parts, but out of combinations of each other. For example, suppose that you have three types of particles, A, B and C. Then, an A particle might be composed out of one B and one C, a B would consist of an A and a C, and a C would consist of an A and B. Chew called this the bootstrapping hypothesis, in reference to von Munchhausen’s trick. Indeed, explaining one particle in terms of another seems equivalent to keeping yourself supported by pulling on your own bootstraps. The idea isn’t as silly as it looks, though. Indeed, understanding the properties of elementary particles means in the first place understanding how they differ from each other. If there was only one type of particle, there wouldn’t be much to explain. Properties, as we saw, are essentially distinctions. The bootstrapping hypothesis neatly accounts for the fact that the particles are distinct. Indeed, a combination of an A and a B is by definition different from a combination of an A and a C, or a B and a C. This was simply a more modern version of Leibniz’s theory of the monads, which get their identity from the way they reflect other monads.

In spite of its elegance, Chew’s model was only partly successful. Since then, it has been supplanted by other theories. This happened in part because the quark hypothesis allowed physicists to explain apparently elementary particles as combinations of still more elementary quarks. This reduced the number of particles needed. In the meantime, the number of quarks and related particles has again grown to worrying proportions, though, and further reductions are not in sight. Some features of Chew’s theory survive in the presently popular string theories, which see particles as different vibrations of a one-dimensional string. Although such theories may eventually explain the differences between particles, they still assume the independent existence of space and time. The need to fit the zoo of particles, interactions, and other weird quantum phenomena into the already complicated mathematics of space and time may explain why string theories are so frighteningly complex: it turns out that space-time must have at least 10 dimensions for strings to be able to exist!

There may be another way, though. In the relational picture, space and time emerge as relations between elementary processes or events. It may seem that these processes then take over the role of foundations. However, in my approach, these elementary processes have no intrinsic properties. Therefore, they don’t need any
explanation! In other words, taken on their own, these processes cannot be distinguished. Distinctions only arise out of the pattern of connections that weaves these processes into a whole. This is expressed by the bootstrapping axiom. It simply says: a process can be distinguished only if the processes that it is connected to can be distinguished. Two processes that have exactly the same connections must be identical. This is yet another formulation of “the identity of the indistinguishables”, or “the difference that makes a difference”. What is new is the focus on mutual connection: a process A is defined in part by the process B to which it is connected, while B is defined in part by A. This is an extension of Chew’s bootstrap hypothesis. The practical use of this axiom is the following. Given a network of connected processes, it allows you to simplify the network by merging all indistinguishable processes. On a higher level, this network then allows you to distinguish more complex structures, such as the light cone that defines relativistic space time. By considering ever higher, more abstract levels, you can hope to reconstruct all fundamental types of systems that science has discovered.

At present, this relational reconstruction of space and time is merely a proof of concept, not a real theory. It does not say anything about the different particles that exist, though it in principle allows you to distinguish different particle types. It is presently not my ambition to develop this framework into a competitor for string theory—although I would be very happy if someone else did. The point here is just to show that a bootstrapping philosophy can be formulated mathematically, and therefore could possibly replace the relics of Newtonian mechanics that are still at the core of particle physics. My broader ambition is to apply this relational approach to the whole of reality, not just to elementary particles, and show how it can help us to build an integrated world view. To start that program, we first need to understand change and evolution in a relational manner.

**Causality: micro or macro?**

Until now, we have discussed distinctions and connections without any reference to change. Therefore, our analysis of complexity remained general, applying to static as well as to dynamic situations. If we wish to understand how complexity evolves, though, we will have to look more explicitly at change. The Newtonian world view reduces all change to causation: a process that conserves distinctions. As we noted, distinction conservation merely exemplifies the principle of the “difference that makes a difference”. More precisely, consider two causes or two different initial situations. Suppose that they
would lead to identically the same effect. In that case, from the point of view of that final effect, the original causes can no longer be distinguished. Ockham’s razor, or Leibniz’s principle of the identity of the indistinguishables, therefore compels us to conclude that these causes must be considered identical as well. There is no sense in distinguishing them, since whatever distinction we make has become unobservable.

This conclusion may seem weird. After all, what if those two causes were really different? How can we simply define that difference away? An example may illustrate the subtleties of the argument. Remember the box with two compartments, which we used to explain the second law of thermodynamics. One compartment contains a gas, the other one is empty. When we open a door between the two compartments, the gas diffuses until it evenly fills both. Now, in this situation, it does not make a difference whether the compartment that was originally empty was the left one or the right one. Therefore, our relational principle would tell us that there really never was a difference between the left and the right compartment. “But I saw with my own eyes that only the left one was full!”, you might object. If that is true, the identity of the indistinguishables does not apply. Indeed, seeing a full compartment means that some signal coming from the left box was picked up by your brain and stored in your memory. Therefore, the difference between left and right did make a difference in your memory, and thus there is distinction conservation.

But what if no one was present when the boxes were filled? Most likely, there still remain traces of the original situation. For example, the filled box would have been a little warmer, and therefore it would have radiated more heat. If we trace back the radiation emitted by the two boxes, we could reconstruct which box was initially empty. Even if we assume that the two boxes were perfectly insulated, so that no radiation could escape, it is theoretically possible to trace back the difference. Indeed, if we would measure the precise trajectories of every single gas molecule in the box, we could use the laws of Newton to reverse all those trajectories, and thus calculate which box any molecule initially came from. In conclusion, if in principle there was a difference in initial situations that difference can still somehow be found in the final situation. If there is no difference at the end, it means there cannot have been a difference to start with. That is the way it should be, according to both Newton and Leibniz.

This reasoning is clearly not realistic, though. In practice, it is impossible to reverse the trajectories of all molecules, or to trace back heat radiation. The different non-classical theories—thermodynamics, relativity theory, quantum mechanics and chaos—would all agree: given only the final situation we could never gather the information needed to reconstruct the initial situation. Similarly, given the initial situation we can
never gather all information needed to predict the final situation. The perfect predictability and reversibility of Newtonian causality only works in an ideal realm, where everything is known about everything.

Since it implies that we should know all the tiniest details about the tiniest particles, I will call this interpretation microscopic causality. In contrast, when we reason about the things around us, and try to make predictions, we assume the existence of macroscopic causality. Macroscopic causality means the conservation of large-scale, observable distinctions, independently of what happens to all the invisible, microscopic differences.

Microscopic causality says that two different initial states for the universe can only give rise to two different final states for the universe. However, such a “state of the universe” includes everything that could ever be known, including the movement of a butterfly’s wings on the other side of the ocean, the position of a rock on the moon, the fleeting thought that passes through your loved one’s mind and even the precise quantum state of an atom somewhere in the Andromeda galaxy. Since every part of the universe is directly or indirectly connected to every other part, all these phenomena matter if you wish to predict the future state of the universe. In conclusion, if we believe in the identity of the indistinguishables, then the principle of microscopic causality must be true. At the same time, the principle is totally useless, since you cannot derive any prediction whatsoever from it. In conclusion, microscopic causality is a tautology, an empty, trivial truth.

The only reason that Newtonian mechanics could be based on this principle, is because it assumes that you can completely separate a system from its surroundings, so that you don’t have to care about Andromeda or your partner’s mood if you want to calculate the trajectory of a planet. Quantum mechanics and systems theory, among others, have shown that such perfect isolation is in principle impossible. This leaves us to explain why causality nevertheless often works. In other words, how does macroscopic causality emerge from microscopic causality? At the same time, we will have to formulate an alternative principle, for those cases where causality clearly does not work.

Some examples will clarify the issue. Consider a coin standing vertically on its edge. This is a highly unstable position, which won’t last. You cannot predict to which side the coin will fall, though. When it falls, you will not look for the cause of it falling down this side rather than that side, but consider it a chance event. The reason is that the coin is so sensitive to initial conditions that the tiniest fluctuation will affect it: a little bit of wind, a vibration of the table, an irregularity in the surface, or even a random distribution of air molecules, so that a few more molecules bump against one side rather
than against the other. The coin obeys the microscopic causality principle: the slightest difference in initial conditions makes a difference in results. However, on the macroscopic level you cannot find a causal relation between initial and final state.

Consider now a heavy desk. No amount of wind, vibration, irregularity in the surface or chance concentration of air molecules will affect its position. The only thing that will make it move is a strong and persistent force. If you come back to your office, and you see that your desk has moved, you will immediately start to look for a cause, and probably assume that, in the absence of earthquakes, some person has been pushing it. On the other hand, you can predict that if you push sufficiently hard against your desk, it will move in the direction of your push. Now, you are using the principle of macroscopic causality: clearly distinguishable causes have clearly distinguishable effects. Here you apply causality as a practical principle to explain or predict events.

To show that this argument is not limited to physical systems, I can add a psychological example. Suppose that you have a friend with a calm, balanced character. If she gets upset, you will assume that there must be a clear, recognizable cause for her anger: perhaps somebody offended or deceived her. Now, imagine a neurotic, mentally disturbed personality. Such a person can get upset by the most complicated combinations of the most subtle factors. Both the number of factors (the weather, what he dreamed last night, which color of shirt you are wearing, whether a passer-by smiled or not, ...), and the sensitivity to each single factor will be so large that in practice you won’t be able to explain or predict his bursts of anger. In practice, you would consider such a person “crazy”, “irrational” or “insane”, because he does not behave in a predictable, causal manner. Still, you could argue that his behavior is not random, but governed by the microscopic causality principle.

To clarify the difference between these two types of situations, we can write down the microscopic causality principle with symbols: $C \rightarrow E$ (cause $C$ produces effect $E$). $C$ and $E$, as we noted, represent the complete state of the universe, respectively before and after the process. Let us assume that we have a (very) partial knowledge of that state of the universe. For example, we know a few features of the situation, such as the position of the desk and the force with which it is being pushed. Let us represent these known features by the letter $K$. However, we don’t know any other features, such as the movements of air molecules or the microscopic irregularities of the floor. Let us call the whole of these unknown features $B$, for “background conditions”. The microscopic causality principle can then be written more precisely as: $K + B \rightarrow K' + B'$ (where the prime (‘) distinguishes the situation after the causal process from the one before).
Macroscopic causality then means that you can forget about the background conditions, and limit yourself to those features that you know or can observe: \( K \to K' \). This is possible in the case of the desk, because the background conditions in practice don’t affect the features \( K' \) that you are interested in, such as the position of the desk. In the case of the coin, though, the background conditions will always interfere, and \( B \) cannot be eliminated from the reaction \( K + B \to K' + B' \). As long as the unknown \( B \) remains part of it, this formula is totally useless. Since we cannot distinguish the components of \( B \), there is no conservation of the distinctions that we can observe.

That does not mean that we cannot say anything meaningful about the coin falling on its side, though. Let us look in more detail at what happens. Different background conditions—such as the random movements of air molecules—will affect the coin differently. It can move to the right or to the left, and it can start to move immediately or after a few seconds. Many possibilities open up. A single known cause, the coin standing on its edge, leads to different observable effects. Eventually, though, the coin will end up lying flat. In fact, there are only two final positions: head or tail. The precise movement in between the vertical and horizontal positions does not affect the outcome. We thus see two phases in the process affecting the coin: first, the number of possible states or trajectories increases, then it decreases again, leaving only two. Thus, we can make a non-trivial prediction: the coin will come to rest horizontally. We just cannot say which side will be up.

In fact, this two-phase description of a non-causal process is perfectly general. Consider any phenomenon where distinctions are not conserved. Non-conservation of distinctions can mean two things: new distinctions can be created, or existing distinctions can be erased. In general, there will be both creation and destruction of distinctions. Distinction creation means that the trajectory branches, that one possibility gives rise to several (see fig.). Distinction destruction means that different trajectories come together, that several possibilities are reduced to one. Since distinctions cannot be eliminated before they have been created, the natural sequence for a process that starts from a single situation is the following: creation of distinctions, followed by the destruction of part or all of these distinctions. Of course, a complex, extended process will have many such subprocesses, where some distinctions are created, then some of them are destroyed, then others are again created, etc.

This is the most general description of a non-Newtonian process. Our next task will be to formulate a principle that can help us determine which distinctions will be created, and which ones will be eliminated. This principle should provide us with a more
realistic alternative for the microscopic causality principle. At the same time, it should tell us when macroscopic causality is applicable.
The Trial-and-Error Method

Variation and Selection

What we called creation and destruction of distinctions can be described even more simply. Creation of distinctions means that out of a single starting situation you produce a range of different possibilities or variants. This is merely what Darwin would have called “variation”. Destruction of distinctions means that you eliminate a number of these variants, and thus reduce the range of remaining possibilities. This is what Darwin would have called “selection”. The concepts of variation and selection, as we saw, were introduced to explain the evolution of species. As such, they are associated with a range of biological phenomena, such as inheritance, sexual reproduction, competition between species, adaptation to the natural environment, and so on. Yet, we saw that these same concepts have been applied to an ever expanding list of non-biological phenomena, from molecule design to computer algorithms, and from scientific discovery to business competition. As such, the concepts of variation and selection have acquired a more general meaning, losing most of their purely biological connotations. What I propose here is to set one step further in this process of generalization, and to use variation and selection in their widest possible sense, as the two fundamental phases of distinction creation and distinction destruction.

What name we give to a phenomenon is not so important. Whether we call something “distinction creation” or “variation” does not really matter, as long as we agree on the name. What is important, though, is to understand the phenomenon well. It is here that the theory of evolution can help. In biological evolution, variation is produced basically by mutations: small random changes in the DNA molecules that carry an organism’s genetic information. These mutations are caused by a variety of uncontrollable factors, such as errors in the copying of DNA strings, cosmic rays breaking chemical bonds, chemical reactions involving foreign chemicals or waste products of the cell, or simply random movements of the molecules. These causes belong to what we have called the “background conditions”, when we discussed the microscopic causality principle. These influences are so difficult to observe and so diverse that we can
never come to know all of them. That is why we consider variation to be essentially random or unpredictable.

Yet, variation is not arbitrary, either. What all mutations have in common is that they are small. The genes of a cat will not suddenly turn into the genes of dog, a horse or an oak tree. Only a few atoms will be affected by each mutation. To produce a major change, like the evolution from sea-dwelling fish to land-dwelling reptile, you need a very long sequence of such small mutations. This gradualness of evolutionary change in fact expresses a more universal principle, that of continuity. No change can be instantaneous. The larger the change, the more time and effort you need to bring it about.

This principle is already apparent in Newton’s law of inertia: in order to change the movement of an object you need to apply a force. The larger the change (or the heavier the object), the larger the force you need. Still, you might think that if you would apply an infinite force, you could blow an object from here to Andromeda in zero time. Apart from the fact that infinite force requires infinite energy, this scenario is forbidden by the theory of relativity. According to relativity theory, nothing can move faster than the speed of light. By applying an infinite force, you will only manage to make the object move with the speed of light, not faster. To move an object to Andromeda, this means that you will need at least a few million years, whatever your energy supply.

The impossibility of infinite speed is neatly captured by the concept of the light cone: all movements necessarily take place within the light cone; they can never go beyond. The light cone is perhaps the most fundamental expression of continuity. You don’t need Einstein’s theory to explain continuity, though. Continuity follows from our basic idea of the difference that makes a difference. It expresses connection, which is a defining feature of existence. Unconnected events are like the virtual particles which spontaneously appear and disappear in the quantum void: they are too ephemeral to be seen or manipulated. That is why they can ignore physical laws, such as the conservation of energy or the restriction to movement inside the light cone. But that is also why we call them virtual, and distinguish them from real or existing events.

The light cone can also be seen as an expression of freedom: as the realm of all possible movements. The typical shape of the light cone, starting in a single point, then fanning out, becoming wider and wider, illustrates how the possibilities for change increase as time goes by. You can see the same structure in a non-causal process during the phase of distinction creation (see Fig. 1): the longer the time that has passed, the larger the number of distinct possibilities. This is merely another way of formulating the second law of thermodynamics: the longer you wait, the more uncertain the state of the system becomes, and therefore the higher its entropy.
In biology, continuing variation without selection is called “drift”. Its effect is similar to the effect of wind and waves on a bottle thrown into the ocean: the bottle moves aimlessly, now going in one direction then in another, while slowly getting further away from the place it started. The only thing you can be sure of is that the longer the bottle has been drifting, the more difficult it will be to find it. Thus, “drift” is just another word for small, unpredictable changes, which gradually increase your uncertainty.

The concept of variation has helped us to understand distinction creation, and its characteristics of unpredictability, continuity and entropy increase. Similarly, the concept of selection may help us to grasp the properties of distinction destruction. In biological evolution, selection is carried out by the environment in which the organism lives. The factors in the environment that determine whether the organism will survive have nothing to do with the factors that produce mutations inside the cells. Therefore, variation and selection are independent. Even if you knew which mutations would take place, you would not be able to predict which of the resulting variations would survive. In the more likely case where you would know what features an organism must have to survive in a given environment, you would not be able to predict which of those features will be produced by variation. It is this separation between variation and selection which makes evolution both unpredictable and inherently creative.

This applies to non-causal processes as well: in general, the phases of distinction creation and distinction destruction are independent. “In general” does not mean “always”, though. Sometimes, they exactly counterbalance each other. This leads us back to the special case of macroscopic causality, where a single initial state leads predictably to a single final state, irrespective of any intermediate states that may have appeared during the process. Let us illustrate this situation with an example from quantum mechanics.

The state of a quantum particle, for example an electron, can be represented by a wave function. The wave function determines the probability that the particle will be found in a particular place. The more dense the wave in a particular region of space, the more likely it is that you will find the particle in that region. Suppose that you start with an electron localized in a particular point. The moment you let the electron evolve on its own, its wave function will start to spread, like the expanding wave created by a stone thrown into a pool. This means that the number of possible positions for the electron increases as time goes by. This is the phase of distinction creation, characterized by increasing uncertainty. Suppose now that you make a detection of the electron’s position, for example by letting it pass through a screen which lights up in the precise spot where it was hit by the electron. The moment the electron hits the screen, the number of
possibilities for the position is reduced to one, the spot on the screen. Physicists call this event “the collapse of the wave function”. It corresponds to the phase of distinction destruction: one position is selected, all others are eliminated. This selection is largely independent of the variation which created the different possibilities. That is why the positions of particles in quantum mechanics are basically unpredictable.

However, not all possible positions have the same probability. In fact, you can look at the spreading out of the wave function as if the electron’s trajectory would be diverging. Instead of a single trajectory, like in classical mechanics, a quantum system follows many parallel paths or trajectories at once. The spreading wave function can be seen as a superposition of all these parallel trajectories. (this is the “path integral” formulation of quantum mechanics). From those different trajectories, the ones that are closest to the classical trajectory are the ones most likely to be selected when you make an observation. The spread of trajectories around this classical trajectory determines your uncertainty about the electron’s position: the wider the spread, the more difficult it is to predict where the electron will hit the screen. This uncertainty or spread depends on the size of a quantum system: the larger or the heavier the system, the smaller the divergence from the classical trajectory. In the so-called “classical limit” of a macroscopic object (say a planet or a football), there is no perceptible uncertainty left. This means that the different possible trajectories in practice reduce to a single trajectory. Thus, the distinctions created by variation are completely eliminated by selection, leaving only one possibility. We are back at macroscopic causality, and reliable predictions.

This was an example of how selection completely suppresses the effects of variation, so that no variation is left. In general, though, selection will be independent of variation, sometimes suppressing it, sometimes reinforcing it. The selection of internal mutations by the external environment illustrates this independence. Yet, we don’t need to assume that variation happens inside the evolving system and selection outside of it, as Darwin did. The same principles apply to external variation and to internal selection, as we will discuss later. But what principles govern selection then? To better understand selection, we need to discuss the irreversibility of change.

The Ratchet Effect

What is the difference between variants that are selected and variants that are eliminated? At the most basic level, the difference is obvious: selected variants are retained as such, eliminated variants are replaced by something different. Natural selection in this sense is merely the observation that some things change, while others don’t. Although this may
seem trivial, the distinction between change and persistence is crucial to our understanding of evolution. The fact that there is a difference between a statement and its negation—between A and not A—seems trivial as well. Yet, the whole of logic and mathematics is built on this distinction. Similarly, we may build a “logic of evolution” on the distinction between stability and change.

But what is so new about that? Newton’s mechanics already described that difference very well. A stone can be in equilibrium on the ground, or falling down. A comet can revolve around the sun in a stable orbit, or it can enter the solar system, hurtle past the Earth, and disappear in the interstellar void, never to be seen again. The problem is that the two types of trajectories described by Newtonian mechanics, stable and changing, always remain separate. The one can’t turn into the other. However weird this may seem, if we take Newton’s laws literally, the stone that falls down can never reach an equilibrium. It will bounce up and down like a yo-yo until the end of time. That’s in a way what happens with a satellite orbiting the Earth: it is forever falling, and will never come to rest. The reason for this counter-intuitive prediction is that Newtonian mechanics is fundamentally reversible: every change can be undone. If movement can turn into equilibrium, then equilibrium should be able to turn into movement. But the definition of equilibrium is precisely that it does not lead to movement. The stone lying on the ground will not spontaneously jump up. If we believe that stability or equilibrium exists, in other words that a system in a particular state will remain in that state, then mechanics forces us to conclude that no system could reach that state without already being there. In other words, stable systems remain stable, and changing systems remain changing.

This conclusion is clearly unrealistic. The fact that falling stones do come to rest makes it clear that we have to abandon the principle of reversibility. Different non-classical phenomena—entropy increase in thermodynamics, black holes in general relativity and attractors in chaos theory—all illustrate how, in general, change cannot be reversed. In contradiction with Newton’s mechanics, it turns out that the future is essentially different from the past. As we saw, the increasing number of possibilities associated with distinction creation may be sufficient to explain entropy increase in thermodynamics. In order to explain irreversibility in general, though, variation is not sufficient.

An attractor (including a black hole) is defined by the fact that you can enter it, but not leave it. As such, it generalizes what we called equilibrium or stability. You can enter an attractor from many different positions (which together form its basin), but once you are in, the number of positions you can reach is limited to those inside the attractor. Therefore, entering an attractor reduces the number of possibilities. In other words, it
destroys distinctions. Reaching an attractor is just another description of what we have called *selection*: the positions inside the attractor are retained, all the others are eliminated. Thus, selection is possible because there exist attractors: states you can enter but not leave.

