INTRODUCTION

There is a strange paradox in modern biology. On the one hand, new discoveries are made at such a high rate that our science of life appears full of surprises and in a constant state of flux. On the other hand, all new findings are apparently accommodated within a theoretical framework that remains remarkably stable. Present-day biology, in other words, seems to be in that phase of development that Thomas Kuhn referred to as *"normal science"*, a phase in which an endless stream of novelties is smoothly accounted for by an unchanging paradigm. And this is definitely not for want of alternatives. No efforts have been spared to provide different explanations of life, but none has withstood the test of time. What makes us feel good about our present paradigm (which many call *universal Darwinism*) is that only the truth – or something very near the truth – can resist so many assaults and outlive generations of critics. In such a situation, I find it almost embarrassing to suggest that our beloved paradigm is not as perfect as we like to think. But this is the message that is coming from nature, and I had better tell you straight away the reasons that lead to this conclusion. The main points are three: the existence of organic codes, a mathematical model of epigenesis and a new theory of the cell.

The organic codes

From time immemorial it has been thought that codes, or conventions, exist only in the mutable world of culture, while nature is governed by immutable laws. The discovery that a genetic code is at the very heart of life came therefore as a bolt from the blue. And people rushed

to anaesthetise it. The genetic code was immediately declared a *frozen accident*, and the divide between nature and culture remained substantially intact. The existence of other organic codes is, in principle, as natural as that of the genetic code, but its implications are perhaps even more revolutionary. The genetic code appeared on Earth with the first cells, while the linguistic codes arrived almost 4 billion years later, with cultural evolution. These are the only codes that modern biology currently recognises, which is tantamount to saying that in 4 billion years no other code appeared on our planet. And if codes are relegated to the beginning and to the end of the history of life, we can safely say that 4 billion years of biological evolution went on with the sole mechanism of natural selection. In this book, however, we will see that there are many other organic codes in nature, and that they appeared not only throughout the history of life but marked the main steps of that history, the steps which brought about the great events of macroevolution. But if codes exist, they must have had origins and histories, and above all they must have had a specific mechanism. Languages evolved not only by chance mutations of letters in their words but also by changes in their grammatical rules, and the same would apply to living organisms. We must conclude, in short, that biological evolution was produced by two distinct mechanisms: *by natural selection and by natural conventions.*

From a logical point of view this is a straightforward conclusion, but unfortunately theory and practice do not always go hand in hand. The idea of evolution by natural conventions was proposed for the first time in 1985, in a book of mine entitled *The Semantic Theory of Evolution*, but it did not have any significant impact (even if I am pleased to say that in a private letter Karl Popper called it *"revolutionary"*). Regrettably, people do not seem to associate the existence of organic codes with a mechanism of natural conventions, as if one could exist without the other. Edward Trifonov, for example, has been campaigning in favour of sequence codes since 1988, and in 1996 William Calvin wrote a book entitled *The Cerebral Code*, but nobody called for anything different from natural selection. And there is a reason for that. The reason is that the word *code* has largely been used in a metaphorical sense, as have so many other words which

have been borrowed by molecular biologists from everyday language. It is imperative, therefore, to realise that there are organic codes which are not metaphorical but real, and to this purpose we clearly need to prove their existence. A code is a correspondence between two independent worlds, and a real organic code requires molecules that perform two independent recognition processes. These are the codes' fingerprints, and it is they that we must look for and bring to light. In the genetic code these molecules are the transfer RNAs, but we will see that equivalent *adaptors* (the word that Francis Crick initially proposed for the tRNAs) exist in at least two other processes (signal transduction and splicing) and are expected to turn up in many other cases. And luckily this is beginning to happen. In the year 2000, for example, Gabius provided evidence for a *sugar code*, while Strahl, Allis, Turner and colleagues discovered a *histone code*. The more we learn about organic codes, in conclusion, the more they turn out to be every bit as real as the genetic code. Sooner or later, therefore, biologists will have to come to terms with the theoretical implications of this extraordinary experimental fact.

A mathematical model of epigenesis

Embryonic development was defined by Aristotle as an *epigenesis*, i.e. as a sequence of one genesis after another, a step-by-step generation of new structures. Today epigenesis is often referred to as an *increase of complexity*, but when we use this expression we should always add an important qualification. We should say that epigenesis is a *convergent* increase of complexity, in the sense that its outcome is neither random nor unexpected. This is what makes it so radically different from the *divergent* increase that takes place in evolution. The distinction between convergent and divergent phenomena is particularly relevant today that the study of complexity has become a research field in its own right. Many interesting ways of obtaining *"order out of chaos"* have been described and have found applications in various disciplines, but the expectation that they could apply to embryonic development has been an illusion. Embryos are not chaotic

systems, and embryonic stages are not phase transitions.