In a way, the attractor states are privileged: they have an edge over the other states. That is why selection happens naturally or spontaneously: it is sufficient that a system ventures into an attractor state for it to be retained there. There is no need for an external agent, such as a breeder or designer, to make choices. The attractor states are automatically separated from the non-attractor states, without any planning or foresight. The mere existence of a difference between stable and unstable is sufficient for natural selection, as it leads to a long-term weeding out of the unstable states.

Although the attractor concept derives from chaos theory, it describes biological evolution just as well. Darwin’s natural selection is based on the survival of the fittest. Imagine a species of plants or animals undergoing different variations. Some of these variants are “fit”: they survive and reproduce without further change. Others are not: either they die out, or they undergo further mutations until they happen to chance on a fit configuration. The fit configuration functions as an attractor: a species can evolve towards it, but not away from it.

Let us try to understand more precisely how such selection works. The “black holes” formed by gravitation suggest an analogy: an attractor is like a hole in the ground. When you come too close, you start to slide into it, until you reach the bottom. Once there, you may perhaps move about a little, but the walls are too steep to climb out, and you are stuck. Both physicists and biologists use this metaphor to describe irreversible processes. Imagine a landscape with mountains and valleys. Because of gravitation, an object such as a ball will roll downwards, away from the peaks and into the valley. Once it has reached the lowest point, it stops moving. It now is in an attractor (see Fig.). The height of a position is a measure of its stability: the lower the position, the less likely that the ball will continue rolling.
When we describe processes other than balls rolling down hills, we need a factor more general than height to understand what will happen. In physics, this factor is called potential energy. In the case of a ball, potential energy is proportional to height: the higher the ball, the higher the speed it will reach by moving downwards, and therefore the higher the energy of its movement. Potential energy can be defined and calculated for much more complicated situations, which have nothing to do with height above the ground. For example, a gas that is compressed has a high potential energy. Like all physical systems, it will evolve to a state with less potential energy, for example by expanding—using its energy to push a piston.

All physical systems try to minimize their potential energy. Yet, we know from physics that energy is always conserved! This can be explained by dissipation: the excess energy is simply lost into the environment. In the case of the ball rolling down the hill, the potential energy is converted into the energy of movement. That movement is slowed down by friction, which makes the ball stop eventually. Friction means that when the ball rubs against a rough surface the energy of movement is turned into heat. The heat simply diffuses into the surface and the air. In physics, energy is defined as the capacity to perform “work”. More generally, we can define it as the capacity to produce changes. Systems with such capacity will sooner or later use it: otherwise it does not make sense to speak of a “capacity”. However, by exerting that capacity, they lose it, since the energy dissipates by usage. Thus, physical systems tend to change their environment, while becoming themselves more stable.

Fig.: a fitness landscape with peaks and valleys: the arrows denote the directions in which a system will move. The valleys A, B and C correspond to local minima or attractors where the system will come to rest. The height of a position corresponds to its potential energy or to its lack of fitness. Thus, A has a higher fitness (or lower energy) than B.
An example from quantum mechanics will show us in more detail what happens with the energy in such an irreversible process. A quantum system, such as an atom, can exist at different levels of energy. Unlike the energy of a ball, the energy of an atom can only have a few, discrete values, with no values in between. (The energy landscape for an atom would look like a staircase, not like gently rolling hills and valleys.) Like all physical systems, an atom at a high energy level tends to go back to a lower energy level. When it makes such a “quantum jump” to another level, it emits the excess energy in the form of a photon—a particle of light. That photon may be absorbed by another atom, making it jump up to a higher level. More likely, though, the photon will disappear into the depths of space, never to be seen again. Since the universe is expanding, the photon may fly forever and never encounter another atom. Thus, the universe provides an infinite sink, always ready to absorb the waste energy produced by irreversible processes!

The idea of a landscape with peaks and valleys to describe irreversible changes extends far beyond physics. In biology, an organism does not try to minimize potential energy, but to maximize fitness. The low points of the landscape, to which everything moves, then correspond to high fitnesses. If we want to keep the image of the ball rolling down the hill, we will have to equate height with negative fitness or lack of fitness. In biology, though, people tend to equate height with positive fitness. This would mean that the system tries to climb up to the peak, rather than roll down into the valley. Remember our discussion of artificial intelligence. There too, problem-solving was viewed as “hill-climbing”, that is, trying to reach the highest point in the problem landscape. This image is similar to the fitness landscapes used by biologists, but opposite to the energy landscapes used by physicists. Although this may seem confusing, the two conventions are equivalent. The one type of landscape is like the other one turned upside down, with peaks becoming valleys, and valleys becoming peaks. You should just remember which of the two conventions you use: climbing up or sliding down.

As we will argue later, fitness is a more general concept than potential energy. You can use it not only in biology but in any situation where there is a mechanism similar to the variation and selection of evolution, or the generate and test of problem-solving. Therefore, we will from now on only speak about fitness landscapes, and consider the energy landscapes as a special case of this general concept.

Although the movements in a fitness landscape are usually characterized as “irreversible”, that term is not very accurate. Indeed, a movement down an energy or a fitness valley can always be reversed. For example, the atom that emits a photon while falling down to a lower energy level, can always absorb another photon, and get the energy to jump back to the higher level. Similarly, the ball in the pit can always get a
kick, propelling it out of the hole again. The point is that such a movement upwards is much less likely than the movement downwards. For an atom to lose a photon, nothing special needs to happen. The photon is emitted spontaneously. For an atom to absorb a photon, on the other hand, the right type of photon needs to arrive at exactly the right spot. This is a quite unlikely event. Similarly, it is much more likely that a ball would spontaneously roll down a hill, than that it would be pushed up by a bored child or by a gust of wind. The inverse movement is not impossible, it is simply improbable.

Such movements are *asymmetric with respect to time*: one direction is more likely than the other. In this more general view, selection is due to the fact that some changes are more probable than their reverses. If the transition from state A to state B \((A \rightarrow B)\) is more likely than the inverse transition \((B \rightarrow A)\), then systems coming from A will tend to accumulate in state B. B is not strictly an attractor, since you can leave it, but it is a kind of “approximate” attractor: leaving it is more difficult than entering it. It is this asymmetry which again distinguishes the more “privileged” states, like B, from the less persistent ones, like A. However, this separation will be much more fuzzy than the distinction between attractors and non-attractors. Instead, we will find a gradation, where B may be more “attractive” than A, C more attractive than B, D more attractive than C, and so on. This is well expressed by a fitness landscape, in which valleys can be deeper or shallower. The deeper a valley (or the higher the surrounding hills), the less likely it is that a system will be able to escape from that valley, and therefore the more “attractive” the position at the bottom.

You should note, however, that the absolute height or fitness is less important than the height with respect to the surrounding territory. In Fig. there are many points in valley A which are lower than the lowest point in valley B. Yet, only the bottom of valley A will be effectively more “attractive” than the bottom of valley B. All points higher than the bottom will automatically lead to the bottom itself, and are therefore intrinsically unstable. The bottom of valley B is a so-called “local minimum” of the energy (or a local maximum of the fitness). This means that it is the lowest point with respect to its surroundings, but not the overall lowest. The overall lowest point is called the “global minimum”. Since it is difficult to get out of a valley, especially if its flanks are steep, systems tend to remain stuck in such a local minimum. They thus will never have the chance to reach the global minimum. In the next sections we will explore how we can help a system to get out of this rut, and discover deeper valleys.

Since movement upwards rarely happens and will any way be followed by downward movement, we might say that in the long term, a system can only move down. This is not a continuous movement, since each time the system reaches a local minimum
it will tend to stay there for a long period. The long-term change will rather look as if the system is moving down a staircase, pausing a while at each step, but then inexorably continuing its descent. Thus, in the long term, changes will only take place in one, privileged direction, the one of increasing fitness. We may call this the ratchet effect.

A ratchet is a mechanism, used in tools such as wrenches, that only allows movement in one direction. For example, when you pedal on a bicycle, you can only make the wheels move forward; pedalling in the other direction simply has no effect. Ratchets normally consist of a series of teeth or steps, which prevent movement going in the wrong direction. Consider a surface that looks like a somewhat slanted, horizontal staircase (see Fig.). A ball in position A cannot move to the left, because it is blocked by a steep, vertical wall. On the right, there is a gently inclined plane. A slight push to the right will make it cross into the next position B. From B, it cannot go back to A, but another slight push will make it cross to C. Thus, the ball will tend to continuously move to the right, in the direction of the arrow, without ever being able to go back.

![Fig. a schematic ratchet mechanism. Movement over the toothed surface is only possible in the direction of the arrow.](image)

Although a wrench may seem a purely artificial mechanism, we can find a similar structure in nature. Some seeds, such as those of rye, have a shape that looks like the inclined teeth of a ratchet (see Fig.). The difference is that there is not a ball or other object moving over the teeth, but rather the teeth themselves moving against a surface. Indeed, if you walk through a rye field, it is quite likely that one of these seeds will enter the tissue of your pull-over or trousers. With respect to the tissue, the toothed surface can move only in one direction, the one of the seed’s top. Thus, the seed will tend to penetrate ever deeper into your pull-over. It is not possible to get rid of it by pulling it out. Pulling againsts the direction of the teeth will only damage the tissue. The only way to remove the seed is to pull it forward, and make it pass completely through the tissue. The function of this strange contraption should be clear: by attaching themselves to clothes or the fur of animals, seeds like this can hitch a ride and find new grounds to germinate and grow.
Maxwell’s Demon

We have discussed the main properties of variation and selection. Variation is the unpredictable and continuous effect of unknown background factors. As such, it increases the number of possibilities and produces a growing uncertainty. Selection derives from the fact that changes are in general asymmetric: it is easier to go from an unstable state to a stable state than the other way around. That is why systems tend to concentrate in the regions of high stability or high fitness, the “attractors”. This spontaneous concentration is the opposite of the diffusion caused by variation. That is why selection decreases our uncertainty. The longer we wait, the more likely it becomes that the system we are interested in has settled into one of these attractors. If we know where the attractors are, this also means that the longer we wait, the better we know where the system is. Variation and selection are two complementary forces, the one increasing the number of possibilities, the other decreasing it.

At first sight, this conclusion seems to contradict the second law of thermodynamics. The second law says that—overall—entropy cannot decrease. In the interpretation of Boltzmann, entropy is the same as uncertainty. How can selection then decrease uncertainty? The way I see it, Boltzmann’s view is not as general as the original definition of entropy in terms of heat. An evolutionary process can reduce uncertainty, but it will still dissipate heat, and thus produce thermodynamic entropy.

Let me go back to my favorite example of the box with two compartments to illustrate the difference. If an opening is made in the wall separating the two compartments, the gas from the left compartment will irreversibly diffuse into the empty right compartment. Shortly after Boltzmann discussed this situation, the British physicist James Maxwell proposed the following counter-example. Suppose that next to the hole you post a guard, a microscopic “demon”, who is able to see individual molecules. When a molecule enters the hole from the left, the demon’s task is to stop it. When a molecule enters from the right, though, the demon lets it through. The demon performs a selection, deciding which molecules to allow where. With such an arrangement, all molecules from the right will sooner or later move to the left, but once on the left, they will have to stay.
In conclusion, instead of diffusing, the gas will concentrate in the left compartment. This contradicts Boltzmann’s interpretation of the second law.

Numerous scientific papers and even complete books have been devoted to Maxwell’s demon. Some have argued that such a demon could never exist. Others have proposed that if such a demon existed it would still need light to see the molecules, and the entropy created by the interaction of the light with the molecules would compensate for the decrease in entropy caused by separating the molecules. More sophisticated interpretations have considered the demon as an information-processing system. All interpretations seem to agree that the demon cannot decrease the overall thermodynamic entropy.

To clarify the issue, I propose to look at a simpler set-up, without an intelligent demon that observes and decides. We can just replace the demon by a little door that only opens in one direction. Physically, designing such a door is easy. We just need to imagine a large molecule that covers the hole. The molecule is attached to the wall in one point, which functions as a hinge (see fig.). When a gas molecule hits the door molecule from the right, the door swings open into the left compartment and lets the molecule pass. An attractive force, functioning like a spring, then pulls it back to its original “closed” position. In that position, it can only open to the left, as it is blocked by the wall on the right. The door molecule functions like a ratchet, allowing movement in one direction but not in the other. The net effect of the door is the same as that of the demon: gas molecules will concentrate into the left compartment, decreasing our uncertainty about where they are. The left compartment functions like an attractor, which the molecules can enter but not leave.
Although this thought experiment may appear paradoxical, we have transgressed no laws of physics. Energy is still being dissipated. Indeed, each time a gas molecule hits the door molecule, it passes on some of its energy to the door. When the door swings shut, it passes on this energy to the wall, which dissipates it further into the box and into the environment. If you would make the calculations, you are likely to find that the overall thermodynamic entropy is increasing, not decreasing as you would expect if you only take into account your uncertainty about the position of the gas molecules.

If the second law would be transgressed, then you could use this set-up to build a *perpetuum mobile*, a machine that runs forever. It would suffice to make another hole in the wall, this time without a door. The door would produce a constant movement to the left. The hole, because the concentration on the left is higher, would produce a constant diffusion to the right. Together, they would produce a cycle of molecules moving from left to right and back. That movement cannot go on forever, though. Indeed, each time a molecule bumps against the door, forcing it to open, it will lose some of its energy to the door. Thus, it will slow down. If enough molecules have in this way lost energy, they will no longer be able to open the door. Indeed, in order to swing open, the door molecule
must overcome the force of the “spring” which pulls it back. Without such a spring, the
door would stay open, and there would be no ratchet effect, and therefore no cycle. In
conclusion, although the one-way door does decrease our uncertainty about the gas
molecules, it still allows energy dissipation and therefore the increase of thermodynamic
entropy.

More generally, we may conclude that although variation and selection to some
degree *compensate* for each other, they do not *reverse* each other’s effects. In some
respects, they rather reinforce each other. Indeed, more variation will in general lead to
more selection. In the case of the box filled with gas, more variation means faster
movement of the gas molecules, which is the same as higher temperature in the box. The
closer the molecules move, the faster they will leave the right compartment and
concentrate in the left, “attracting” compartment. In other words, the more quickly right
positions will be eliminated, while left ones are selected.

**Order from Noise**

The above principle is universally true: the higher the variation, the more quickly a
system will discover the high fitness, attracting regions. Indeed, more variation means
more different states or configurations that are explored per time unit. And the more
places that are explored, the higher the probability that at least one of these places will
turn out to be part of an attractor.

Note that this principle does not depend in any way on the direction of variation.
Whatever way variation takes you, the faster it goes, the more quickly you will end up in
an attractor. Remember the example of the bottle thrown into the ocean. Whatever
direction the winds blow, more wind will make it end up more quickly on the shore. The
wind does not need to know in which direction the coast lies; the bottle will reach it
eventually. The only thing that can stop it is a complete calm: no wind or current in any
direction. This explains why *blind* variation is sufficient for evolution. Variation will
result in selection while being completely independent of that selection. Variation does
not need guidance or foresight; it just needs to blindly try out different possibilities in
order to eventually hit on one that is retained. Variation and selection is just another
expression for *trial and error*. The more you try, the quicker you will find a solution.

Different scientists have formulated this principle in different ways. Heinz von
Foerster, as we saw when discussing self-organization, called it “order from noise”. By
“noise” he referred to the blind, random, unpredictable character of variation, similar to
the noise that disturbs messages sent over telephone lines. By “order” he meant the result of selection, the attractor state in which the entropy has been reduced. To illustrate the principle, he proposed the following example. Imagine a box with small, rectangular magnets. The magnets are just dumped into the box, without any kind of order or arrangement. Now shake the box energetically. The different magnets will be thrown around, and collide with each other, like the molecules in a gas. Unlike gas molecules, however, two magnets that make contact tend to stick together, through the attraction between their north and south poles. Sticking together is an irreversible process: once the poles touch, they will remain attached, independently of further shaking. In the end all magnets will be attached to each other, forming complex three-dimensional structures, very different from the original random heap. This is the “ordered”, attractor state.

Since you are unlikely to possess a large collection of powerful magnets, I would propose you to try out a different experiment. Instead of magnets you can use paper clips. Just open each paper clip a little so that another clip can slide into it. Now vigorously shake the box with the opened paper clips. If you shake long enough, you will find that the paper clips have arranged themselves in long, branching chains. The mechanism is the same as with the magnets: when through a random movement one paper clip slides into another, the process is asymmetric. It is possible that another shake would undo the resulting connection, but that is much less probable. The state where all the clips are attached is an attractor relative to the state where they are separate. Therefore, sufficient variation will make the clips all (or almost all) end up in an attached state. Notice how the amount of shaking affects the time needed to form chains: strong shaking will lead to results much more quickly. The direction of shaking, on the other hand, has no noticeable effect.
Experiments such as these may seem like nice games, but without any practical use. Yet, if you look around you will find that the “order from noise” principle has plenty of applications. Instead of paper clips, fill a box or jar with any type of small pieces, for example nails, beads, coffee beans, tea leaves or salt crystals. Normally, you would like to get as much as possible into the jar. The trick is to shake or stir the jar repeatedly after you have filled it. You will notice that the jar which seemed full at first, now has quite some space left, so that you can add more nails, coffee or salt. The reason is that at first the different components are spread randomly through the jar, in an arrangement which leaves many small spaces unused. However, since the nails or coffee beans have weight, their most stable position is the one where they are closest to the bottom of the jar. But not all nails can lie on the bottom. Since their position depends on the position of all the others, it is not obvious to find an arrangement where all nails are optimally positioned. By shaking, the system goes through many different arrangements. The more stable the arrangement, the less it will be affected by further shaking, and the more will remain of it after you have stopped shaking.

Random shaking is a universal method to approach problems where you don’t know how to reach the desired effect. If your TV or radio suddenly stops working, hitting or shaking may well bring it back to its normal mode of operation. Of course, if one of the fuses has gone, shaking will not be able to restore it. But if there is merely a bad connection, shaking is often sufficient to make things fall back into place.

Such home remedies may be useful, but they surely cannot compete with high-tech methods? Yet, the same order from noise principle is used in many old and new technologies. Centuries ago, sword smiths discovered the method of annealing to make their blades stronger. The same technique has been perfected in the present manufacture of glass and steel. The principle is that the metal is first heated to a high temperature, then allowed to cool slowly. Metals consist of atoms arranged in a crystal structure. The more regular the structure, the less likely it is that the piece of metal will break. Indeed, when the piece is put under stress, cracks first appear in the places where there are small irregularities or defects in the crystal. The perfectly ordered configuration is the one with the lowest energy. It is therefore also the most stable one.

You can achieve that configuration in a way similar to the shaking of the jar with nails: by adding strong random variation. The equivalent of strong shaking for crystals is strong heating. The hotter the metal, the faster the random movement of the atoms, and the more quickly they will discover stable positions. However, as long as the temperature stays high, no position will be really stable. Indeed, fast moving atoms may have
sufficient energy to knock other atoms out of their positions. Similarly, while you are shaking the jar strongly, the nails will never be able to really settle in their most stable positions. The trick is to gradually reduce the variation, by decreasing the temperature, or by shaking more and more lightly. You are likely to achieve the most stable configuration by maintaining this gradual reduction over a long period, until variation has practically stopped.

To understand the mechanism, you should think about the fitness landscapes, with their local and global minima. To produce the strongest steel, you need to reach the global minimum of the energy. However, you are likely to remain stuck in one of the many local minima, which are less stable, but still easier to enter than to leave. The strong variation caused by the high temperature will shake the system out of any minimum, local as well as global. However, the system will still prefer regions with lower energy to regions with higher energy. That is to say that it is more likely to move to a deep valley than to a shallow valley. When the variation is reduced, the system will no longer be able to escape the deep valley, but it can still escape shallow valleys. The first, very energetic shaking can find the largest and deepest overall valley, but it does not allow the system to settle while the shaking continues. The later, less energetic shaking helps the system to avoid more shallow subvalleys within this deep valley.

Each level of variation makes the system sink deeper and deeper into the energy landscape. At first, it moves with large, violent jumps. Then, it gradually slows down, moving deeper with less speed, but more precision. Thus, the system gradually zooms in on the lowest spot of all. If we had started with gentle shaking, it would never have been able to overcome the steep inclines separating the major valleys. If we had continued with violent shaking, it would never have been able to settle down in the most stable position. That is why the gradual reduction of variation, from very high to practically zero, is the best strategy for exploring a fitness landscape if you don’t know anything about its geography.

This strategy has many applications in addition to metal annealing. Computer scientists use a technique they call “simulated annealing” to find the optimal solution to a problem about which they know very little. The principle is the same: at first, the computer program makes large random jumps, trying out different possibilities that have little to do with each other, but concentrating in those regions which are better with respect to the preferred solution. Then the variation is reduced, and the possibilities that are explored remain closer in the vicinity of the best solution yet. By gradually decreasing the length of the jumps, the program can zoom in on the local (and hopefully also global) optimum.
A similar strategy has been discovered by bacteria, the simplest living organisms. The bacterium, which lives in our gut, E. Coli, has a surprisingly smart method for finding food. Bacteria don’t have eyes that would allow them to see food from afar. They can only “smell” food molecules that happen to pass in their neighborhood, without knowing where those molecules are coming from. How then does an E. Coli know in which direction it should move in order to reach the food source? The principle is simple: the bacterium “swims” in a random direction. If the concentration of food molecules diminishes, that is, if the smell becomes weaker, the bacterium suddenly changes course, and starts moving in a different random direction. If the smell now becomes stronger, it will continue in that direction. Otherwise, it will make another random jump. In fact, the more the smell decreases, the quicker the bacterium will change course; the more it increases, the longer the bacterium will wait before it makes another random jump. This is the same gradual reduction of variation as in (simulated) annealing: the closer you come to the target, the less random changes you allow. Calculations show that such a “blind” method of navigation is quite efficient: on average, the bacterium needs hardly double the time to reach the food than if it were able to see the source and move straight toward it.

Bacterial navigation and annealing differ in an important respect, though. With annealing, the “best” or most stable configuration is always the same. For a bacterium, the place with the highest food concentration changes when one source gets exhausted and another appears. Therefore, a bacterium needs constant variation. It can never come to rest in a final state. Its rate of variation will continuously increase or decrease depending on the availability of food. In situations such as these, where the fitness landscape is changing, variation can never be completely phased out. The order from noise principle then becomes the law of requisite variety. This law was proposed by W. Ross Ashby to describe cybernetic control. We will describe this law in more detail when discussing control. At present, though, we will consider a more general version of the principle: you should maintain enough variation so that you would be able to cope with unforeseen circumstances. This is merely another way of saying that you should not put all your eggs in one basket.

The classic example is the danger of monoculture. Suppose that you are a farmer growing one particular variety of one particular crop. Suppose that an unexpected problem arises, for example a drought, a flooding or a parasite attacking your crop. If you are unlucky, your crops will not be able to resist the infection or flooding. Since the plants are the same, if one lacks resistance, all lack resistance, and all will be wiped out. Suppose now that you have diversified and are growing different crops or different
genetic varieties of the same crop. The same disease or weather change will now affect the different plants differently. Some will die, some will survive, some may actually do better. Though your harvest may not be as plentiful as usual, your farm will be able to survive. The danger of monoculture and lack of biological diversity is well-known. It is a particular problem for Third World countries whose economy is often based on one major crop, such as coffee or bananas. An unseasonal frost, a parasite invasion or merely a drop in coffee prices on the stock market may be sufficient to create wide-spread poverty.

Increasing the diversity of crops is not only a precaution against unforeseen circumstances: it can actually increase the yield. Different plants will use the minerals in the soil differently. Growing only one type of plant will quickly exhaust the particular minerals that plant needs most. That is why farmers have learned to grow different crops on the same patch of land, either simultaneously or the one after the other. Biologists have shown by experiments how diversity increases yield. On similar patches of land, they sowed different combinations of grass seeds. The patches where several species of grass were combined yielded significantly more than the patches with only one or a few grass species. The larger the diversity, the larger the yield, although the effect levelled off if more than a dozen or so species were combined.

The necessity of on-going variation in a changing environment explains many puzzling phenomena. For example, the main effect of sexual reproduction is to increase the genetic variability of the species. Sex, which requires the presence of two partners and a complicated ritual of fertilization, seems much less efficient than asexual reproduction, where a parent can produce offspring just by splitting or budding off. The fact that all unicellular organisms, most plants and many animals can reproduce asexually shows that this is a very basic way of reproduction. Then why did sexual reproduction ever evolve? The detailed answer to this question would take us too long here. Yet, it is clear that the constant variation created by recombining the genes of father and mother in the offspring makes it much easier for a species to adapt to changing circumstances. Since the members of a sexually reproducing species all have a different combination of genes, a new type of parasite or disease is unlikely to wipe out the whole population.

Francis Crick—who together with James Watson discovered the structure of the DNA molecule—has suggested an altogether different application of the order-from-noise principle. After his ground-breaking investigations into molecular biology, Crick turned to brain research. One of the great mysteries in that domain is why people dream. REM sleep, during which most of the dreaming takes place, is characterized by large bursts of seemingly random activity in the neural circuits of the brain. Dreams don’t have
an obvious physiological function. Yet, if you keep people from dreaming—by waking them up each time your EEG equipment registers the onset of REM-sleep—they develop severe mental and physical disturbances. Studying the way neural activation affects our perception and memory led Crick to the following hypothesis.

During the waking hours, your brain accumulates a myriad of impressions, which it tries to fit into a coherent memory trace. However, since new impressions come in continuously, it does not have the time to find an optimal arrangement, where everything would fit into place. This leads to many irregularities and malfunctionings in the neural network—which Crick calls “parasitic modes”. Like the irregularities in an iron crystal, they can only be eliminated by strong random variation. The brain produces the equivalent of heating for metals, or of shaking for a jar of coffee-beans, by releasing strong random excitation during REM sleep. This electrical activity diffuses from neuron to neuron, activating many of the memories that are stored there, but without following any logical sequence. Each time a memory fragment is activated, it will feel as if you relive the experience that created it. Subjectively, you will interpret this seemingly irrational sequence of experiences as a dream.

In conclusion, according to Crick, dreaming is a form of mental hygiene: a way of ordering and cleaning up your store of impressions. This would make your mind more efficient and more creative, by eliminating meaningless or misleading connections. As yet, we don’t know enough about brain functioning to be able to confirm or refute Crick’s hypothesis. Still, the universality of the order from noise principle makes it likely that some form of random variation would be good for the brain. And dreams might well be the shape that such variation takes.

**The Stepping-Stone Principle**

The order from noise principle states that we can always increase the chances of finding a fit state by increasing variation. However, that is not always a practical solution. If you increase variation too much, you destroy the order that is already there. That would not be a problem if you would have a guarantee to find an even better state. But increased variation will only increase the *probability* of finding a solution; it does not provide certainty. If the probability you start with is near to zero, increasing it tenfold or hundredfold may still leave you with virtually no chance to ever find anything. When we consider the evolution of complex systems, such as organisms, instead of simple systems, such as crystals or jars with coffee beans, the number of possibilities becomes
astronomically large. And the fraction of “fit” possibilities, those that correspond to organisms capable of survival, becomes exceedingly small.

This fundamental problem is often mentioned by people who don’t believe in the power of variation and selection. One of their favorite examples is a tornado passing through a junkyard. Like the magnets being shaken in the box, the whirlwind will pick up different pieces of scrap metal, nuts and bolts, and make them collide. Even if we would assume that those pieces of metal would tend to stick together, like the magnets, it is exceedingly improbable that they would assemble in the shape of a complex system, such as a Boeing 747. Since living organisms are orders of magnitude more complex even than a Boeing 747—so the argument goes—it should be clear that no process of random variation could ever produce them by shuffling different molecules together.

Indeed, even the simplest organisms have DNA strings that contain millions of codons—the “words” of the genetic language. Let’s consider a hypothetical primitive organism, a beast so simple that it only needs a thousand words to describe its complete organization. The vocabulary of genetics consists of 20 words, representing the 20 different amino acids which form the building blocks of life. This means that in order to discover the correct 1000 word description of the beast, you would need to guess which out of the 20 amino acids is used in the first place of the description, which is used in the second place, and so on, up to the 1000th place. If you would be guessing at random, you would have one chance in twenty to have the first one correct, one in twenty to have the second one correct, and so on. However, the probability to guess the whole description correctly would be one in twenty to the power 1000, a gigantic number. This number is about 1 followed by 1301 zeros!

This number is much larger than anything you could ever imagine. If you would make one guess per second, and would have been guessing for the whole time since the universe came into being, your number of guesses could still be written down with less than 17 zeros. This means that it would be totally insignificant compared to the number of guesses needed to find the correct description of our primitive beast. In other words, if you were guessing at random, the odds that you would ever come up with the correct description is as close to zero as you could practically get. Increasing variation, that is, increasing the number of guesses per unit of time, would obviously be futile: no variation mechanism can be so fast that it could explore all possibilities in a time less than the age of the universe. If there were only random variation to guide evolution, it is clear that no organism, simple nor complex, could ever have spontaneously assembled.

The fallacy with this argument is that it assumes that selection only takes place at the end of the process: first, all possibilities must be explored, only then will the correct
one be known. All trials that do not result in the correct description are wasted. The only information they give you is that the search must go on. They don’t tell you whether you are making progress or, on the contrary, are getting farther away from the solution. This is obviously not how evolution works. In practice, selection will take place during the whole of the search. Each step will give you a little information, some hints about the direction in which the solution might lie. These hints will make the search much, much easier.

Herbert Simon is a Nobel prize winning economist, but also a path-breaking psychologist, philosopher, management theorist and computer scientist. The thread that runs through the whole of his research is problem-solving: how can people, organizations or computers cope with the complexity of the problems that confront them? One of the fundamental methods to reduce problem complexity is factorization. Simon illustrated this principle with the following example. Imagine that you are trying to crack a safe, by guessing the correct combination of numbers that opens the lock. Suppose you need to turn 6 dials until each of them shows the correct number. Each dial goes from 0 to 9. This means that there are in total 10 to the power 6, that is, one million possible combinations. Only one of those will open the safe. This number is much smaller than the number of possible descriptions of our hypothetical beast, but it is still large enough to make it in practice impossible to open the safe without knowing the combination. How do safecrackers then manage to open safes without using explosives?

The trick is to try out the dials one by one. If the combination of dial settings is used often, it is likely that the dial positions are slightly worn out, so that you can hear a small “click” when the dial passes through the correct setting. That is why safecrackers use a stethoscope, so that they can hear very faint sounds while turning the knobs that govern the dials. Now, the problem becomes much easier. For each dial, the safecracker needs to go through maximum 10 positions until he hears the “click”. Then he can move on to the next dial. For the whole safe, he needs to try out a maximum of 60 settings in order to find the correct combination. The “click” provides the crucial information that he is on the right path, and can move on to the next dial. Without the “click”, he would need to explore all one million combinations to be certain of success.

The essential effect of such information is that it reduces the number of possibilities from a product to a sum. Instead of multiplying the number 10 with itself 6 times, producing one million, you can now add 10 to itself 6 times, producing the number 60. In other words, the number of possibilities will increase proportionally to the number of dials, instead of exponentially. This makes a crucial difference. If you could apply the same trick to find the DNA description of our hypothetical beast, you would have to add
20 to itself a 1000 times, resulting in the number 20,000. If after each time you guess one word of the description correctly, you would hear a “click” to confirm that you are on the right path, you should be able to find the complete description in not more than 20,000 trials. This is a very small number compared to the millions of years that evolution has had to try out different variations. In other words, in this case you can be certain that evolution would find the correct solution.

Of course, the evolution of life is not a simple guessing game in which one amino acid is filled in after another one. Yet, it has the crucial property that there isn’t just one final solution, but a multitude of intermediate solutions. None of these intermediate solutions is optimal, but they all are sufficiently good to be used as “stepping-stones” on the road to the final solution. Imagine that you need to cross a river that is six meters wide. There is no way in which you could step or jump over such a distance without getting wet. Imagine now that different rocks or stones are scattered throughout the river bed, each forming a small island around which the water flows. If the average distance between these rocks is about a meter, you should have no difficulty jumping from the one onto the other without touching the water. Now you can cross the river in six or seven steps. What was impossible without the stepping stones, now has become child’s play. From each stone, you just need to carefully reach out to the nearest stone that lies in the direction of the other bank, testing whether it is stable with one foot, and then transferring your weight onto it.

A similar role is played by the intermediate states in the evolution of a complex system. Each is an approximate attractor, a situation that is sufficiently stable that you could use it to explore the neighborhood, until you find one that is even better. Like in the ratchet effect, evolution will move down the fitness valleys step by step, pausing a little on each intermediate situation, until it finds a way to go even deeper. The steps are like the “clicks” of the dials: they indicate that you already have a part of the solution right, and can build on that for the following variations.

Another example favored by the critics of evolution may illustrate the role of the stepping-stone principle. When Darwin proposed that all complex adaptations had evolved through natural selection, he met with plenty of disbelief. His critics could not understand how a complex organ such as the eye could have come into being through blind variation. Yet, the eye is a particularly clear example of how the stepping-stone mechanism can produce a complex system through a sequence of small improvements. Obviously, the intricate machinery of our eye could never have been discovered by randomly shuffling components in the hope that they would suddenly produce a fully working eye. Yet, there are plenty of intermediate steps that provide enough benefits to
be retained by selection. The biologist Richard Dawkins, in his book “The Blind Watchmaker”, has described these steps in much detail. I will here go through the sequence more quickly, just to give you a taste of the argument.

Imagine a primitive animal, such as perhaps a worm. If by a mutation, some of the cells in its skin would become sensitive to light, then this would benefit the worm, since it now could distinguish between night and day, or between the earth and its surface. This distinction between light and darkness is the first stepping stone on the road towards the eye. The worm would not be able to see any images, though. Imagine now that the patch of light-sensitive skin would sink a little deeper than the surrounding skin, so that it would become part of a cavity. Different parts of the patch would now receive light from different directions, and the animal would be able to at least distinguish some kind of a two-dimensional pattern. The deeper the cavity, and the smaller the hole where it would open to the light, the sharper the image that would be projected into the cavity. This is step two: the formation of a primitive pinhole camera. In step three, some further mutation may produce a patch of transparent skin to cover the hole, so that the light-sensitive cells are protected from water, wind and dust. The “eye” has now become an internal organ, instead of just a part of the skin. In step four, part of the transparent skin may get the shape of a lens, being thicker in the middle than on the border. This further increases the clarity of the image. In step five, small muscles may attach themselves to the primitive lens, so that they can pull on it to make it either more flat or more thick. The lens can now change its focus, depending on the distance between the animal and the object that it is looking at. This leads to a further improvement in sharpness of the image. In a final step—which in fact might as well have occurred at an earlier stage—the light-sensitive cells may mutate into different varieties, so that some are sensitive primarily to green light, some to blue, others to red. The eye can now perceive colors.

In conclusion, these six steps, each of which is relatively simple, have produced an organ that is essentially as powerful as our own eye. That this scenario for the evolution of the eye is realistic, is shown by the fact that there exist animals exhibiting each of the six stages of eye development. Some worms and jellyfish have only the light-sensitive skin, while the Nautilus has just the cavity. Some snails have the protected cavity but without a lens, while many nocturnal mammals have the complete focusing lens, but without the capacity to distinguish colors.

These six steps in the evolution of the eye are like the steps the safecracker follows when searching for the secret code that opens the safe. Each one is a definite step forward, a ratcheting upwards which sets in place part of the solution. Without these intermediate steps, the probability that blind variation could ever have produced the fully
functioning human eye would indeed have been zero. But because each step in the right direction spectacularly diminishes the number of combinations that must be tried out further, the overall probability to develop a complex organ becomes quite high—given enough time to explore the different possibilities.

These examples should not mislead you to think that variation follows some preordained course towards a final solution. In the case of the eye, the overall direction of evolution is clear: to increase the clarity and detail of visual information. In most cases, though, there is not such a clear direction. And there certainly is no final solution, in the sense of an optimal state that cannot be improved. The river-crossing analogy is misleading because it makes us think that there is an end-point: once we are on the other side, the problem is solved. Unlike crystals, complex systems such as organisms don’t have an ideal state, a “global optimum” to which all other states converge. Everything can always be improved further. Moreover, as in the example of the bacterium, the fitness landscape itself changes. Thus, what used to be a good solution may no longer be so in the future.

One of the ways in which this variability of direction shows itself is in the pattern of “stepping-stones”. Steps that seemed to go in one direction often end up somewhere completely different. Many of our organs were initially selected for functions that had nothing to do with the ones they perform now. For example, when the first fishes started to adapt to life on the land, you would expect that they would have used the same organs for breathing as the ones they already used in the water: their gills. Yet, the organs which their descendants use for breathing, the lungs, evolved out of something completely different: their swim-bladder. The gills were specialised to extract oxygen from water but could not adapt to breathing air. The swim-bladder, on the other hand, already used air, but for a wholly different purpose: to regulate the vertical position of the fish in the water, by making it either float or sink. This function became useless for life on the land. Thus, the direction of further developing the gills was abandoned, while the swim-bladder became a stepping stone towards the development of lungs, veering off altogether from its initial course.

Most of the great puzzles in evolution can be solved through a similar reasoning. Often, there does not seem to be a path that would explain how an existing system could have evolved out of a simpler system through a series of fitness-improving steps. The fact that we cannot think of such a path does not mean that none exists, though. Understandably, our bias is to look for direct, straightforward paths, that go from A to E via B, C and D. Evolution, however, does not work that way. It has no global sense of direction. It cannot see where it is going to. It just will reach whichever state seems
locally more fit than the present state, meandering through a space with an infinite number of dimensions. Thus, it might well reach E from A by first going through Z, V, M, O, and I. However, like the bacterium’s steering mechanism, these random changes of direction are quite efficient to steer the system towards ever increasing fitness. Our mind is much too limited to conceive of the virtually infinite number of detours and side roads that evolution explores. That is one reason why many people still find it so difficult to grasp the power of variation and selection.
Fitting in

The Survival of the Fit

The meandering course of evolution may create the impression that it is totally directionless. Yet, the idea behind the ratchet effect is that there is a preferred direction. How can evolution at the same time wander unpredictably and have a preferred direction? An example will clarify the situation. Imagine that you let a ball roll down from the top of a steep mountain. On the rugged terrain, the ball will constantly encounter obstacles: rocks, trees, crevasses, ravines. Depending on the precise angle and speed with which it hits the obstacle, it will be deflected either in one direction or in another direction. Thus, its downwards course will be very irregular and difficult to predict. If you repeat the experiment with different balls, you are likely to find that the positions where the balls finally come to rest are miles apart. Yet, one thing will be certain: their final position will be lower than their initial position at the top. Although you cannot predict the movement in the horizontal direction, you know that there is only one possible direction in which a ball can move vertically: downward.

To find the equivalent of such a movement for evolution, we need to discover the evolutionary equivalent of the height or vertical direction. In other words, we need to find a property that can only decrease (or increase) during evolution. As we noted earlier, in physics, that property is potential energy, which will always decrease, or thermodynamic entropy, which can only increase. In biology, though, that property is fitness, which tends to increase. It is now time to define the more subtle concept of fitness, while showing how energy can be seen as a special case of this more general idea.

“Fitness” as a term describing evolution comes from biology. It was first used by Herbert Spencer, who described the result of selection as “the survival of the fittest”. This phrase was later taken over by Darwin himself, and by the following generations of evolutionary biologists. At first, the meaning of fitness was purely intuitive: the “fittest” individuals were those strong enough to win the “struggle for life”. Strength is not a particularly good indicator of evolutionary success, though. Many of the most numerous species, such as bacteria, algae, sponges, mushrooms, ..., can hardly be called “strong”. Their success derives from different features. In addition to “strong” or “in good
condition”, the English word “fit” has also another meaning: “well-suited”, “adapted” or “fitting”. Whereas the first meaning is *absolute*, emphasizing the intrinsic worth or strength of something “fit”, the second meaning is *relative*, emphasizing that something is “fit” only because it suits or agrees with something else. Thus, later generations of biologists stressed that fit organisms survive because they are adapted to their environment. They don’t need to be intrinsically robust or vigorous, as long as they have found a little place, a “niche”, in which they snugly fit.

The increased understanding of genetics brought with it a mathematical approach to evolution. Knowing how the genes of mother and father combine to produce the genes of their offspring allowed geneticists to calculate how many genes of each type there would be in each generation. Since in these abstract models it was not possible to distinguish genes that are stronger or better adapted than others, the geneticists needed to define fitness in another way. They defined fitness simply as the average number of offspring that an individual with a particular gene combination is expected to have. If that number is larger than one, then those genes will become more frequent in the population. If it is smaller than one, then they will become less frequent. When you consider many generations, then you should expect genes with fitness lower than one to gradually disappear from the population. In other words, fitness lower than one means eventual extinction. Fitness equal to one means survival. Fitness larger than one means growth or proliferation.

This definition of fitness is perhaps the most general, since it does not depend on the specific causes why a particular type of organism would be more or less likely to survive. An organism can be fit because of intrinsic strength, good adaptation, or any other reason. The basic idea is that fitness distinguishes those that are selected from those that are eliminated. We have argued earlier that *stability* distinguishes retention from elimination. This is the physicist’s view. What biology has added is the possibility for *multiplication*: fit systems not only persist, they tend to become more numerous. Since particles do not reproduce, physicists normally ignore that possibility. Yet, multiplication is not limited to the realm of biology.

Chemists are well-acquainted with the phenomenon where the presence of a few molecules of a particular type leads to a growing number of these molecules. They call such molecular reproduction *autocatalysis*. Catalysis is the process where the presence of a “catalyzing” substance makes it easier for a reaction to produce new substances. Autocatalysis is when the catalyzer facilitates its own production. But in physics too we can find examples of self-reproduction. For example, when you drop a salt crystal in a saturated salt solution you will see the crystal grow. The reason is that the salt molecules
attach themselves to the crystal, because these positions have a lower potential energy. But the more molecules become part of the crystal, the larger the crystal grows, and the more positions there are for further molecules to attach themselves. Thus, we can say that the regular grid structure of the crystal reproduces itself throughout the solution, becoming ever larger, until it has absorbed all available salt molecules.

To arrive at a fundamental definition of fitness, I want to make one further generalization. In biology, the only way to increase the number of individuals is by reproduction. Without parents, you cannot have offspring. In the case of the salt solution, though, crystals will start to appear and grow even without a “parent” crystal that is added from the outside. Seeding the solution by dropping in a crystal will accelerate the process, but it is not necessary to make it happen. The crystalline state is fit because it spontaneously appears and grows, taking over from the state where the salt molecules are dissolved in the liquid. In the right circumstances, crystal patterns are produced in large numbers. It is not so important whether they appear on their own, or whether they “reproduce” from a parent crystal. What is important, is that they grow in numbers, while their competitors, the dissolved molecules, decrease in numbers. Thus, the crystal patterns are selected, while the dissolved patterns are eliminated. In sum, within the given circumstances the crystal patterns are more fit than their competitors.

Fitness in this case means simply the number of patterns we can expect at the next “generation”, divided by the number of patterns there are now. Like in the case of the geneticist’s fitness, fitness larger than one means increase in the number of patterns; fitness smaller than one means decrease; and fitness equal to one means maintenance of the present number. This definition of fitness applies to all possible systems, patterns or configurations, to anything we can distinguish. It applies to atoms, molecules, crystals, snow flakes, plants, animals, societies, inventions, ideas or habits. It generalizes the concept of stability, which describes the situation where fitness is smaller than or equal to one. Instead of distinguishing merely between configurations that maintain and configurations that disappear, we can now also describe configurations that appear or multiply.

Note that this implies that there are two basic strategies for achieving fitness: you can either aim for fast multiplication or for high stability. This is what biologists call respectively $r$ and $K$ selection. Organisms like bacteria, mice and weeds follow an $r$ strategy: they produce many, short-lived offspring. Elephants, trees and people, on the other hand, follow a $K$ strategy: their offspring are few, but long-living. If we extend this concepts to the realm of physics, we might say that rain drops follow an $r$ strategy, while stones follow a $K$ strategy. Both types of systems are fit, since they are unlikely to be
eliminated. The reason is that the lack of stability is compensated by a surplus of multiplication, and vice versa.

**Fitness and Causality**

In this most general sense, fitness is a measure of the likeliness that we will encounter a configuration in the future. The higher the fitness, the higher the likeliness. Natural selection then means that not all configurations have the same fitness: some will become more frequent, others will become less frequent. This has an important implication for the things we see around us. Since natural selection has been operating for a very long time, we should expect most of the things with low fitness to already have been eliminated. This means that the majority of phenomena we see around us will have a fitness at least equal to one. Of course, variation will continuously produce unfit configurations, but those, by definition, won’t last, and will have little effect on their surroundings. What remains must be fit.

Look around you and consider all the different kinds of things that make up your environment: stones, clouds, the sun, household objects, flowers, trees, animals, people, institutions... If my assumption is correct, they must all be fit under the above definition. This means that either they must be very stable—stones, the sun, some institutions—or they must be (re)produced frequently—all living things, but also unstable phenomena such as clouds, rain drops or snow flakes. Although clouds, for example, are far from stable—a bit of wind or sun can make them diffuse—they are recurrently produced in normal atmospheric circumstances. Therefore, although they cannot survive long, and cannot reproduce, they are still fit under my definition.

The fact that most phenomena are fit can help us to explain why the world is understandable. According to the Newtonian world view, we can understand the world because it is governed by causality. Observing causes allows us to predict their effects. In the new world view, we can no longer take causality for granted: the world is intrinsically unpredictable. But if the world is so chaotic, then why do our theories work most of the time? Why are most of our expectations confirmed? The answer is that most phenomena are fit, and fit phenomena—because of their stability or recurrence—are predictable. Although the universe is infinitely complex and changeful, the parts that are irregular, chaotic or surprising are in essence fringe phenomena. The general rule is fitness. The unfit phenomena are the exceptions: they will sooner or later be eliminated by selection.

Cognitive scientists have tried to understand how people, using very simplified theories of how things function, can still reason adequately about the complex world in
which they live. One of the basic mechanisms they discovered is default reasoning. Default reasoning means that although you may be aware that exceptions exist, you will first of all assume that the things around you follow the general rule. Since the general rule is true most of the time, this default assumption saves you the time and effort needed to consider all possible special cases that might apply. Only if the default assumption turns out to be wrong will you start considering alternative possibilities. For example, if you see a bird, you will assume by default that it can fly. Only after you observe that it does not seem to fly, you will start thinking that it may be ill, or have a broken wing.

In my paper “Fitness as Default”, I have argued that default reasoning only works because the world is organized in such a way that the general case is much more likely than the alternatives. This means that the entropy (in the sense of uncertainty) of the world must be low. Otherwise, the different possibilities would be about as probable the one as the other, and none would stick out as “the general rule”. Such a low entropy distribution can be explained by natural selection eliminating all but the few unfit possibilities.

Our natural tendency is to assume that a phenomenon we newly encounter must be fit, until proven otherwise. For example, we expect by default that stones are hard, that buildings are stable, that people are healthy, that trees are tall, and that bacteria multiply quickly. Indeed, by now most soft stones have eroded or crumbled while the unstable buildings have been destroyed. People who are ill either are cured or die. They rarely remain ill for a long time. Trees in a forest that don’t grow tall enough are overshadowed by their competitors and die from lack of sunlight. Bacteria that don’t reproduce quickly enough would by now already have lost the competition with more prolific species.

Moreover, the fact that fitness is common even may explain why causality works. Remember the distinction we made between macroscopic causality and microscopic causality. Microscopic causality, the principle that every unique state of the universe C causes another unique state of the universe E, \( C \rightarrow E \), is necessarily true, but totally useless, since we can never know C and E. Macroscopic causality, the idea that a known cause K produces a known effect K’, is the basis of all predictions. Yet, it is only justified if we can eliminate the background factor B from the formula K + B \( \rightarrow K’ + B’ \). There are at least two cases in which we can ignore the unknown background factor.

The first case is when K, the factor we are interested in, is not affected by B. This will often be the case when K is fit, because fitness implies that K is stable and can resist different kinds of perturbations. For example, a heavy desk is intrinsically stable and will not be affected by wind, vibrations or random distributions of air molecules. Therefore, we can predict that the position of the desk will only be affected by a push strong enough
for us to observe it. We don’t need to consider undetectable background effects. The coin standing on its edge, on the other hand, is in a definitely unfit state, which will be quickly eliminated. Because of this intrinsic instability, it is sensitive to the slightest perturbations, and therefore macroscopic causality will not help us to predict its trajectory. Because of natural selection, the stable and predictable state of the desk will be much more common than the unstable and unpredictable state of the coin.

The second case in which we can forget about the background is when that background itself is stable. If the background conditions are always the same, then we don’t need to know how they affect the system K in which we are interested. We just need to observe a few instances of the phenomenon and derive a general rule. For example, by watching apples falling from a tree we may derive the rule that heavy objects fall to the ground. If we observe carefully, we may even derive the precise acceleration with which they fall. However, we don’t need to know that what determined their fall is an invisible force called “gravity”. Gravity only becomes important for predictions when it varies. For example, the gravity on the Earth is different from the gravity on the Moon or in space, and therefore objects will fall differently in those different places. Gravity can be ignored on the surface of the Earth because it is invariant: it is everywhere the same and it never changes.

Each time we predict that a heavy object will fall to the ground, we unconsciously make a lot of assumptions about the background conditions. We assume that gravity will remain constant. We assume that there won’t be a tornado sucking in objects and blowing them high up in the sky. We assume that there won’t be a magnetic field counteracting the force of gravity. We assume that the medium around the object (normally air) is lighter than the object itself. (Otherwise the object would float upwards, like a balloon, rather than fall.) We assume that the object is not self-propelled, like a helicopter or a fly. We assume that the object will not be picked up during its fall by a person. We assume that the object will not suddenly change its physical structure and become lighter than the air. We assume that the laws of physics will remain the same...

There are an infinite number of such assumptions that we need to make if we want to justify our prediction. Yet, in practice, we can forget about all these conditions by simple assuming that the situation is normal. “Normal” means fit, stable, or “the way things usually are”. If things were constantly changing, so that we could not trust whether gravity would still be there the next we time we looked, then predictions would be impossible, and macroscopic causality would not exist. It is because natural selection prefers stability that we can ignore the background factor B, and focus on K. In other words, macroscopic causality is a product of natural selection.
This argument also explains why causality is more reliable in physics than in the social sciences. When physicists express physical laws, such as “massive objects in a gravitational field fall”, or “opposite electric charges attract”, they are much more categorical than social scientists expressing the laws of their domain. Social scientists tend to accompany any rule or prediction with the phrase *ceteris paribus*, which is Latin for “all other things being equal”.

For example, a psychologist might predict that if you ask people whether they would prefer going to a wild party or staying at home to read a good book, then the extroverted will prefer the party while the introverted will prefer the book. This is a causal relation: the distinction introvert-extrovert is mapped upon the distinction book-party. Yet, we all know that predicting people’s behavior is not that simple. Perhaps the extrovert would be tired and prefer to stay at home, while the introvert would just have been studying for exams during the last couple of weeks and be ready for some action. Or the introvert might hope to run into an old friend at the party, while the extrovert would be afraid to meet an ex-spouse. You will never know all the factors that influence the decision. That is why such a prediction needs a qualification: *all other things being equal*, an introvert will prefer the book, while an extrovert will prefer the party.

Of course, we will never find two people whose situations are completely equal except for their degree of introversion. Therefore, the psychologists’ prediction is only statistically valid: *on average*, introverts are more likely to prefer the book. Social science predictions are less reliable because the background factors in social situations are much more variable than the background factors in physical situations. The social domain is still in full evolution, while in the physical domain, things have had plenty of time to settle down and reach stable configurations. This has not always been the case, though: shortly after the Big Bang, the physical domain of particles and forces was in full upheaval and the “laws of Nature” that we rely on now would have been essentially useless as tools for making predictions. We will discuss later why the more recently evolved domains tend to be more complex and variable than the older domains.

But even in physics many phenomena, such as the weather or the behavior of quantum particles, remain unpredictable. One reason is that there will always be variation. No configuration is absolutely stable. There will always be exceptions, or chance fluctuations. Another reason is that fitness is not just stability. Fitness is also growth, multiplication or proliferation, and that can lead to chaos. To understand that better, we will need to look at the roles of positive and negative feedback.

In conclusion, the concept of fitness can help us to explain not only the organization of the universe, but the organization of our own knowledge about the
universe. In order to really understand, though, we need to know more about what makes a system fit. Our definition says that fit systems survive and multiply, but it does not yet tell us why some systems survive or multiply better than others. To unravel this mystery, we will need to revisit the phenomenon of self-organization.

Self-Organization or Selection?

Darwin’s theory of natural selection has always been controversial. Evolution through variation and selection contradicts both the older world view, where complex phenomena are produced by supernatural powers, and the Newtonian world view, where everything is determined once and for all by the laws of nature. Both views are still very influential. This explains why many people find it so difficult to accept Darwin’s ideas. More surprisingly, however, Darwinism as an explanation for complexity is also criticized by a growing number of post-Newtonian scientists. Perhaps the most well-known of these critics are the complexity theorist Stuart Kauffman, the cell biologist Lynn Margulis, and the paleontologist Stephen Jay Gould. It is worth looking into their arguments.

Kauffman starts by contrasting self-organization and selection. Self-organization, as he formulates it, produces “order for free”. In other words, organisms don’t need to undergo selection by their environments in order to become complex: self-organization will produce such complexity on its own. A theory of evolution which ignores these intrinsic capabilities will never be able to explain all the richness and sublety of biological organization.

Margulis, on the other hand, stresses symbiosis. Complexity evolves not only when an individual organism undergoes variation and selection, but when different organisms come together in a state of cooperation or synergy. Margulis is most famous for her theory that the complex cells of plants and animals arose through the symbiotic merging of more primitive bacteria-like cells. The integration of different systems or parts of systems in a larger whole is an essential component of creative evolution.

The “punctuated equilibrium” theory of Niles Eldredge and Stephen Jay Gould proposes yet another criticism of traditional Darwinism. Eldredge and Gould observed that evolution tends to happen in fits and starts. Variation, as we saw, tends to be small. Therefore, Darwin saw evolution as a slow, continuous process, without sudden jumps. However, if you study the fossils of organisms found in subsequent geological layers, you will see long intervals in which nothing changed (equilibrium), “punctuated” by short, revolutionary transitions, in which species became extinct and replaced by wholly
new forms. Instead of a slow, continuous progression, the evolution of life on Earth seems more like the life of a soldier: long periods of boredom interrupted by rare moments of terror.

Various other criticisms have been raised against the Darwinian picture of evolution. One important observation is that variation is in general not random. For example, when you put bacteria in an environment to which they are poorly adapted their mutation rate increases. Because of the order-from-noise principle, more mutations produce faster evolution. This allows the bacteria to adapt quickly to their new environment. Yet, if mutations were purely random, they should not be influenced by what happens in the environment. A related criticism notes that not all properties of an organism can vary freely. There are intrinsic constraints on the development of an organism, which make certain variations simply impossible.

Although these different criticisms object to the variation and selection picture of evolution, their observations point towards a very different shortcoming of Darwin’s theory. Like the other non-classical theories, Darwinism introduced a revolutionary new idea, but inherited most of its other concepts from the then prevailing Newtonian world view. That means that its general outlook is reductionist: it analyses the world into atomic components. These components are similar to the particles and forces of Newtonian mechanics: the individual organisms play the role of the particles, and the environment plays the role of the force field, which directs the evolution of the particles. In the more modern, neo-Darwinist approach, the role of the particles has been taken over by the genes, but the rest of the theory has basically remained the same. As the Austrian biologist Rupert Riedl proposed, it is time to replace this reductionist approach by a systems theory of evolution.

As you will remember, the main idea of systems theory is that you cannot understand a system in isolation. Every system interacts with other systems. Together they form a whole, a supersystem. In turn every system is composed of interacting components or subsystems. Thus, to understand how a system functions you should look both inside, at its subsystems, and outside, at its environment and the larger wholes to which it belongs. Let us apply this approach to variation and selection. For a given system, there will be both internal variation, affecting the system’s components, and external variation, affecting the system’s relations with other systems. There will also be internal selection, eliminating or retaining some of the changes in the system’s components, and external selection, eliminating or retaining the system as a whole.

In the traditional Darwinian picture, all variation is internal, caused by mutations or recombinations of the genes, and all selection is external, governed by the
environment. By adding *external variation* and *internal selection* to this picture, we can counter all at once the different criticisms raised against Darwinism. Self-organization is simply the result of internal selection. The basic ideas underlying variation and selection that we discussed, such as attractors, order from noise, and the stepping-stone principle, apply equally well to self-organization and to Darwinian evolution. Most of the examples we discussed, such as crystallization or paper clips arranging themselves in chains, are in fact typical examples of self-organization. The only difference between self-organization and Darwinian selection is that the former does not need an external environment. Self-organization occurs when a system discovers a configuration that is *intrinsically* fit or stable, independently of the environment.

Internal selection also explains the fact that variation is constrained. Most mutations of the DNA molecule that carries genetic information are intrinsically unfit, and will be eliminated long before they come into contact with the environment. We can imagine many different variations that will never get the chance to be expressed in the environment: the mutated DNA might be unstable and fall apart; the mutated DNA might not make any sense with respect to the genetic code, and would therefore not be interpreted in the form of amino acids; the mutated DNA might code for a sequence of amino acids that is unstable or that has no function in the cell; the sequence of amino acids might be poisonous and kill the mutated cell before its DNA could be passed on to daughter cells; etc. Moreover, the cell has various repair mechanisms that will eliminate or correct DNA errors. These repair mechanisms will be better at eliminating certain types of mutations than others. This determines another type of internal selection. Finally, even if the mutation would survive all these forms of selection internal to the cell, it is still likely to fail when trying to grow into a complete organism. For example, the mutated embryo may be spontaneously aborted, and never get the chance to be born and produce offspring of its own. None of these selection processes involves the environment.

Still, the environment could affect these internal selection mechanisms indirectly. This may help us to explain why the mutation rate increases for bacteria in a stressful environment: the repair mechanisms that eliminate mutations might simply become less effective under stress. You could expect that when a bacterium has difficulty surviving, it won’t be able to put a lot of energy into repairing DNA errors either. Alternatively, you might assume that the bacteria have evolved their own, in-built version of the order-from-noise mechanism, and that their repair mechanisms are programmed to become more tolerant of errors in difficult circumstances.

External variation, on the other hand, is sufficient to explain the emergence of symbiotic systems. Instead of varying the internal structure of an organism, you might as
well vary the network of interactions in which it takes part. If the new network of relations between different interacting organisms is more fit, that network will be selected as whole. If that whole is sufficiently stable and tightly integrated, then we might consider it as an organism in its own right, rather than as a mere assembly of organisms. This is what happened with the appearance of complex, nucleated cells, which integrated more primitive cells. This is also how multicellular organisms evolved out of an assembly of unicellular organisms. In a later section we will discuss in more detail how such supersystems emerge through the integration of subsystems.

Finally, a systems approach can explain why evolution tends to move in fits and starts. Later we will discuss how negative and positive feedback can either slow down or speed up evolution, resulting in an extremely variable speed. At present, it is sufficient to consider a typical fitness landscape, in which there are valleys separated by ridges. If the evolving system has reached the bottom of a deep valley, there will be almost no change, since variation will fail to pull the system out of that hole. On the other hand, if there is only a small ridge separating the valley from a neighbouring, deeper valley, then a chance event may be sufficient to push the system over the edge so that it enters the other valley. Such a lucky variation will become increasingly likely when the fitness landscape changes so as to reduce the height of the ridge. Once over the ridge, the descent into the new valley will go very fast. This means that the system will evolve very quickly to a new, fitter configuration. If we would check the evolution of the species in the geological record, we would find many fossils corresponding to the position at the bottom of the valley where the organism remained for so long, but few or none corresponding to the crossing of the ridge, which happened very fast on the geological time scale.

The systems approach can help us to understand more profoundly how a small variation can produce a major change. Indeed, organisms, like all systems, are organized in levels, corresponding to their subsystems and subsubsystems. Each subsystem is described by its own set of genes. A mutation in one of the components at the lower levels will in general have little effect on the whole. On the other hand, a mutation at the highest level (the recently discovered “homeobox” genes), where the overall arrangement of the organism is determined, may have a spectacular impact. For example, a single mutation may turn a four-legged animal into a six-legged one. Such high-level mutations are unlikely to be selected, but potentially they can lead to revolutionary changes.

In conclusion, it is clear that a multi-level, systems view of variation and selection solves many of the riddles about evolution. Just considering the possibility of both internal and external processes already integrates selection with self-organization and symbiosis. You don’t need to introduce separate principles to describe these processes:
everything can be explained by variation and selection. Yet, you might wonder whether this isn’t just a play with words, where I replace the term “self-organization” by the term “internal selection”. Perhaps, what I called internal selection is really so different from traditional Darwinian selection that we might as well stick to the name “self-organization” in order to emphasize the contrast. This is probably what Kauffman would argue. Using the systems approach, however, it is easy to show that there is no fundamental difference between internal and external selection. The only difference is the point of view.

Let us consider a typical example of external selection. An animal, in order to survive, needs oxygen. The amount of oxygen in the air is a selective factor. Variations which reduce the capacity of the lungs to extract oxygen will be eliminated. Thus, animals living at a high altitude, where there is less oxygen, will have larger lungs and more red blood cells in order to use the available oxygen more efficiently. The oxygen concentration is a property of the environment. Therefore, the selection which it exerts is a typical example of external, Darwinian selection. On the other hand, oxygen is produced from carbon dioxide by plants. For plants, the concentration of carbon dioxide is the more important selective factor, while oxygen is merely a waste product.

But where does the carbon dioxide come from? It is produced mainly by animals! Thus, you find mutual dependency: the animals, by producing carbon dioxide, determine which kinds of plants will survive, while the plants, by producing oxygen, determine which kinds of animals will survive. If you consider plants and animals together, that is, if you consider the global ecosystem, then the concentrations of oxygen and carbon dioxide are determined internally. The ecosystem of the Earth can survive only because the two halves of the cycle are in balance: the amount of oxygen consumed by animals matches the amount of oxygen produced by plants. Without this equilibrium, the world might run out of oxygen and both plants and animals would perish.

Where does this equilibrium come from? It is obviously not the result of some kind of selection external to the Earth. There are no cosmic factors which determine the amount of oxygen on the Earth. The oxygen concentration is determined purely internally, by the self-organization of the global ecosystem. This capacity of the Earth to regulate its own environment has led James Lovelock and Lynn Margulis to propose that the Earth itself should be seen as a living organism. They call this global organism “Gaia”, after the Greek goddess of the Earth. You don’t need to believe that the Earth is alive to agree that it is a self-organizing system, though. But note that what we now call self-organization, we first described as external selection. What is the difference? Only the point of view. First, we considered animals, for whom the oxygen concentration is
external. Then, we considered the Earth, for which the oxygen concentration is internal. But the mechanism that keeps animals, plants, carbon dioxide and oxygen in equilibrium is always the same: selection of a fit configuration. Each of the components in the cycle exerts a selective influence on the other. Together, their configuration is selected by its overall stability: unstable configurations are simply eliminated and replaced by new configurations. Later we will discuss the feedback mechanisms which help to reach such global equilibrium.

**Co-Evolution: fitting together**

To better understand the complex interaction between selection processes in different systems, we need to introduce a new concept: *co-evolution*. The ecologist Paul Ehrlich was one of the first to emphasize that different plants and animals always evolve together. The buffalo, which lives from grazing, needs to adapt to the specific types of grass that grow in its environment. In turn, the grasses need to adapt to the buffalos eating them. Plants that cannot survive if their upper parts are regularly chewed off will simply be eliminated. Those that remain have been selected for their resistance to grazing. In turn, the buffalos are selected for being able to digest the hard and dry tissues of grass. The end result of such co-evolution is mutual adaptation: the grasses can live with the buffalos; the buffalos can live with the grasses.

Such mutual evolution cannot be understood by assuming a fixed selection criterion, which either the buffalo or the grass tries to fulfill. There is no given fitness landscape in which the buffalo tries to find the deepest valley. Since the fitness of the buffalo depends on the grass, and vice-versa, every movement of one species in its fitness landscape will change the fitness landscape of the other species. If grasses become stronger to better resist being pulled out by buffalo, then the buffalo will need to develop better teeth and stomach to chew and digest the grass. If the buffalos get stronger teeth, the grasses in turn may need to grow closer to the earth in order to avoid being pulled out altogether. This will in turn trigger a new series of adaptations in the buffalos.

It is as if the two fitness landscapes are made out of rubber and attached by strings, so that as one organism walks around on its rubber sheet, the other sheet gets deformed. Co-evolution makes us understand the limitations of fitness landscapes as a representation of selection. Fitness landscapes are only useful insofar that the fitness function remains fixed (or changes in a predictable way). This is in general not the case with co-evolution. However, that does not mean that we should abandon the concept of fitness altogether.
In fact, fitness can help us to understand the mutual dependence of co-evolving systems. Remember the two meanings of the word “fit”: fitness as intrinsic strength, and fitness as suitability or adaptedness. At first sight, these two meanings—the one absolute, the other relative—seem to have little in common. Yet, from a systems point of view they are merely the two sides of the same coin. Intrinsic strength refers to the result of internal selection, to the property of forming a robust, well-connected whole. Adaptedness refers to the result of external selection, to the property of “fitting in” to the environment. The systems approach is ultimately also a relational approach: all systems are relations between their input and their output. In this most general sense, fitness is a characteristic of relations. A fit relation is simply a relation that is selected; in other words, a relation that maintains or becomes more numerous. When such a relation has evolved, we may say that the objects between which the relation holds “fit”: they stick together; their connection or mutual arrangement is stable. Assemble two pieces of a jigsaw puzzle. If they fit, they will be difficult to dislodge. If they don’t, they will separate easily. Thus, “fitness” in essence describes the strength of a connection. This includes both the first meaning, which emphasizes the strength of the internal connections between a system’s components, and the second meaning, which emphasizes the strength of the external connection between the system and its environment.

We are lucky that the English language can express these two aspects of the same function by the single word “fit”. Otherwise, we would have had to invent a new term for this fundamental concept. Unfortunately, most other languages, such as French, Russian or German, don’t seem to have an equivalent word. I could only suggest to eventual translators of this work that they keep the English term “fitness”—just like many languages have adopted the English word “computer” without translation.

Let us look in more detail at the process of co-evolution, and the discovery of fit relations. One way to understand this process is to look at it as if it were a conversation between two partners—the co-evolving systems. Imagine that you and your partner are trying to decide what to do tomorrow. You might suggest to go and visit your aunt. This suggestion can be seen as a variation, as a new idea offered for evaluation. Your partner will either accept the suggestion, or reject it, for example by remarking that visits to your aunt are so boring. This is selection performed on the variation. If the suggestion is accepted, the process stops and you have reached an agreement. If it is rejected, either you or your partner will make a new suggestion, for example “Let’s go to the movies”. This new variation will again undergo selection. Perhaps you’ll note that no good movies are being shown at the moment, thus rejecting your partner’s suggestion. In turn, you may suggest to drive around in the countryside. This time, perhaps, your partner may be
more positive, and reply that that’s not a bad idea, but that it would be nice as well to visit your friends. You may reply that visiting your friends would fit in with in the country-side drive as long as you go in the direction of the North. Your partner may again note that if the plan is to go North, then you might as well visit that nice little village which is not far from where your friends live. You might add that it would be better to first visit the village, and use the occasion to have a drink at the famous local pub, and then go to your friends and spend the rest of the day with them. Your partner might conclude that that seems like a good way to spend the holiday.

Thus, you have reached an agreement, which is accepted by both parties without need for further variation. The agreement is like a “fit” relation, binding you and your partner together for the next day. The conversation that produced it followed the typical pattern of an evolutionary process: first, large, rather disconnected variations, which are quickly eliminated, followed by smaller variations which build on what was already achieved, thus gradually zooming in on a final solution. The difference with the examples we discussed earlier is that there are now two parties involved, mutually performing variation and selection. Thus, it is an example of co-evolution towards a collectively fit configuration. Yet, if we consider the couple formed by your partner and you as a single system, then variation and selection are purely internal to the couple, and the process could be seen as self-organization. On the other hand, if we would consider you as an individual system in its own right, and your partner as the environment to which that system tries to adapt, then we could see the process as Darwinian selection, where you try different variations until you find one that fits the environment.

All co-evolutionary processes between two systems can be seen as a kind of “conversation”, where each of the partners offers “suggestions” until a configuration is reached that is accepted by both. Of course, the partners may never reach an agreement, and continue to try to get some advantage which the other partner is not willing to grant. This is typical for competition, where the goals of the two systems are mutually exclusive. For example, the buffalo’s goal of eating as much grass as possible cannot be reconciled with the grass’s goal of not being eaten. At best, they reach a “compromise” where the buffalo eats a certain part of the grass, while leaving enough of the plant so that it can grow back. In this case, the compromise solution will be selected, but it will remain intrinsically less fit than a potential “synergy” solution, where both partners would profit from their relation. Synergy is characteristic of mutualist symbiosis, like in the animals and the plants producing respectively the carbon dioxide and the oxygen that the other party needs.
Feedback as Fuel

Viewing co-evolution as a conversation or, more generally, an interaction, can help us to understand some further puzzling features of self-organization. Imagine that one of the parties initiates variation, that is, takes a particular action. The other party may accept that action as such, without further reaction. In that case, the effect of the action is retained. However, if the other party is not perfectly “happy” with the action, it will react, and propose another variation. This may in turn be accepted as such, or trigger another action, and so on. Each reaction can be seen as feeding back the other party’s appraisal to the party that initiated the action.

Imagine that you are keeping a big dog on a leash. Suppose you want to go in a particular direction, so you pull the dog in that direction. If the dog accepts your lead, it will follow. However, if the dog would like to go somewhere else, it will pull you in a different direction. Suppose that neither of you is strong enough to force the other one to follow. In that case, your overall trajectory will be a give and take between the places where you would like to go and those where the dog would like to go.

In general, the reaction will be independent of the action. In that case, the overall pattern of movement will look like a random walk, constantly changing direction. Like the bacterium searching for food, these erratic movements may slow down and zoom in on an attractor, or they may continue without apparent limit (see Fig. a). However, there are two special cases where action and reaction are somehow “parallel”. The reaction can point in the direction opposite to the action (Fig. b), or in the same direction (Fig. c). The first case, where every action is immediately counteracted, corresponds to negative feedback. The second case, where every action is on the contrary reinforced, corresponds to positive feedback.
Both types of feedback produce selection, and therefore self-organization. Negative feedback is the simplest to understand. Imagine that the dog doesn’t want to go anywhere. Whichever direction you pull, it will pull in the opposite direction. The result is a balance of forces: in spite of all the energy invested, nothing moves. Both you and the dog remain in the same position. That position is stable. It functions as an attractor. Thus, situations characterized by negative feedback are directly selected.

With positive feedback, the selection is more subtle. Suppose that your pull stimulates the dog to start moving. Instead of just following your lead, it moves ahead, pulling you behind. Seeing the enthusiasm of the dog, you start to walk faster, pulling it forward more. The dog, in turn, pulls even harder. In the end both of you are running. What was selected now, was not your initial position, but the initial direction of movement. Whatever appeared or increased in the first move, was increased even further in the second move, and so on. This is selection, not for stability, but for multiplication. The typical example is reproduction: the more offspring an individual has, the more individuals there will be to produce further offspring. As we saw when discussing economic development, such runaway processes must come to an end when the resources for further growth are exhausted. If there are too many rabbits, all the grass will be eaten, and the rabbits will starve.

Paradoxically, when such a growth process reaches its limit, the positive feedback that fueled it will turn into a negative feedback. Imagine that the rabbit population
increases for as long as there is plenty of grass. The moment the amount of grass being eaten by the rabbits is equal to the amount of new grass that grows, there is an equilibrium. If the rabbit population increases further, the rabbits will starve and the population will decrease. If the population becomes smaller than the equilibrium value, there will again be a surplus of grass, and the population will increase again. Therefore, any deviation from the equilibrium population will be counteracted, and the system will remain close to the equilibrium. However, imagine now that the restriction is lifted. For example, more grass might become available, or the rabbits might learn to eat another type of plants in addition to grass. The negative feedback, which kept the rabbit population at the equilibrium value, will now again turn into a positive feedback, and the rabbits will multiply until they reach a new limit.

In general, such self-organizing processes will experience both negative and positive feedback. Negative feedback will stabilize equilibria, suppressing variation. Positive feedback processes will speed up variation, pushing the system to quickly reach a new attractor. This is what Prigogine called “order through fluctuations”. A fluctuation is a chance variation that is so small that it normally should have no effect. However, if the system is far from equilibrium, such tiny variations may be amplified by positive feedback. Thus, small causes can have large effects. Like the order from noise mechanism and the stepping-stone principle, positive feedback is a way to spectacularly increase the speed of evolution. The succession of episodes with positive feedback and negative feedback may explain the “punctuated equilibrium” pattern of evolution, where long periods of equilibrium (negative feedback) are interrupted by bursts of frantic activity (positive feedback).

Positive and negative feedback can be present simultaneously. Imagine that you are walking with your dog on the leash. Every step forward is reinforced by the dog pulling you forward: positive feedback. However, for every step sideward the dog pulls you back: negative feedback. This is the same kind of movement as the one we discussed for a pendulum: some changes are amplified, some are dampened. The result is that the system—instead of staying put in an equilibrium state—will continue to move along an attractor. For the pendulum, this attractor is simply a limit cycle. For more complicated systems, the attractor may have a very complex, even fractal shape. This leads to a trajectory which is at the same time unpredictable, and non-random. Thus, instead of sitting quietly at the lowest point of its fitness valley, a self-organizing system may exhibit the most complex types of behaviors. Some of the pattern-forming chemical reactions studied by Prigogine fall in this category. For example, a “chemical clock” is a reaction where two stages—say, one dominated by a blue coloring and one dominated by
a red coloring—alternate periodically. This can be understood as a limit cycle, where the concentration of a blue substance increases, then decreases while the red substance increases, then again increases while pushing out the red, and so on.

**The Cycle of Life**

The way I have explained it, selection through feedback may seem rather trivial: action-reaction. In practice, there will be more than two parties involved. For example, a chemical substance A may act on another substance B, which in turn acts on a third substance C, which then acts on a fourth substance D, and so on. Until now, this is just like a classical chain of cause and effect. However, imagine that at some point in the chain one of the earlier molecules appears again. For example, suppose that substance H acts in turn on substance A. The linear chain of causes and effects has now turned into a closed loop (see Fig.). Every action of A will—through the rest of the chain—feed back into A itself. The same applies to all other elements of the chain: B, C, D, ..., up to H. Each component will indirectly affect itself.
Again, this feedback relation can be positive or negative. There is a simple way to find out the overall “sign” of the causal loop. Suppose that for each step in the causal chain, for example $A \rightarrow B$, you can determine a sign, “+” for positive, “–” for negative. The relation is positive if an increase in $A$ also produces an increase in $B$ (or a decrease in $A$ produces a decrease in $B$). The relation is negative if an increase in $A$ produces a decrease in $B$ (or a decrease in $A$ produces an increase in $B$). For example, the effect of the number of rabbits on the amount of grass is negative: an increase in the number of rabbits will lead to a decrease in the amount of grass. Now, to determine the overall sign, positive or negative, for the loop, you simply need to multiply the signs for each of the steps. For example, the relation rabbits $\rightarrow$ grass is negative, while the relation grass $\rightarrow$ rabbits is positive (more grass provides more food to grow more rabbits). Positive times negative is still negative. Therefore, the overall sign of the loop rabbits $\rightarrow$ grass $\rightarrow$ rabbits is negative. This is a negative feedback relation. This means that without outside interference, the relation between rabbits and grass will reach a stable equilibrium.

Calculating the sign for a longer loop is extremely simple. Since multiplying with a positive value does not change the sign, either negative or positive, you only need to count the number of negative relations. Negative times negative gives positive. Therefore, an even number of negatives will cancel itself out, and the overall loop will be positive. On the other hand, any chain with an uneven number of negative links will be negative overall (see fig.). This gives us a simple way to determine whether a causal loop will lead to stabilization (negative feedback) or on the contrary, to a runaway, explosive process (positive feedback).

Positive loops are intrinsically risky, because they accelerate out of control. However, explosions remain rare events. That is because fast growth quickly runs out of resources. In that case, a positive feedback turns into a negative feedback, as we discussed earlier. The result is a cycle that is both intrinsically stable, and able to grow quickly the moment more resources become available. Stability means that it can resist most outside perturbations. If, for example, a large number of rabbits are killed by hunters, or by a flood, the population will quickly grow back to its initial level. If a fire destroys most of the grass, most rabbits will die from starvation, thus helping the grass to grow back more quickly, and feed a resurgent rabbit population.

The system formed by the different components connected together by such a feedback cycle is both flexible and self-reliant. It regulates itself, adapts to changing circumstances, and uses an opportunity when it encounters one. Under the right circumstances, it can even grow and multiply. This sounds suspiciously like the description of a living being. It will come as no surprise that the most popular theories
about the origin of life propose exactly such a positive feedback cycle of chemical reactions as the first “spark”, the precursor of all living organisms. Such chemical loops are called *autocatalytic cycles*. As we noted earlier, catalysis is the process where a chemical substance A stimulates the production of another substance B. This is a positive causal connection. You get an autocatalytic cycle when you chain together a series of catalytic reactions such that the end product is the same as the substance you started with. Such cyclical, non-linear reactions show many unexpected types of behaviors. The self-organization they exhibit inspired Prigogine to develop his theory of dissipative structures. Another Nobel prize winner, the biochemist Manfred Eigen, suggested that in order to develop primitive living systems, you need more than a single cycle. You would rather need cycles within cycles, so that some of the products participating in one cycle would also be regulated by another cycle. For example, the cycles $A \rightarrow B \rightarrow C \rightarrow A$, and $B \rightarrow D \rightarrow E \rightarrow B$ intertwine through their common component B. Thus, several autocatalytic cycles are braided together in a complex, self-regulating network. Eigen called such a network a *hypercycle*.

Many scientists have suggested that life has an intrinsically cyclic organization. The Chilean biologists Humberto Maturana and Francisco Varela formulated the defining feature of life as *autopoiesis*, which is Greek for “self-production”. Indeed, all living organisms produce their own components: their cells, organs, tissues and circuits. Since living matter is dynamic, it is not sufficient to produce these components once and for all: living tissues are produced, grow, decay and die, to be replaced by new components, in an on-going cycle that only stops at the death of the organism. In this view, the outside world is merely a source of building materials, in the form of food and oxygen, and a place to dump waste, in the form of excrements and carbon dioxide. The essence is the cycle of self-production, whose only purpose is to keep itself running. Therefore, Maturana and Varela call an organism “organizationally closed”. Although living organisms are open systems, with an input and output of matter and energy, their organization is determined purely internally. It is not the environment that instructs an organism how it should organize itself. As far as the organism is concerned, the environment is important only to the degree that its fluctuations may perturb the inner functioning of the autopoietic loop.

This property of *closure* characterizes all cyclical processes. Although cycles usually do need some input and output outside of the cycle, they are in an important respect self-sufficient—indeed, independent of the environment. That is what makes them fit, capable to maintain and grow in spite of external perturbations. Because they are in a way separated from the rest of the world, and because they behave very differently from linear
cause-and-effect chains, theorists such as Robert Rosen have suggested that living systems belong to a wholly different category, where Newtonian causality no longer applies. I would rather say that Newtonian principles don’t even apply to simple, linear systems. Moreover, the strict separation between linear, mechanical systems and cyclical, living systems seems to imply that the one could never evolve out of the other. Yet, the transition from a linear chain to a closed loop is an extremely simple, almost trivial event, which has happened countless times.

Just consider an unlimited causal chain, where A produces B, B produces C, C produces D, and so on, until the the end of time. It suffices that, just once, the chain would produce a product—say J—that was already present at an earlier stage (see Fig.). Since equal causes produce equal effects, the next step after J would again have to be K, just like when J first appeared. From that moment on, the chain would start repeating itself. Depending on the sign of the overall loop, the cycle would immediately stabilize, or grow until it ran out of materials—which would force it to stabilize too.

The probability of getting into a loop becomes much greater if we allow more than one effect to result from a cause, as is common for chemical or physical reactions. For example, substance A might produce substances B1, B2, B3, and B4. B1 might in turn produce C11, C12, C13 and so forth. The total number of products thus grows exponentially. With such a continuing explosion of new products, it would be very unlikely that there would never appear a product that had not appeared at some earlier
stage in the process. The moment such a repetition occurs, an autocatalytic cycle starts up (see Fig.).

The complexity theorist Stuart Kauffman has turned this intuitive idea into a more detailed mathematical model. He estimated the probability that certain products would catalyze certain other products. Using a computer program to simulate the chain of reactions, he found that once a certain threshold of complexity was passed, autocatalytic reactions were bound to emerge. The threshold basically depends on how many products are available to start with, and how likely it is that they would catalyze the creation of further products. The organic chemistry that existed on the Earth’s surface before the appearance of life was almost certainly complex enough to produce this endless generation of ever more complex autocatalytic reactions. Therefore, Kauffman concludes that the origin of life was not an accident: it was bound to happen. We just don’t know the exact details of the different reactions that produced the first “living” system yet.

Of course, the evolution of life required much more than the emergence of a closed cycle of reactions. The simplest living organism is orders of magnitude more
complex than an autocatalytic cycle. To understand more about the origin of life we will have to examine in more detail how evolution creates complexity.

The Growth of Complexity

Closing the Loop

We now have gathered all the concepts and principles that we need in order to explain the origin of complexity. The most important concept is the one of fit. Imagine two systems that interact, for example, two atoms. They will generally pass through different variations until they reach a mutual arrangement that is selected. This particular assembly will remain, without further variation. For example, a sodium atom and a chlorine atom suspended in water may interact until they reach a stable state, where the distance between the two atoms is fixed. The two atoms are now bounded together, and can no longer move independently. They form a new system: a molecule. The properties of this new system are in general quite different from the properties of its components. Both sodium and chlorine are very reactive, poisonous substances. The molecules they form by bonding together, however, are completely harmless. The substance produced by their chemical reaction is ordinary salt, the one we add to food, and which is an essential ingredient of our blood.

What has happened to cause such a dramatic change? The properties of a system are determined by the way it interacts with other systems. Sodium reacts violently with most other substances, including the substances that make up our own body. This makes it very dangerous, since such a reaction destroys existing arrangements. However, once such a reaction has taken place, for example a reaction with chlorine, the atom’s capacity for further reactions has been “used up”. The sodium atom fits snugly into its new assembly with the chlorine atom, and is no longer capable of interacting with other atoms. That is why salt, the result of that assembly, is quite harmless. Having reached a fit state, an attractor, the sodium atom has lost its capacity for further variation. It is now constrained by its new-found relationship with the chlorine atom. Of course, the salt molecule as a whole can still vary, and enter in different interactions with other molecules. But these interactions are of a different, less violent type.
It is because of this common constraint that we call the molecule a “system”. If we just put a few atoms or other components together, without them reaching a stable state, then we would not call the assembly a “system”. As long as the components can move freely with respect to each other, nothing fundamental has changed, and the assembly remains a mere aggregate, a collection of independent pieces. The moment the components reach a collectively stable state, on the other hand, the situation changes dramatically. The components will now directly depend on each other, and the one will no longer be able to move without affecting the others. They now all fit together. For example, mix some dry sand and cement. Nothing happens. The grains of sand can still move independently, and will separate the moment you pass your finger through the mixture. Then mix in some water, and wait. Now the result is very different: the reaction of the cement with the water has created a substance that glues all grains of sand together. You end up with a piece of concrete, rather than a mixture of cement and sand. No amount of pushing or pulling with your fingers will separate the grains of sand. The grains now belong to a coherent system: a block of concrete.

It is the shared constraint holding the grains of sand together that turns the assembly into a “system”. The constraint imposes order on the assembly. Each grain now has its fixed place, from which it can no longer move. If we know the constraint, we can predict where each grain of sand will be found in the block of concrete. Of course, the arrangement of grains in a block of concrete is rather irregular and therefore difficult to determine. We can find the same type of collective assembly in crystals, though. As we discussed earlier, salt molecules suspended in a solution will tend to settle down and fit into a crystalline pattern. Since all salt molecules are exactly the same, they will all tend to settle in exactly the same position with respect to their neighboring salt molecules. The distances and relative positions at which salt molecules fit together are always the same. The result is that the molecules will assemble themselves in a neat, regular grid.

The spontaneous assembly of components into stable systems is a universal principle, applicable to all components that can interact with each other. It does not matter whether the components are different, like the sodium and chlorine atoms, or the grains of sand and cement, or similar, like the salt molecules. It also does not matter what the components themselves consist of: they might be particles, atoms, molecules, cells, organisms, people, or organizations. As long as two systems interact, their variation relative to each other will tend to settle into a stable state, defining a new, emergent system. What does matter, though, is whether the new, coupled system can undergo further interactions of this kind. In the case of the sodium and chlorine atoms, the direct interactions between atoms have been “used up” in the salt molecule. No further atoms
can come and attach themselves to the assembly. In the case of the crystal, on the other hand, further molecules can always be added, making the crystal grow without bound.

A good way to understand the difference is the concept of closure. We already encountered this idea when discussing feedback cycles, and their role in the origin of life. A “closed” cycle of chemical reactions does not need any further chemicals to sustain itself: it produces its own components. Such a cycle is intrinsically dynamic: it constantly produces and consumes molecules. However, we can use the closure concept as well to describe static assemblies of molecules, or any other systems. Imagine that an atom can attach itself to two other atoms of the same type. When it encounters another atom, they will form an assembly, a molecule. However, for each of the two atoms in the molecule, there is still opportunity for another atom to join, say, one on the left of the assembly and one on the right. Thus, the molecule can grow to four atoms. But the two newly added atoms again have “space” each for another atom to join. Thus the chain of atoms can continue to grow and grow. This is the basic mechanism behind the growth of polymers: molecules in the form of long chains. If more than two atoms can attach themselves to a given atom, the assembly does not need to be restricted to a one dimensional chain: it can have complex two dimensional or three dimensional shapes. This is the mechanism underlying crystal growth.

But let’s go back to the chains. Suppose that the last atom of the chain attaches itself to the first one. Now both “free ends” have used up their capacity for bonding; each is attached to two atoms: the one to which it was attached before, and the one it joined at the opposite end. The chain has been closed, and turned into a loop. This makes it impossible for further atoms to join the assembly. The assembly has been “closed off” from the rest of the world. The creation of a closed loop could have happened at almost any stage of the process: after two, ten or thousand atoms had joined. However, imagine that there are certain restrictions on how new atoms join. For example, two atoms may only join such that they form a 60 degree angle between them. In that case, the number of closed loops is severely limited. The simplest one consists of 6 atoms, each turned 60 degree with respect to the previous one, so that overall they form a 360 degree, hexagonal loop. This is the structure of a benzene molecule (see Fig.), which consists of 6 carbon atoms (each with an additional hydrogen atom attached to it).
For a simpler example, you may remember the magnets that stick together after being shaken. For magnets to get attached, there are again restrictions: the south pole of one magnet is only attracted by the north pole of another magnet. If these magnets are rectangular in shape, they can only get attached at right angles (90 degrees) or at 180 degrees. The latter orientation will produce long straight chains of magnets. However, the 90 degree orientation can produce a very simple, closed assembly, where four magnets are arranged in the shape of a square (see Fig.). Such an assembly is quite stable, since each magnet is attached in two points. Magnets arranged in a chain are much more likely to disintegrate, since the ones at the ends of the chain are only attached on one side, and thus likely to fall off. This exposes the one but last magnet, which is again prone to get disattached, and thus the whole chain may lose its components one by one. Assemblies that are not closed always have “loose ends”. The advantage of these loose ends is that they provide an opportunity for further growth, since they can function as mooring points for further components. The disadvantage is that the loose ends are not as strongly attached, and therefore more likely to break off. Thus, closed assemblies are more stable than open ones, but less likely to grow.

More generally, an assembly is closed if its components have “used up” their capacity to get attached to further components. This means that elements outside the assembly can no longer interact with separate components. The components no longer react individually.
They all move together. Since distinguishing phenomena means interacting with them, you can no longer distinguish the individual components from the outside. You can only distinguish the assembly as a whole. That is another reason why we call the assembly “closed”: it is as if its components are hidden behind a wall or boundary that separates them from the rest of the world. Therefore, we call the closed assembly a system, that is, a coherent whole, separated from its environment by a boundary.

The Architecture of Complexity

The fact that individual components have stopped interacting with the environment does not mean that the system has become sealed off. The system as a whole will still interact with its environment through its input (external events affecting the system) and its output (internal events affecting the environment). These interactions can again give rise to fit linkages between systems, where the interaction has stabilized so as to connect systems together into higher order assemblies. If such an assembly reaches closure, it will determine a system of a higher order. This supersystem will contain the original assemblies as subsystems. Such supersystems can in general get interconnected themselves, forming assemblies of a yet higher order. This produces an ever higher “pyramid” of systems and supersystems. This explains the hierarchical “boxes within boxes” architecture, that is typical of complex systems (see Fig.).
For example, the human body consists of organs, which consist of cells, which consist of organelles, which consist of polymers, which consist of monomers, which consist of atoms, which consist of elementary particles. Each of these levels has an approximately closed organization. The organelles within a cell form a tightly integrated whole, that is separated from the surroundings by a membrane. Therefore, you might say that they fit together much more strongly than cells fit together with other cells. Similarly, the monomers in a polymer fit together more strongly than the polymers in an organelle, or the organelles within the cell. At the lowest level, the quarks in an atom fit together much more strongly than any other system we know. In general we can conclude that the lower the level, the stronger the fit, the more stable the assembly. The general principle is that closed assemblies are like stable building blocks, out of which larger blocks are assembled.

Thus, the hierarchical architecture of complex systems, with their distinct levels, can be explained through the stepping stone principle. Different components undergo
relative variation, until they hit upon an arrangement that is stable. This is a sudden, discontinuous transition, a step towards higher fitness. The resulting assemblies continue to vary relative to each other, again until they reach a stable configuration. This is another discrete step resulting in a more fit, but also more complex, system. This higher order configuration could not have arisen without the prior emergence of its less complex subassemblies. For example, molecules could not have evolved without the prior emergence of atoms, and complex organisms could not have evolved without the prior emergence of cells. Each of these levels of complexity—particles, atoms, molecules, polymers, cells, etc.—is a stepping stone in the evolution towards ever more complex systems.

This insight, that the stepping stone principle underlies the hierarchical organization of complexity, is due to Herbert Simon (whose safecracker example we used to originally illustrate the stepping stone principle). However, Simon’s original proposal was more limited, because it lacked the concept of closure. Developing the safecracker argument, Simon concluded that assemblies could only evolve if they were very simple, that is, if they contained few building blocks. Simon argued that the more blocks you need to produce a stable assembly, the smaller the probability that these blocks would arrange themselves in the right configuration. But this assumes that the building blocks would have to fall in place all at once, which is indeed a very unlikely event. Our example of the growing polymer, on the other hand, makes it clear that building blocks can be added one by one, and that this process can go on indefinitely, until it reaches closure. Moreover, the process may be characterized by positive feedback, so that the addition of building blocks goes faster and faster. For example, a salt crystal growing in a solution becomes larger and larger, and therefore it provides more and more space for further salt molecules to get attached. Thus, the number of building blocks in any given stable assembly can become very large indeed.

Simon’s reasoning can only explain very “steep” pyramids, where there are many levels and few subsystems on each level. Our present argument can also explain “flat” pyramids, with few levels and many subsystems. However, the multiplication of subsystems does not really make the system more complex. Indeed, adding subsystems, like adding molecules to a crystal or to a polymer chain, is in general mere repetition of a pattern. Complexity, as we defined it, is *distinction plus connection*. Adding links to a chain may increase connection, but as long as the links are all the same, it does not increase distinction. What really creates complexity is the nesting of closure within closure: the components in a closed assembly are tightly connected, but at the same time
distinguished from their surroundings. In that sense, Simon’s argument is really the beginning of a theory to explain the evolution of complexity.

The closure concept can even explain system architectures that are not strictly hierarchical. The traditional idea of a hierarchy is that for every group of “subordinates” there is a unique “supervisor” or “next-in-command”. In systems terms, this means that a system can contain many subsystems, but can only belong to a single supersystem. In real life, systems can take part in different supersystems. For example, John can at the same time be an employee of a particular firm, a player in a football team, and a member of a family (see Fig.). All these are supersystems of the same subsystem. We could say that the supersystems, although different, overlap through their shared subsystem.

![Fig.: Overlapping systems: John is simultaneously a component of a family, a firm, and a football team.](image)

Such “multiple allegiances” can still be seen as closures if we consider different types of interaction. There is a particular way in which employees of a firm interact, and there is a totally different way in which football team players or family members interact. Each of these interaction types can reach closure. For example, employees will only pass confidential business information to employees of the same firm. Football players will
only pass the ball to members of the same team. Such interactions do not leave the system. They therefore define the system as a closed entity, with a clear separation between the “in” and the “out”, between “us” and “them”. But because the same component, in this case John, can engage in different interactions, that component can be part of different closed assemblies.

In practice, closure is never complete: only part of the interactions will remain inside any given system. As we noted earlier, a system that does not interact with the outside world might as well not exist. Closure, albeit incomplete, is a useful concept because it allows us to make distinctions by grouping closely interacting components together. Since the same components can undergo different types of interaction, they may take part in different groupings. However, once a particular type of interaction has reached closure, it will put a constraint on the component, thus restricting its freedom to vary. This will make it more difficult for the remaining types of interaction to reach closure. Moreover, closure will typically bring together similar components, that interact in the same way. If one of their shared interactions has reached closure, thus binding the components tightly together, the remaining interactions will no longer distinguish between the individual components, but rather act on the system as a whole. For example, protons (a type of elementary particle) will interact through the “strong” nuclear force to form the nucleus of an atom. That nucleus will then interact via the remaining electromagnetic force with electrons to form an atom. However, the electrons are bound to the nucleus as a whole, not to individual protons. Therefore, all components of the nucleus are also components of the atom. The nucleus and the atom are not overlapping supersystems: they are nested (see fig.).
We may conclude that Simon’s proposal for a hierarchical architecture of complexity is in many cases a good description. Most physical, biological and social systems indeed consist of boxes nested within boxes. We must just remember that hierarchies can also be flat, rather than narrow, and that supersystems sometimes overlap, especially when their components are capable of many different types of interaction. We may call the resulting architecture a “quasi-hierarchy”: it is almost, but not completely, hierarchical. Similarly, the systems at the different levels are mostly, but not completely closed. That is why Simon calls them “nearly decomposable”: you can analyse or decompose a system into individual subsystems, but then you will lose the interactions that tie them together. But since the interactions inside a system are normally stronger than the interactions that connect it to other systems, what you will lose will be less stable or less fit that what you will keep. That explains why Descartes’ method of analysis is useful most of the time. But it also explains why analysis must be complemented by synthesis if you want to get a complete picture of a system.

Finally, Simon’s argument explains why complexity grows during evolution: whatever the level you start with, on-going variation and selection will tend to group systems at that level into closed assemblies at a higher level. This adds both distinction
and connection, and thus complexity. Another way to see this process is integration of systems into a supersystem, followed by differentiation between the new supersystem and its environment. Thus, the relative variation of systems and selection of closed assemblies leads to the spontaneous, on-going creation of complexity.

**In Control**

The emerging complexity we have discussed until now is purely static. Systems are bound together once and for all by fit linkages, forming a stable supersystem. This allows us to explain the evolution of atoms from particles, molecules from atoms, polymers and crystals from molecules, rocks and clouds from polymers and crystals, planets and stars from rocks and clouds, solar systems from stars and planets, and galaxies from solar systems. All of these are basically rigid, unchanging, “lifeless” systems. To explain active, “living” systems we need to go back to dynamic closure.

Remember how so-called “autocatalytic” cycles of chemical reactions can be seen as the precursors of life. By producing their own components, these cycles become to some degree independent of their environment. This makes them into organizationally closed systems, distinct from their surroundings. However, an autocatalytic cycle typically requires many different chemicals to sustain the reaction. Some of these it produces itself. Most others, however, must be provided by the environment. We have shown that such cycles will arise spontaneously as soon as a sufficiently large variety of different chemicals is present. However, the more chemicals to be produced by the cycle, the smaller the probability that such a cycle would arise. Variation will normally only create relatively simple cycles, that produce few chemicals, and that depend on the environment for all other chemicals.

We also argued that the interplay between positive and negative feedback will tend to stabilize the cycles, so that they can survive fluctuations in their environment. If the input decreases, the cycle will slow down and produce less chemicals. If it increases, it will accelerate and produce more. However, because they remain dependent on their input, autocatalytic cycles will never be as stable as closed, static assemblies, such as molecules or crystals. It is enough for one of the input chemicals to be temporarily unavailable for the cycle to break down. Another danger is that one of the input chemicals might be substituted by a chemical that reacts differently. This will produce chemicals different from the ones needed to sustain the cycle, and thus abort the cycle. The substitute behaves like a poison.
If we imagine different cycles trying to survive and grow in the same environment, the probability of such perturbation seems rather high. One cycle may absorb all the input chemicals (“food”) that another cycle needs for its survival. Similarly, the output chemicals of one cycle (“waste products”) may act as a poison for the other cycle. If cycles emerge easily, there will be many of them. If many cycles compete for the same food molecules, few will manage to survive. To survive for a longer time, an autocatalytic cycle needs more than negative feedback: it needs a reliable mechanism for neutralizing the perturbations in its environment. As we noted earlier when discussing cybernetics, counteracting perturbations defines the problem of control. You are in control when you can achieve your goals in spite of various disturbances that would otherwise make you deviate from your preferred course. For an autocatalytic cycle—as for any evolutionary system—the fundamental goal is survival and growth, that is, fitness. The disturbances it needs to compensate are lack of food and presence of poisons. There are two basic mechanisms that will help a cycle achieve control.

The simplest and most obvious method to reduce perturbations is to create a protective wall or encasing that will keep poisons and other dangers out, and food in. For example, the shell of a tortoise will protect the tortoise from predators and violent blows, will still allowing the animal to eat and to breathe. For simple, unicellular organisms, such a semi-permeable barrier is called a membrane. The membrane surrounds and protects the chemical processes that take place inside the cell.

It is not particularly difficult for an autocatalytic cycle to evolve a membrane. It suffices that some of the molecules produced by the cycle would stick together, forming a two-dimensional sheet that surrounds the space where they originated. As more of these molecules are produced, the sheet grows, creating more space for other molecules to react inside the sheet. This allows the system to build up an internal reserve of “food” molecules, which is safe from competitors, and which can be used when outside food becomes scarce. Chemically, there exists a simple type of molecules, “bilipids”, that will spontaneously form such sheets in a watery solution. This would turn the autocatalytic group of molecules into a primitive cell. When the number of bilipids in the cell becomes large enough, they can even produce an internal wall, splitting the cell into two, independent parts. This is the beginning of biological reproduction.

Membranes or walls reduce the effect of perturbations, and thus make the system more stable. They moreover separate autocatalytic systems in space, so that they can develop without interfering in each other’s activities. They do not provide active control, though. Negative feedback is a more active mechanism, that constantly adjusts the various concentrations of molecules back to their equilibrium level. However, as we
noted when discussing the example of supply and demand, this equilibrium itself depends on outside factors that vary uncontrollably. Ideally, the system would adjust the different concentrations to the values that are best for the system itself, independently of the outside situation. These optimal values constitute the system’s own goal. To reliably achieve this objective, the system needs some further mechanisms. First, it must have an accurate memory or representation of its goal, so that however severely it has been perturbed, it knows what to go back to. Second, it must be sufficiently sensitive to recognize perturbations before they have become too large. Third, it must be powerful enough to effectively suppress external influences.

These requirements do not only apply to autocatalytic cycles: they are universal. You are in control if you know what you want (have a goal), if you have the power to overcome the obstacles that keep you from achieving that goal, and if you are sufficiently perceptive to discern those obstacles. The more sensitive you are, the quicker you will recognize any possible deviation from your goal, and the less power or energy you will need to use in order to correct that deviation. Thus, sensitivity can to some degree compensate for lack of power: if you are not strong, you need to be perceptive!

This simple understanding of control was analysed in great detail by the theorist Bill Powers. Powers sees a control system as a negative feedback loop with two inputs: external disturbances, and an internal goal (see Fig.). Both inputs influence the state of the system. The overall result will be a give-and-take between the effects of the disturbances and the effect of the system’s own desires. The system will be in control if it can impose its own standards, that is, if the result it gets is closer to its goal than to the disturbances.
Let us explore the simple control scheme in Fig. 1: a control loop according to Bill Powers: the external disturbances together with the system’s own actions determine the situation that the system perceives. The difference between this perception and the system’s goal determines the system’s subsequent action to correct the situation. The disturbance brings about a situation which is in general different from what the system wants, that is, from the system’s goal. The system must first perceive that situation, compare it with its goal to determine how large the deviation is, and then take a counteraction that will suppress the deviation, and bring the situation back to the goal. The larger the disturbance, the larger the deviation, and therefore the more forceful the action must be. If the action is forceful enough, the deviation will be suppressed and the goal will be restored. It is not grave if the action is too forceful: since the feedback loop is negative, an overreaction will merely create a deviation in the opposite direction, which will itself be suppressed. Using simple calculations, Powers showed that the stronger the action the better, since it will bring the situation back more quickly to the goal. However, to function well, this requires great sensitivity on the side of perception: if the action is too strong, the system should immediately sense that it is going too far. Only then can it stop or reverse the action before the deviation has grown out of control.
The example of the thermostat, which we used to illustrate the principles of cybernetics, will clarify the relation between power and sensitivity. Suppose you have a powerful heating installation: it can heat a cold room in a very short time. The moment the thermostat senses that the temperature drops below the goal temperature, it switches on the heating. This makes the temperature increase very quickly. Suppose that the thermostat would not be very sensitive to temperature changes, and therefore slow to react. In that case, the room would become too hot before the thermostat would switch off the heating. Similarly, after the heating was switched off, the room would cool down a lot before the thermostat would take corrective action. A sensitive thermostat, on the other hand, would switch the heating off and on very quickly. In practice, the combination of sensitivity and power would keep the room very precisely at the desired temperature, independently of the outside weather.

Another way to understand the relation between sensitivity and power is what Powers calls “amplification”. If you follow the signal that cycles through a control loop, you will notice that a small deviation is perceived and immediately compensated by a strong counteraction. The signal is very small on the perception side of the loop, but very large on the action side. This asymmetry is what distinguishes control from mere negative feedback. If both signals had a similar value, there would not be any fundamental difference between the goal and the disturbance. Both inputs would affect the looping signal to the same degree, and the system would not be able to impose its “will” onto the signal. To achieve real control the loop must function as an amplifier, boosting small sense signals into large action signals. This requires very finely tuned sensation, but also a large store of energy to draw power from.

Note that in the control scheme sensitivity is more important than power, since a very sensitive controller does not need much power to counteract the as yet small deviations it senses. A powerful controller that lacks sensitivity, on the other hand, will in general “overshoot” its target, and this may increase rather than decrease deviations. Not only sensitivity can compensate for lack of power: if you can reduce the strength of perturbations, you will need less strength in the counteractions. One way to achieve this is a protective wall, as we discussed before. The thickness of your house’s wall determines how well your room is insulated against changes in outside temperature, and thereby how much energy you need to spend in order to keep the room at its ideal temperature. Another example is the reserve or store of food: the larger the reserve, the less the system will be affected by external fluctuations in food availability. Both are examples of a buffer: a mechanism that dampens the effect of fluctuations without
actively intervening. Although a buffer is in itself not a control mechanism, it makes control much easier to achieve.

Let us apply our new understanding of control to the origin of life. In an autocatalytic cycle, you could support a stable goal or target by a special type of catalysing molecule. Such a molecule would affect the cycle of reactions, but would not itself be affected by it. Because this molecule remains intact even when the cycle is interrupted or severely disturbed, it can function as a reliable standard, a reminder of how things should be. In the living systems we know, this role is played by DNA. DNA can store huge amounts of information, containing a blueprint for all the different chemical processes that produce a complex organism. If any of these processes is interrupted, for whatever reason, the DNA preserves all information needed to start it up again, or to start up a wholly new process tailor-made to the particular circumstances.

The second requirement for control, sensitivity, is implicit in the positive feedback. Indeed, positive feedback means amplification of small deviations into large changes. Similarly, amplification will provide power, since it can produce strong counteractions. Of course, amplification requires energy, and this means a constant input of “food” molecules that can be digested and stored in an energy-rich form. To achieve control, though, the overall sign of the feedback loop must be negative. This means that the amplified action must itself be suppressed as soon as it overshoots its target. We noted that positive feedback turns into negative feedback when it runs out of resources. However, the whole point of control in a living system is not to run out of food or other vital resources.

There is another way that lack of a “resource” can dampen positive feedback, though. Imagine that the cycle must be catalysed by DNA in order to continue. DNA can be in an “active” or in an “inactive” state. If it’s active the cycle continues and more “actions” are produced. If it’s inactive, it is still there but it does no longer participate in the reaction, thus making the cycle halt. Suppose that the goal is to keep the concentration of a certain chemical at its ideal value. If the concentration is too low, the DNA will be active and help produce more of that chemical. However, when the concentration becomes too large, the “overdose” of the chemical will trigger a reaction that deactivates the DNA. This stops the further production of the chemical, and lets its level stabilize at the ideal value. Thus, a chunk of DNA can represent a target value, controlling the presence or absence of different chemicals in the cell, while remaining itself in constant supply.
Knowledge and Action

Powers’s simple picture of a control loop has allowed us to understand how an autocatalytic cycle can turn into a living organism. Control explains how directed action can make an organism autonomous, independent of the fluctuations in its environment. Instead of remaining passive, like a crystal or a rock, an organism constantly adapts, changing its inner configuration and taking appropriate action to salvage its goal of survival and growth.

The picture moreover helps us to understand how variation and selection can lead to the emergence of life. Variation will constantly produce different values for the properties of the control loop: amount and type of buffering, power or sensitivity, choice of the specific target values, etc. Some of these values will lead to better control. Organisms with better control will be better able to cope with disturbances or changes in the availability of food. They therefore will survive and grow more easily and thus become more numerous. In the beginning of life on Earth, such variation would typically affect all the different types of molecules that took part in an autocatalytic cycle. By the time DNA appeared, however, the metabolic cycle stabilised under its control, and only DNA would continue to vary. This is the way biological evolution still functions nowadays: what actually evolves is the control structure of an organism, not so much its anatomy or appearance.

To understand this better, we must expand our notion of control. The present picture is still too simple to fully explain the complexity and adaptability of life. When discussing artificial intelligence, we defined a problem as a difference between the present situation and the desired situation. Problem-solving means finding a sequence of actions that leads from the present situation to the goal. For the kind of control systems we discussed until now, problem-solving is trivial: if the temperature is too low, you heat; if the concentration of chemical X is too low, you boost the reaction that produces X. In these cases, there is little difficulty in choosing the right action. When the perturbations become more complex, though, there won’t be such an obvious recipe for action. The organism is then confronted with two difficulties.

First, it must be able to perform the appropriate action. For a weed that is going to be pulled out and eaten by a buffalo, the most adequate action might be to run away, but plants are simply not capable of running. For a buffalo dying of thirst in a muddy plain, an appropriate action might be to grow roots and extract water from the soil, but buffalos are incapable of growing roots. Second, assuming that the organism has the necessary capacities, it must still choose the right action to execute. If a hedgehog confronted with a
fox that would like to eat it, curls up into a ball, that action is likely to save its life. If the hedgehog performs the same action when confronted with a speeding car, that action may well cost it its life. On the other hand, if the hedgehog would run away from the car (and hedgehogs can run much faster than you might think!), it would survive unscathed. In sum, an organism confronted with a problem must have sufficient capacity for action, and sufficient knowledge to select the right action. Let us look in more detail at these two requirements.

The first requirement is merely an expression of Ashby’s “Law of requisite variety”, which we encountered when discussing the order from noise principle. Put more precisely, Ashby’s law states that, to be in control, the variety of actions a control system is able to perform must be at least as great as the variety of perturbations that it is likely to undergo. For example, suppose you are driving a car. Your goal is to keep the car on the road, independently of curves, gusts of wind, approaching vehicles or other disturbances that make you deviate from your course. Imagine that your steering wheel is stuck so that it can only turn to the left. If the road curves to the left, your left turn would allow you to compensate for the deviation, and keep you on course. If the road would suddenly turn to the right, though, you would be unable to take the correct action, and as a result you would end up in the ditch. A car that can only turn leftwards does not offer the requisite variety of action to drive safely. Similarly, a car without brakes does not provide the necessary variety to control your speed.

The requisite variety of actions will depend on two things: the environment and the goal. Environments that are complex and variable are likely to put up many different obstacles and opportunities for the organism. For example, the arctic environment may be harsh for the ice bears that live there, but it is simple and predictable. The temperate forests with their many species of plants and animals are a much more complex environment for the ice bear’s relatives that live there. To survive, an ice bear only needs a single basic type of action: catching and eating seals. The brown bear needs a much larger variety of skills to provide in its food: fishing salmon, collecting berries, finding honey, etc.

The second thing to determine the variety of actions is the organism’s specific goal, its way of making a living. Every evolutionary system strives to survive and grow, but there are many different strategies that lead to this common goal. Every strategy determines a number of more specific subgoals. For example, a weed in the forest tries to survive and grow by collecting enough minerals and water from the soil, and light from the sun. There are relatively few ways in which changes in the environment can perturb these functions. Lack of water or minerals can trigger the weed’s roots to grow deeper.
Sunlight being blocked by the leaves of competing plants may force the weed to grow taller or in a different direction. The requisite variety of actions for the weed is relatively small. A bear roaming freely through the same environment, on the other hand, will encounter a much larger variety of opportunities and dangers, that require a much larger variety of actions to turn them to the bear’s benefit.

The combination of environment and goal that determines the requisite variety of actions can be summed up by the biologists’ term *niche*. A niche is the segment of the environment that a particular type of organism interacts with, and whose resources it needs to survive and grow. The organism does not care about the aspects of the environment with which it does not interact. Most of the time it will not even be aware of these aspects. The weed is not aware of the bear that passes next to it. The bear most likely does not notice the weed. Natural selection will produce organisms sensitive to those aspects of their environment that may either help them or hinder them in their struggle for survival. As we saw, sensitivity to deviations is the first step towards achieving one’s goals. Sensitivity to other aspects does not provide any benefits. It is therefore extremely unlikely to evolve. Physically, the bear and the weed, the hedgehog and the mushroom, the snail, the oak and the robin may all share the same forest. Subjectively, though, they live in different worlds. The biologist von Uexküll coined the term “Umwelt” (German for “surrounding world”) to describe the individual reality experienced by each species.

Just like every type of organism needs its own repertoire of actions to cope with problems, it needs its own ways of sensing what constitutes a problem or perturbation. The only thing more it needs to be in control is a way of mapping its sensations to actions. That is where knowledge comes in. As we noted when discussing Ockham’s Razor, all knowledge can be expressed as a collection of “if...then...” rules. In the simplest case, these rules tell the organism directly what to do: if deviation A is sensed (for example, *it is too cold*), then perform action B (for example, *heat*). In a more complex case, a sequence of rules is needed: if A is sensed, then B is the case; if B is the case then C is the case; if C is the case, then D is bound to happen; if D is to happen, then perform action E. Such a chaining of rules is called “inference” or “deduction”. In an even more complex case, different sensations and rules may be triggered in parallel, and combine with each other in order to conclude into an action.

Sometimes, different sensations and different rules will trigger different actions. If these actions cannot be performed at the same time (for example, running away and curling up into a ball), the organism must decide which option to take most seriously. To help such decision-making, different rules will typically have different strengths or
degrees of reliability. Some rules will be practically certain (for example, if a heavy object is dropped, then it will fall), others will be merely suggestions (for example, if a berry is red, then it can be eaten). Depending on the strength of the sensations and of the different rules that support them, different recommendations for action will garner different strengths. The organism will then choose the action that receives the overall largest strength.

These different strengths of rules give us a hint of how knowledge could have evolved. We can view an organism without knowledge as one that does not have any preferences for actions. All possible rules connecting sensations directly or indirectly to actions have the same strength. Whatever its sensations, the different possible actions seem equally appropriate or inappropriate. In those circumstances, it can only try out an action at random and hope that it will turn out to be adequate. If it is inadequate (for example, curling up in front of an approaching car), the organism will be killed. Imagine that different individuals, because of variation, have somewhat different preferences for action. Those whose preferences tend to be more adequate are more likely to survive and have offspring. That offspring will inherit those preferences. Thus, fit preferences will spread, while unfit ones will get eliminated. In other words, natural selection will select the right type of rules. These rules themselves will select the right type of actions. If the rules did not eliminate inadequate actions, natural selection would eliminate the organism that followed those rules. Thus, the knowledge embodied in the rules works as a “stand-in” for natural selection. Donald Campbell, the evolutionary theorist, saw knowledge as a vicarious selector: a delegate mechanism, representing or substituting for the effect of natural selection.

Knowledge or vicarious selection makes an organism immensely more effective at solving problems. Instead of having to try out actions blindly, it can now make informed decisions. It can consider the different alternatives, and choose the one that receives the best overall support from its different rules. Those rules can be trusted because they have been selected for their adequacy. This makes the organism much more likely to survive and thrive in an environment that constantly confronts it with new problems. Rather than running the risk of being killed each time it tries out a possibly inadequate action, it can plan its actions with confidence. Blindness and uncertainty have been replaced by foresight.

Given the spectacular improvement in problem-solving ability brought about by knowledge, it is difficult to imagine how complex problems could ever have been solved without it. Yet, we must not forget that all knowledge is itself the result of blind variation and natural selection. By producing living systems capable of goal-directed and
knowledge-supported action, natural selection has provided a shortcut for its own slow and meandering operation. As we will discuss in more detail now, this led to an enormous acceleration of evolution.

**The Red Queen Principle**

We have seen that living organisms are control systems built on top of an autocatalytic cycle. The cycle constantly grows and renews its own components, ingesting food, and extruding waste products. The control mechanism copes with perturbations, such as changes in the availability of food or presence of poisons and other dangers. It does this by accurately sensing the nature of the perturbation, and applying its knowledge to choose an appropriate counteraction. The larger the variety of actions the system is able to execute, the larger the variety of perturbations it can cope with. This implies that evolution will tend to increase the variety of actions. Indeed, imagine two organisms A and B. Assume that A’s repertoire contains the same actions as B, plus one. In most circumstances, this additional action won’t make any difference. However, if a situation arises for which none of B’s actions is appropriate, for example a new type of poison or a new type of food appears in the environment, then A’s action may be just the one needed to tackle the problem. In that case A will thrive, while B may be eliminated. Thus, A, the organism capable of more diverse actions, will be selected.

This example may seem rather theoretical. However, precisely this kind of phenomenon has been observed many times with bacteria. If bacteria adapted to a particular environment are put in very different environment, most of them will not survive. However, through a mutation, some of them may have developed a new capacity that will allow them to thrive. For example, bacteria that can digest a particular type of food, will starve when put on glass plate that only provides a different type of food. A mutation in one of the bacteria’s DNA may now enable it to produce an enzyme capable of digesting the new food. That bacterium will not only survive, but grow and reproduce quickly in an environment that provides plenty of food. Its brothers and sisters, meanwhile, will die of starvation, and their place will be taken by the mutant’s offspring.

The more complex and variable the environment, the larger the variety of actions an organism needs to be successful. Even if the environment is relatively simple and stable, a major shakeup now and then, such as extreme weather, famine or epidemic, will eliminate all organisms that lack the requisite variety. Since no environment is perfectly stable or predictable, this means that an organism would do well to have the maximum variety that it can sustain. Thus, natural selection will foster variety of action. However,
there are limits to an organism’s internal variety. These limitations have two different origins: the material support and the knowledge needed to choose the right action.

Let’s first look at the physical support. To perform an action, an organism needs to have the right kind of “organ” or “tool”. For example, to lift a weight you need particular muscles, such as biceps and triceps. For a bacterium to digest a certain type of food, it needs the right kind of enzyme to break the food molecules apart. In systems theory, such “organs” used by a control system to perform an action are called effectors. Complementarily, before it can take the appropriate action, the organism must sense the problem. For example, the bacterium must recognize the unusual type of food before it can activate the appropriate enzymes. This requires appropriate sensors. Sensors and effectors are the concrete organs or molecular structures that an organism needs to realize its variety of actions. Increasing the variety of actions therefore means increasing the number and diversity of sensors and effectors.

To contain more of such organs, the organism must grow in size and in complexity. In particular, its DNA must grow in length, because each new molecule or organ must be coded by its particular stretch of DNA. A larger and more complex organism will require more resources to grow and maintain. All of these organs must be maintained in good working order, and be protected against perturbations. Thus the organism will become physically more vulnerable and dependent on its environment.

Moreover, its complexity may make it more vulnerable on the organizational level. Indeed, the larger the variety of molecules or organs, the higher the probability that these different functions would interfere with each other. For example, an enzyme useful to digest a particular food may react with another enzyme that functions to neutralize a particular poison. This reaction may cripple the enzymes or even create dangerous products as side-effects. The interaction between organs can also be more indirect: the effect of one action on the environment may change the situation in such a way that another action is no longer appropriate. Computer programmers have learnt a long time ago that the more complex a program, the more difficult it becomes to avoid “bugs”: unforeseen effects that appear only in very specific combinations of circumstances, but that may completely block the computer. Similarly, complex organisms are more likely to develop “bugs” in their operation. While natural selection will eliminate buggy organisms, this may take a very long time, since the particular circumstance that triggers the bug may be very unusual.

More generally, a larger variety of possible actions will simply make it more difficult for the organism to decide on the right sequence of actions, that is, to coordinate its actions. The reason is that the number of possible action sequences increases
exponentially with the number of actions. Remember the example of the safecracker: without help from the “clicks” to show that he is on the right way, finding the right sequence of 6 numbers to open the lock would require going through 1 million possible combinations. In many cases, there will be the equivalent of a “click” when the organism searches for the right action, but in some situations, the actions will depend on each other in a subtle way, requiring the organism to consider all possible combinations of actions.

In conclusion, when considering the evolution of an organism, we can expect a trade-off between the fitness gained because of increased variety, and the fitness lost because of greater complexity of functioning and the resulting difficulty of making decisions and avoiding bugs. In practice, the organism’s complexity will increase up to the trade-off point and remain more or less stable from then on. The trade-off point will depend first of all on the complexity of the environment: the more opportunities and dangers the environment presents, the more selection there will be for complex functions. Let us then consider which factors influence the complexity of the environment.

Remember our discussion of macroscopic causality: we can predict effects from their known causes because these effects are largely independent of the unknown background factors. If the background were more variable, macroscopic causality would not work, and it would be impossible to make predictions. The relative stability of background and known causes can be explained by natural selection. Macroscopic causality is not only necessary for scientists to make predictions, but for organisms to know which action to take when confronted with a problem or opportunity. Without some degree of predictability, control would not work, because each situation would be unique, and an action that was successful in the past may be wholly inappropriate in the present. For the same reason that macroscopic causality works, we may conclude that an organism’s environment will be stable and therefore predictable most of the time. Deviations from this “normal” situation will be rare. Thus, an organism can initially do well without much complexity. Through evolution, it will acquire additional sensors, effectors and rules, so that it can cope even with the more unusual perturbations. If the environment would remain the same, this increase in complexity would slow down as the more common perturbations have been taken care of, and the selective pressure to adapt to the remaining perturbations diminishes. It would finally come to a halt at the trade-off point.

The problem, though, is that an organism’s environment largely consists of other organisms. These others include predators that try to eat you, prey that is eaten by you, parasites that live off your body, competitors that consume the same resources as you, and finally “cooperators” or symbiotes that exchange resources with you. All these
opponents and companions will themselves become more complex in their actions. This means that you will need more complexity too. For example, if your competitor discovers a new way to make use of an opportunity, that opportunity may pass you by, and you may lose the competition. If your prey discovers a new action that will stop you from eating it, you will have to find a counteraction, or you may starve. If a predator develops a new technique to catch you, you will have to devise a counter-strategy or you may be eliminated.

Every advance made by one of the organisms in the ecosystem puts pressure on the others to advance as well. This is a general evolutionary principle. It does not only apply to the variety of actions, but to any improvement or fitness increase. The biologist Van Valen called it the “Red Queen” principle. The name comes from Lewis Carroll’s novel “Through the Looking Glass”, the sequel to “Alice in Wonderland”. In the novel, the Red Queen observes to Alice that “in this place it takes all the running you can do, to keep in the same place.” This is an accurate description of what happens in co-evolution. If an organism would not advance at the same rate as its evolutionary competitors, it would lose fitness, since the competitors would get away with an ever larger share of the common resources. To remain at the same fitness level, the organism needs to advance at least as fast as its opponents. Thus, continuing progress is necessary just to keep in place relative to the others.

A typical illustration of the Red Queen Principle is the “arms race” between predator and prey. Two countries engaged in a cold war must constantly increase their armament just to maintain the balance of forces. Similarly, if a prey species improves its defenses, then the predator must improve its offensive capabilities. For example, if rabbits learn to run faster so that they can better elude their pursuers, foxes too will have to run faster if they want to keep up their rabbit diet. While faster rabbits will merely maintain their fitness relative to faster foxes, they are likely to increase their fitness relative to other dangers. For example, a faster rabbit is less likely to be killed by an avalanche or a falling tree. However, more typical arms races will decrease overall fitness. The reason is that, like all positive feedback loops, they tend to exhaust resources. This is obvious for an arms race between countries: more money spent for military build-up means less money available for health care, education or infrastructure, and therefore a society that is overall weaker.

An example from biology is the arms race between trees. Trees in a forest compete for access to sunlight. If one tree grows a little taller than its neighbors, it can capture part of their sunlight. This forces the other trees in turn to grow taller, in order not to be overshadowed. The net effect is that all trees become taller and taller. Yet, on
average they still gather the same amount of sunlight, and thus they get no benefit from their effort. On the other hand, the increase in height requires more resources (water, minerals, carbon dioxide, etc.) for its build-up and maintenance. These can no longer be used for other purposes, such as producing more seeds. Tallness also makes the tree more vulnerable to storms, earthquakes or lightning. The result is that taller trees lose overall fitness while trying to keep up their fitness relative to their competitors. Yet none of the trees can afford not to participate in the race for ever increased height: if one tree would remain at what would be its ideal height without competition, it would be completely overshadowed by its neighbors.

Not all arms races are negative. Arms races for increased efficiency, that is, using less resources to get the same results, will be to everybody’s benefit. This is one of the advantages of market competition where companies manage to produce goods for ever lower costs. But it seems that all arms races, while accelerating evolution, must come to a halt at some point. This is obvious for arms races that waste resources: they reach their limit when all available resources are used up. But also arms races that increase efficiency must reach a limit of maximal efficiency: you cannot produce goods without using any resources!

But do these limits apply to the race for complexity? We just saw that there are practical obstacles to increasing complexity. When first defining it, though, we noted that there are no in principle limits to complexity. Matter and energy are conserved: although you may use them more efficiently, you cannot create more of them than there is. Complexity, on the other hand, is not conserved: you can increase it without bound. The question then is whether the pressure to increase complexity will be strong enough to overcome the obstacles. This means that, first, we must look at how strong the pressure really is, second, at how difficult it is to overcome the obstacles.

Like all positive feedback loops, the arms race for complexity can produce a strong push for change indeed. In normal feedback loops, this pressure diminishes as the process runs out of resources, that is to say, as it starts to cut into its own supply. When trees grow too tall, they will be the first to be felled by a storm or lightning strike, thus making space for smaller trees. Increasing the variety of actions does not have such obvious negative side-effects. Apart from the obstacles we noted, higher variety always seems to imply higher fitness. The fitness increase gained by adding another action to an organism’s repertoire is not just relative, limited to the specific opponent that triggered the new action. If you learn how to escape one predator, for example by digging a burrow, this action may well help you to escape other dangers as well. The more actions you can take, the more problems you can cope with, independently of what the specific
problem is. In the worst case, learning a new action may only be useful in very specific circumstances, but this will hardly harm your fitness. As the order from noise principle taught us, greater variation in general allows you to find better solutions, whatever your problems are.

In conclusion, the outside pressure to build up complexity is in practice unlimited. The next question is how an organism can overcome its internal obstacles.

**Division of Labor**

As we saw, the main difficulties in increasing functional complexity are the facts that different functions tend to interfere, and that more functions require more “organs” and therefore a larger organism. There is a simple way to overcome these problems, though. Instead of one organism growing in size and complexity, you could start with an assembly of different organisms, each of which is relatively small and simple. If these organisms would cooperate, they could each specialize in a particular function. Together, they would be able to perform a large variety of functions. Because each organism has developed autonomously, these functions would initially be independent. Therefore, even after they start working together, there would be little interference. For example, assume that performing a certain action, say, digesting a particular food, requires the use of certain enzymes. The organism responsible for that function would internally produce these enzymes, let them operate on the food, and excrete the digested food for the others to consume. The enzymes it uses would not come into contact with the others organisms, and therefore could not interfere with their enzymes.

The driving principle here is the *division of labor*. Each member of the assembly is responsible for its own actions. It is “functionally autonomous”: its task is to perform a particular function for the collective, but it is free in deciding how it performs that function. Thus, complex problems are split up into simpler subproblems that are tackled by the individual organisms. This is a general method for tackling any complex problem. It is equivalent to what we called “factorization” when discussing the stepping-stone principle. Instead of searching through all one million combinations when trying to open the safe, the safecracker “factorizes” the problem into independent subproblems: finding the right number for the first dial, for the second dial, and so on. Each subproblem is relatively easy to solve. Solving them all together solves the overall problem. The difference between the safecracker and the cooperating organisms is that the safecracker solves the subproblems one by one, while the organisms solve their subproblem in parallel.
Theoretically, cooperation between specialized organisms is a powerful method to cope with complex environments. Practically, the question is: how do you get them to cooperate? The answer is everything but obvious. The difficulty is that natural selection promotes selfishness. An organism is selected for taking care of its own survival, not for taking care of others. Helping others can actually harm your own survival. First, helping another system requires effort and resources. These resources you can no longer use for your own survival. Second, even if it would cost you nothing to help, that help will make the other more capable to survive and grow, and thus to consume the resources that you need. Most relations between organisms are based on some kind of exploitation of the one by the other: as a predator, a parasite, or simply a competitor.

Helping another one will only increase your fitness if you get more in return than you put in. This strongly restricts the possibility for cooperation to evolve. One way to fulfill this requirement is if help would cost you nothing at all. This will be the case if you and the other organism need different resources. You can easily afford to give away resources that are useless to you. Even if your gift makes your partner thrive, you don’t need to worry that this will make him a more fierce competitor: it will at most sharpen his appetite for resources that you don’t care about. If your partner moreover returns the favor by giving you resources that he doesn’t care about, you will have every reason to continue the cooperation. This situation is at the base of most cases of symbiosis in nature. For example, the symbiosis between plants and animals that we discussed earlier is of this kind: the animals don’t care about the carbon dioxide they exhale, even though it is an essential resource for the plants, while the plants care very little for the oxygen they produce, even though the animals need it.

Such forms of “mutualist” symbiosis will evolve easily, since they directly increase the fitness of the partners. However, it is much more difficult to evolve a complex division of labor in this way. Mutualist organisms fit together like a sodium atom and a chlorine atom: they are complementary, and their interactions settle into a stable pattern of exchange. Things become more tricky if a third organism is to join the partnership. The third guy might settle into a stable relationship with the second guy, but it is unlikely that his use of resources would be complementary with the first guy as well. It is more likely that since both the first and the third are complementary to the middle guy, they would be competing for the same resources. Things get even more complicated if a fourth party were to join the fray. The more parties involved, the larger the probability that at least two parties would use some of the same resources, and therefore be competitors. In conclusion, this kind of symbiosis through complementarity is unlikely to lead to very complex arrangements.
In the evolution of life on Earth, we know one example where this mechanism produced a more complex type of organism: the transition from prokaryotes to eukaryotes. Prokaryotes, which include all types of bacteria, are the simplest types of living cells. They have no internal structure: they just consist of a cell wall, filled with protoplasm in which some DNA is floating. Eukaryotes are the more complex cells that we find in our own body, in the tissues of plants and animals, and in unicellular organisms such as amoebas. They are much larger than prokaryotes, and contain a kernel that houses the DNA, plus a few separate subsystems, called organelles, that fulfill specialized functions, such as producing energy. Some of these organelles have their own DNA to direct their activity. It is now generally accepted that certain organelles are the faraway descendants of independent prokaryotic organisms that somehow started to live inside a larger cell.

There are different theories explaining how they got there. Perhaps the simplest explanation is that they were swallowed by the larger cell, the way amoebas swallow bacteria to eat them. However, instead of being digested, the prokaryotic cells adapted to the inner environment. This is not such a fancy idea if you think about the billions of bacteria that live in your intestines helping you to digest food, for example by breaking down cellulose. Another explanation is that the organelles were initially parasites, like the bacteria that make us ill. A parasite that completely depends on its host to survive and reproduce will benefit from anything that makes the host more fit. Therefore, it is to the advantage of the parasite that it would keep its host fit, by producing a minimum of damage and if possible by helping the host with its problems. Thus, such parasites tend to evolve to symbiotes. (The situation is different for parasites that don’t need a fit host to reproduce, such as malaria parasites that are spread by mosquitoes biting sick patients).

In spite of the example of eukaryotes, which are still quite simple, we clearly need more than symbiosis to evolve a complex division of labor. If complementary organisms cannot do the trick, then we must start with similar organisms. The problem with similar organisms is that they use the same resources, and therefore tend to be competitors. There is a way to overcome that rivalry, though. Imagine a group of cells that are all descendants of a single ancestor. If no mutations have occurred in the meantime, this means that they all carry the same DNA. Thus, they share their genetic control mechanism. These cells will be selected depending on how well that control contributes to their fitness. If cooperating among each other makes them more fit, then their genes will be selected for promoting such cooperation. It is not obvious how essentially identical, single-celled organisms can fruitfully cooperate, though. They still use the same
resources, and therefore they would get into each other’s way if they would stay too close to each other, eating each other’s food supply.

Yet, biologists have discovered how bacteria can cooperate through the broadcasting of chemical signals. Such a signal consists of a particular type of molecule. As the molecule is released in the environment, it diffuses farther and farther away from the place it was released. As we saw when discussing the order from noise principle, bacteria are quite skillful in locating the source of such a diffusing chemical. One example of cooperation occurs if a bacterium discovers a food source that is much larger than it could consume on its own. In that case, it will emit a signal which essentially tells its family members to come and join the feast. Another example is when a bacterium is dying because of starvation or poisoning. In that case it will broadcast a signal that tells the others to stay away from that place. A DNA molecule programmed to produce such signals is obviously more likely to be selected: the one DNA copy in the dying bacterium will get lost with or without the signalling action, but because of the warning, the copies in its siblings are more likely to survive. Similarly, the DNA of the feasting bacterium will multiply with or without signalling, but thanks to the broadcasted invitation, its copies are more likely to multiply as well.

The mechanism underlying the evolution of such cooperation is called inclusive fitness. The fitness that determines whether a particular DNA will become more or less frequent in the population is not the fitness of an individual organism, but of the organism and its family members included. In bacteria, inclusive fitness can extend to the millions of individuals in a bacterial colony that all descend from a common ancestor.

In higher plants and animals, however, inclusive fitness is limited to close family: parents, children, and siblings. The reason is that during sexual reproduction, the DNA of mother and father are recombined to produce DNA that is different from either. A child only shares 50% of its genes with any of its parents or siblings, and 25% with its aunts and uncles. The more indirect the relation, the less DNA is shared, and therefore the less the fitness of one DNA copy depends on the probability of survival of another copy. This means that cooperation is much more difficult to evolve for groups of higher animals. There are exceptions to this rule, though, such as the social insects: bees, ants and termites. Because of the special way they reproduce (essentially, all insects in a colony are children of a single queen), the different ants in a colony share most of their genes. Therefore, it is in the inclusive fitness interests of an ant to help any member of its colony.

Such cooperation between family members still does not imply a division of labor. The problem is that cooperation means that everybody must carry the same DNA,
while division of labor requires different individuals to be specialized for different tasks. The solution to the problem may become clear if you remember that DNA is a control mechanism: it reacts to different situations with different actions. A specific situation, say, the presence of a particular type of food or poison, will activate that part of the DNA that can produce the right actions to handle the situation. Irrelevant segments of DNA simply remain inactive. If two identical organisms were confronted with different situations, different parts of their shared DNA would be active, and therefore they would behave differently. Thus, you could achieve specialization by subjecting different members of the family to different conditions. This still does not solve the problem, though, since family members must stick together to cooperate, that is, they must remain in the same environment. But, as we shall see, even if primitive organisms literally stick together, this will create at least one form of differentiation.

**Developing an Anatomy**

Imagine a ball of cells that constantly grows larger because of cells splitting further and further. Though the ball may seem homogeneous, there is an essential difference between the cells on the inside and those on the outside. The ones on the outside are directly confronted with the dangers and opportunities of the environment. The cells on the inside are to some degree shielded: their life may be more secure, but they depend on their outside cousins to let essential nutrients seep through. In such circumstances, the natural division of labor would be for the outside cells to ingest food and ward off dangers, and for the inside ones to digest the food and distribute the resulting nutrients through the cell mass. This fundamental differentiation is at the base of any multicellular organism: its most primitive difference is the one between the cells of the skin and those of the internal organs.

A next type of cells may appear at a second stage. Imagine the ball of cells, instead of floating around in the water, coming into contact with a surface, such as a rock or the bottom of the ocean. This creates an asymmetry between the skin cells in contact with water and those in contact with the surface. The ball will typically flatten, like a pancake. The lower side of the pancake is protected by the rock, and perhaps feeding on the bacteria that live there. The upper side of the pancake is most exposed to dangers. This creates three types of cells: seaside, rockside and inside. Imagine now the pancake curling up until it forms a kind of bag with an opening towards the rock. This produces a hollow cavity, in which food can be gathered, protected from danger. The cells on the
inside of the cavity may now specialize in ingesting and digesting the food, while those on the outside specialize in warding off danger.

In a next stage, the cavity may develop a second opening. One opening can now be used to ingest food, the other to expel waste. The cavity itself may become longer and narrower, to allow more time for food that passes through it to be digested. This is the shape of a primitive worm-like creature. Its primitive “gut”, the canal through which food is digested, can now further differentiate: the cells in the beginning of the canal (the “mouth”) will swallow as yet undigested food, while those further on (the “intestines”) will absorb nutrients from the food, while those at the end (the “anus”) will get rid of the remaining waste. If the worm learns to move, it would rather go with its mouth first, so that it can swallow the food immediately after it finds it. Therefore, the cells near the mouth will become more sensitive in recognizing food. That is why our sensory organs (eyes, ears, nose, tongue) are concentrated near our mouth. The sensory organs must be able to warn the other cells, so that they can react appropriately to the presence of food or danger. Some of the internal cells in between the skin and the gut may specialize in transmitting such signals from the senses to the rest of the body. This is the beginning of a nervous system.

In conclusion, starting from a homogenous mass of identical cells, cells will evolve to perform increasingly diverse activities, depending on which part of the environment they come in contact with. This leads to on-going differentiation, and the creation of specialized organs and cell types. In complex multicellular organisms, such as we are, differentiation has gone so far that the different cell types seem to have nothing in common anymore. Just compare the amoeba-like white blood cells that roam through your blood vessels to seek out invaders, with the long, branching neurons in your brain, and the hard, calciferous cells that form your bones. They look as though they belong to wholly different organisms. In practice, the difference is permanent. Putting a bone cell inside your brain will not turn it into a neuron, and letting a neuron loose in your blood stream will not turn it into a white blood cell. The differentiation between cells no longer depends on the situation they are in. Yet, all these cells have developed from the initially identical cells of an embryo, and they all contain the same DNA.

The mechanism through which embryonic cells specialize into their different roles is as yet only partially understood. It is probably similar to a speeded-up, self-reinforced version of the differentiation process out of which complex organisms evolved. This may explain why embryos, during their development in the womb, subsequently resemble the different stages of the evolution of higher organisms: from a ball of cells to bag-shaped, worm-like, fish-like (including the gills!) to reptile-like and finally human-like. This
observation has become famous as the principle of \textit{ontogeny recapitulates phylogeny}: the development of an individual to some degree recapitulates the evolution of the species.

During this process, most parts of the shared DNA are desactivated, so that only the parts necessary to fulfill the specific function (swallowing invaders, connecting nerves, or producing bone) remain in control of the cell’s activity. Yet, the other parts have not been lost. Dolly the sheep—the first animal to be cloned from an adult cell—has shown how the “sleeping” DNA can be reactivated. To develop a complete animal out of a single cell, the scientists who created Dolly had to “deprogram” the cell, so that it would forget its specialized function. The cell had to stop being a specialized brain, blood, or bone cell. Only then could it produce an undifferentiated embryo, whose developing cells would gradually acquire their different specializations. The technique they discovered is similar to the one used to brainwash people, so that they forget their past convictions: the cell must be put under extreme stress, by cutting it off from contact with other cells, food, and other resources. Apparently, this interrupts the processes that maintain the cell’s basic organization. By then inserting the cell’s DNA into an empty egg cell, which was implanted into the womb of an adult sheep, the DNA recovered its capacity to produce a full, undifferentiated embryo. Thus was born Dolly, a sheep whose DNA was identical to the one of the adult from whose cells it was cloned.

The difference between organs and organelles illustrates the two different routes followed by evolution towards the division of labor: organs, although very different, all contain the same DNA; the DNA of organelles is different from the cell in which they function. The route where initially identical cells differentiate because they are exposed to different parts of the environment is obviously the most powerful one for generating complexity. The reason is that the mechanism can go on indefinitely: as the number of cells increases, so does the number of possible cell types. Moreover, differentiation is self-reinforcing.

Consider two differentiated cell types, A and B. The place where A tissue comes into contact with B tissue is kind of special, since the circumstances there are neither the ones that led to A-type cells nor the ones that led to B-type cells, but a little of both. The cells in this intermediate layer are therefore in a unique situation. They may adapt to that situation by developing their own mode of operation, which is neither A nor B, but something in between: say, AB. For example, the AB cells may specialize in providing an \textit{interface}, passing on the products from A to B after some additional processing. But the new AB layer again creates a unique intermediate situation, in the no man’s land between A and AB. This may lead to yet another specialized cell type A(AB), and so on. Similarly, the cells on the outer layers, that come directly into contact with the
environment, will specialize in reacting to specific aspects of the environment, such as touch, smell, light, or heat.

Building complex organisms through differentiation of initially similar components also works on a different level: segments. All higher animals are segmented: their body consists of a number of distinct sections, such as bones, limbs, or body parts. This is most clearly visible in animals such as millipedes, insects and crustaceans, but it can also be recognized in the skeleton of vertebrates: fish, amphibians, reptiles, birds and mammals. The simplest examples are segmented worms, such as the earthworms you find in the garden. If you cut such a worm in pieces, the different segments can sometimes survive and grow back into a full worm. In a worm, there is little differentiation between segments: there is a head segment and a tail segment, but the segments in between are very similar. The same applies to centipedes and millipedes: each segment has its own pair of legs, but apart from the head and the tail, they are essentially identical. Within each segment, though, there is a full differentiation of cells, with skin cells, muscle cells, gut cells, and so on. Each segment is like a miniature organism, a morsel of worm.

The origin of such segmented creatures seems clear. If an unsegmented organism would try to reproduce by splitting in two, the two new parts might not completely disconnect. If later, each part would similarly try to split, this might add more segments to the chain. Since they derive from the same ancestor, all segments are essentially identical, carrying the same DNA. Therefore, selection will promote cooperation between the segments. If the segments work together in gathering and digesting food, this means that a single segment can count on the others to support it, even if it would for whatever reason be unable to immediately return the favor. This allows individual segments to undergo variation, focusing their energy on activities different from the basic ingesting and digesting. The head segment, for example, could adapt to the peculiar condition of sniffing out and swallowing food, and leave the digestion to the others. The basic idea is that because the others provide a backup, any individual component can afford to neglect the more general duties and concentrate on one, specialized function. Without cooperation between the segments (or cells, for that matter), specialization would be very difficult to achieve.

The differentiation between segments determines the body plan of all higher animals. You can see this most clearly in animals like shrimps and lobsters. Lobsters consist basically of a chain of segments, each with its own pair of swimming legs. The last segments have fused into a tail, and the first ones have fused into a head. Within the head part, the original leg pairs have differentiated into antennas, pincers, and different types of larger and smaller claws, for walking, grasping, cutting and bringing food to the
mouth. The same body plan can be found in insects, except that the multiple swimming legs have made way for three pairs of walking legs, antennas and wings. In vertebrates, you cannot see the segments from the outside, but you can easily recognize them in the vertebrae that make up the backbone. The top vertebrae have fused to form a skull, while two pairs of limbs have remained attached to the backbone. Like in insects, the limbs themselves are segmented into a number of smaller components.

If you look at the DNA of segmented organisms, you will find that it contains a sequence of genes, the so-called “homeobox” genes, that are rather similar. Each stretch of gene is responsible for one segment. New segments can be produced by adding another gene stretch. Through mutations, these stretches can then start to vary relative to each other, so that their corresponding segments also start to differentiate. The homeobox genes control the overall anatomy of the organism. Scientists have shown how radically different body plans can be created by shuffling around these genes. For example, they have manipulated the homeobox genes of fruit flies so as to create monsters that have an extra pair of legs, another tail instead of a head, or legs sticking out of their head where the antennas are supposed to be. The homeobox genes perform their work by switching other genes on or off. This tells the cells in each segment which tissue type they should produce in which place. Because the homeobox genes are different for the different segments, these segments will develop differently. Together, the homeobox genes determine the overall layout of the body: the number and position of bones, limbs, organs, etc.

**Jumping out of the System**

We discussed two routes for creating a complex organism out of a combination of simpler organisms. The first route of symbiosis, which led to the eukaryotes, is similar to the integration of subsystems into supersystems. As we discussed earlier, to produce a supersystem it suffices to try out different combinations of subsystems (for example atoms) until you find one that is stable or closed (for example a molecule). Once the subsystems fit together, the supersystem comes into being and the process stops. If we follow the second route, things only get interesting after the subsystems have been integrated into a cooperative assembly: only then can they start to differentiate. In the first route, you start with different systems and integrate them. In the second route, you start with similar systems, integrate them, and let them differentiate afterwards. Both processes are essential steps to a higher level of complexity. The first one, which creates
a supersystem, may be called a *supersystem transition*. For reasons that will soon become clear, we will call the second one a *metasystem transition*.

Both types of transitions are key events in the evolution of complexity. However, a metasystem transition leads potentially to a much more complex organization. One reason, as we already noted, is that in a metasystem differentiation is unlimited: you can always increase the number of components in the system, and the more components, the more possibilities for specialization. In a supersystem, on the other hand, components cannot be added indefinitely: they must fit into the existing system. Nothing can be added to a closed system. In a symbiotic system, where closure is not so strictly defined, the requirement for fit is mutual benefit for the participating organisms, but that becomes more difficult to achieve as more potential competitors participate.

There is an even more important reason why metasystems are intrinsically more flexible than supersystems. In a supersystem, the differentiation between components is essential static: the components were different before they got integrated into a supersystem and these differences will remain (or evolve very slowly over many generations, as in eukaryotes). In a metasystem, the differences are dynamic. You see this most clearly in embryo development, where initially identical cells gradually differentiate into all their specialized functions.

But even such specialization is dynamic. For example, all muscle cells are similar. You can easily transplant muscle tissue from your leg to your arm, and it will continue to function as it should. However, arm muscles and leg muscles do not perform the same function at the same time. Your leg muscles do not contract when you lift a glass of water, and your arms remain in rest while you pedal on a bicycle. Even though cells may be anatomically the same, at any given moment they are likely to perform different activities. This is obvious when different cells are affected differently by the environment. For example the light-sensitive cells in your eye that receive direct sunlight will produce stronger impulses than those that only receive indirect light. But in the case of the muscle cells, the influence that makes them act does not come directly from the environment. It is the organism itself that controls their activity.

Now, what is the mechanism that makes cells and organs act differently in different places and at different times? Remember the control loop according to Powers (Fig. ): a control system, such as a cell, acts so as to bring its perception as close as possible to its goal. Thus, its activity depends on two independent inputs: the environmental situation that it senses, and the goal that it tries to achieve. Until now, we have always assumed that the situation may change, but that the goal is fixed. However, goals can change when the situation changes. But to change a goal in a meaningful way,
you need another, higher-order goal. For example, if you want to shoot at a target, the position of the target determines your goal, and your actions will be aimed to point and keep the gun in the right position, in spite of perturbing factors such as trembling, wind or the weight of the gun. But if the target moves, your higher-order goal of reaching the target will make you change aim in step with the movements of the target. The Powers control loop can easily be extended to incorporate such a higher-order control: it suffices that the goal of the lower-order control loop be itself determined by the action of a higher-order control loop, as in fig.

Fig.: a second-order control loop: control loop 1 is determined by goal 1, which is itself the outcome of the action 2 of control 2, determined by goal 2.

Such multilevel control loops, where higher-order goals set or change the goals at the lower level, are easily found in the body. For an individual cell, the ultimate goal is to survive and reproduce. For a cell in a multicellular organism, this is no longer true. Most cells in the body do no reproduce. Many will even commit cellular suicide ("apoptosis") when they are damaged or no longer needed for the organism. The goal of a cell is
determined at a higher level: the organism as a whole. It is the organism that must survive and reproduce: its cells are merely tools to achieve that purpose. Because of the division of labour, different cell types contribute in different ways to that overall goal; they have a specialized function. Each function is implemented as a set of goals specific for that cell. For example, the goal of neurons is to transmit signals, the goal of white blood cells is to attack invaders. These goals are implemented by differential activation of the genes in each cell type. The genes for transmitting signals are switched off in white blood cells, while the genes for swallowing bacteria are switched off in neurons. Most of this differential activation has been set during embryological development, and will remain for as long as the cell survives. Some of this “setting”, as we have seen, has been done by the homeobox genes. The homeobox genes function as “higher-order” controllers, that tell other genes whether or not they should be active in controlling the cell’s activities.

However, the genes that actively control a cell will still respond differently to different situations. Thus, the actions of a cell can still vary within the limits of the specialized cell type. Some of these actions will be triggered by the external situation. For example, a very hot environment may trigger the production of “heat shock” proteins, that protect the cell against extreme temperatures. Most actions, though, will be triggered by the actions of other cells. Such communication between cells takes place through the release of chemical signals. Such signals are called “hormones”. For example, if some of the cells in the body sense danger, they will release stress hormones. When other cells sense these hormones, they will reduce most of their normal activities, and prepare for a state of high energy production, so that the organism is able to react powerfully to whatever danger threatens it.

Hormones are quite flexible signals to control the overall activity of the body depending on the circumstances. However, they have a number of shortcomings. First, hormones must diffuse and circulate with the blood to reach all the cells in the body. Such diffusion is not very fast or efficient. Moreover, this means that all cells get the same signals, which is not very handy if you want to precisely target your message. This problem is partly solved because cells will disregard a signal that is not intended for them. For example, the sex hormone testosterone may affect the erectile tissue in the penis, but will be ignored by cells in the feet. Different cell types will have different “receptors”: molecular sensors that only recognize specific chemical messages. But this only allows you to target a message at all cells of a given type. To communicate with smaller groups, or even individual cells, you need a more precise mechanism. This mechanism is the nervous system.
Nerve cells are specialized in the transmission of messages. They receive a message in the form of an electrical impulse from a neighboring nerve cell, and pass this message on to one or more subsequent nerve cells. Electrical impulses travel much faster than blood. Moreover, nerve cells only transmit their message to precisely targeted other cells. This allows very quick and specific reactions. An example can be found in the hydra, a very simple animal that lives in lakes and rivers, and that looks like a small polyp or sea-anemone. When you touch a hydra, it will immediately retract its tentacles and curl up into a ball. Sensor cells on the skin of the hydra will convert the touch into an electrical impulse, that travels via a nerve to the muscle cells inside its body, and makes them contract. Thus, the sensation of a disturbance directly leads to the appropriate counteraction. Therefore, the hydra is a simple and efficient control system.

Such control via the nervous system functions at a much shorter time scale than the control via hormones or homeobox genes. Therefore, it allows the organism to act much more quickly and flexibly, adapting in mere milliseconds to what happens in the environment. Yet, hydra-like control is still primitive. The connection between sensation and action is purely automatic: the hydra’s reaction is a reflex movement. It will always react in the same way to the same stimulus. Of course, we could imagine a more complicated nervous system, where different sensors using different nerves would activate different muscles. For example, a touching sensation on the hydra’s tentacles rather than on its body might make the hydra close its tentacles in order to catch what appears like a prey animal. Different sensations will trigger different actions. But this is still merely a single control level. Every reflex arc has an implicit goal: for example, protecting the body, or catching prey. But because each nerve only gets input from its own sensors, these goals are independent: they cannot be changed or reset.

Now imagine a more complex nervous system, where different nerves join and connect, forming a network. Different sensors will all feed information into this same network. Different muscles will all receive their orders to act from this network. Now, different sensations will combine together, determining an overall body reaction, rather than merely a set of disconnected muscle contractions. For example, if a simple, reflex-type hydra were touched both on its body and on its tentacles, it would both contract its body and close its tentacles around an apparent prey. These two movements are hard to perform together, and would merely result in an uncoordinated wriggle. If the hydra’s nerves would be interconnected, the signal from the body and the signal from the tentacles could interact, resulting in a single, coordinated signal to the muscles. For example, the network might be wired so that the signal from the body suppresses the one from the tentacles, resulting in a body contraction without interference from the tentacles.
In this case, the goal of protecting the body overrides the one of catching prey. Thus, one goal can change or reset another goal.

With many different signals coming together in the network, many more complicated forms of coordination are possible. As we noted in the section on knowledge and action, we can imagine the underlying mechanism as a system of “if...then” rules, with different strengths. Rules with more strength (e.g. “if body is touched then contract”) will override weaker rules (e.g. “if tentacles are touched, then close in on prey”). But a bunch of weak rules together may garner enough strength to override a strong rule. In the nervous system, the strength of a rule corresponds to the strength of a neural connection. A rule is active if the corresponding nerve or neuron is activated by an electrical impulse. Strong connections will react to weak impulses, and still pass on the signal. Weak connections will only fire if they get a lot of incoming activation. Thus, activation from different sensors will be differentially added together, with some signals contributing more than others. This activation will again be distributed over the different muscles, resulting in an overall, coordinated action.

Such a network is capable of more than just distributing signals from the sensors to the various muscles. The activation in the network does not depend only on the present sensation. Activation from previous sensations may continue to circulate in the network for a while, combining with newly incoming activation, and thus affecting the outcome. For example, if a hydra is touched repeatedly, after a while it will stop contracting. This is called “habituation”. You might say that the hydra has gotten used to being touched, and therefore is no longer frightened. You might also see it as the hydra getting tired of contracting again and again. Whatever your interpretation, habituation means that the reaction to a repetitive stimulus tends to die down. This is a useful adaptation, which allows the hydra to save its energy in situations where no new dangers are coming up.

How can habituation be achieved? The simplest mechanism is that the activation from previous stimulation, which continues to circulate in the nervous system, would inhibit or suppress newly incoming activation. As more activation from previous experiences accumulates, the suppressing effect will become stronger, until the hydra stops reacting. This may not be the precise mechanism used by hydras and other simple creatures. The point is that with networked neurons, what counts is not just the present sensation, but the overall activation of the network. This state of activation represents the organism’s awareness of its situation, both present and recent past. Thus, what might have seemed to be an appropriate goal for action a while ago (contracting in order to avoid danger), may no longer be a goal for the organism presently. In other words, goals can change depending on circumstances, past and present.
The direct sensor-muscle connection and the connection via a complex network represent two levels of cybernetic development. Valentin Turchin, who introduced the concept of metasystem transition, calls the first level the *simple reflex*, the second level the *complex reflex*. In the first level, the animal has a reflex-like control over its actions: it knows which muscle to contract when a certain sensor is stimulated. In the second level, it has control over its simple reflexes: it knows which of the different possible movements to perform at any given moment, depending on the different present and previous sensations. Thus, an animal with complex reflexes is much more flexible or adaptive in its responses. It can change its short-term goals depending on the whole of its perception. However, its reactions are still largely automatic.

At the next level, the animal will be able to learn new reactions. Assume that the neural network is no longer hard-wired. Connections between neurons can become stronger or weaker, depending on how often they were used. Connections can even appear or disappear completely. Habituation is merely a temporary phenomenon. If a hydra “tired” of contracting is left alone for a while, and then touched, it will again contract just as forcefully as the first time it was touched. The state of activation, which kept it from reacting, has dissipated. The neural connections that lead from sensors for touch to muscles for contraction, however, are still the same.

[manuscript finishes here]