To my knowledge, there is only one mathematical model which has described how a convergent increase of complexity can actually take place. I developed this model as a special case of the general problem of reconstructing structures from projections, a problem which arises in fields as diverse as radioastronomy, electron microscopy and computerised tomography. The mathematics of the reconstruction problem has been thoroughly investigated, and the minimum number of projections required for a complete reconstruction is prescribed by basic theorems. This allows us to give a precise formulation to a problem which may seem hopeless at first sight: the problem of reconstructing structures *from incomplete information.* We can legitimately say that we are performing this type of reconstruction when we work with a number of projections which is at least one order of magnitude less than the theoretical minimum, i.e. when we use 10% or less of the minimal information. What is interesting about this strange-looking problem is that *a reconstruction from incomplete information* is equivalent, to all practical purposes, to *a convergent increase of complexity*, and so it is a mathematical formulation of the problem of epigenesis (if the starting information is incomplete, the reconstruction must produce an increase of information and this is equivalent to an increase of complexity). Even more interesting is that the problem can actually be solved, as we will see in Chapter 3. And the beauty of the solution is that its logic can be grasped even without the mathematics (which will however be provided). The model employs an iterative procedure that performs in parallel two different reconstructions: one for the structure in question and one for its *reconstruction memory*. The key point is that the memory space turns out to have the surprising ability to provide new specific information about the examined structure, and such information can be transferred from the memory space to the structure space with appropriate codes, or conventions. The conclusion is that a convergent increase of complexity can be achieved if a reconstruction is performed with memories and codes. Which means, in biological terms, that epigenesis requires organic memories and organic codes. We come back, in this way, to the issue of the organic codes. And the message from

mathematics is strong: there is no way that we are going to understand a phenomenon as large as embryonic development without organic codes and organic memories.

A new theory of the cell

The extraordinary thing about codes is that they require a new entity. In addition to energy and information they require *meaning*. For centuries, meaning has been regarded as a spiritual or a transcendental entity, but the very existence of the genetic code proves that it is as natural as information. And in fact we can define meaning with an operative procedure just as we do with any other natural entity. *Meaning is an object which is related to another object via a code*. The meaning of the word *apple*, for example, is the mental object of the fruit which is associated to the mental object of that word by the code of the English language (needless to say, the code of another language would associate a different mental object to the same word). The meaning of a combination of dots and dashes is a letter of the alphabet, in the Morse code. The meaning of a combination of three nucleotides is usually an amino acid, in the genetic code (from which it follows that the meaning of a gene is usually a protein).

We are well aware that it is man who gives meaning to mental objects – in the realm of the mind he is the *codemaker* – but this does not mean that a correspondence between two independent worlds must be the result of a conscious activity. The only logical necessity is that the codemaker is *an agent that is ontologically different* from the objects of the two worlds, because if it belonged to one of them the two worlds would no longer be independent. A code, in other words, requires three entities: two independent worlds and a codemaker which belongs to a third world (from a philosophical point of view this is equivalent to the triadic system proposed in semiotics by Charles Peirce). In the case of the genetic code, the codemaker is the ribonucleoprotein system of the cell, a system which operates as a true third party between genes and proteins. This is why I proposed, in 1981, that the cell is not a duality of genotype and phenotype but a

trinity made of genotype, phenotype and *ribotype*. And I argued that the ribotype is a cell category that not only has the same ontological status as genotype and phenotype, but has a logical and a historical priority over them (hence the title of the paper: "The ribotype theory on the origin of life").

The fact that the ribotype is the codemaker of the genetic code leads necessarily to a change of our traditional view of the cell, but there is also another reason for this theoretical shift. The definitions of life that have been proposed in the last 200 years (for a compendium see the Appendix), have underlined a variety of presumed *essential* features (heredity, replication, metabolism, autonomy, homeostasis, autopoiesis, etc.), but none of them has ever mentioned *epigenesis* as a defining characteristic of life. The reason of course is that epigenesis has been associated with embryos, not with cells, and yet even in single cells the phenotype is always more complex than the genotype. This means that every cell has the ability to increase its own complexity, and so it really is an epigenetic system. We realise in this way that the mere presence of organic codes in every cell, starting from the genetic code, requires a theoretical framework where organic meaning is a necessary complement of organic information. And that is precisely what semantic biology is about. It is not a denial of our Darwinian paradigm. It is a genuine extension of it.

About this book

Chapters 1 and 2 are an introduction to the cell theory and to the theories of evolution at a level that may be regarded as undergraduate or thereabouts. Those who are not concerned with undergraduates may skip them and start with Chapter 3, but should not forget that semantic biology applies to all levels of the life sciences and is not just a section for specialists. Chapter 3 presents a model of epigenesis first in words and then in formulae, and even the biologists who are not devotees of mathematics can follow it from beginning to end. This is highly recommended because the idea that a structure can be reconstructed from incomplete information is still met with incredulity

in our present educational system (if engineers and computer scientists insist that it can't be done, just tell them that embryos do it all the time). Chapter 4 is a biological sequel of Chapter 3 and makes a first excursion into the world of organic codes and organic memories. This is instrumental to the next three chapters which are dedicated to the main events of macroevolution: the origin of life (Chapter 5), the emergence of eukaryotic cells (Chapter 6) and the Cambrian explosion of animals (Chapter 7). These chapters allow us to revisit those great transitions and show how different they look like when organic codes are taken into account. Chapter 8 brings together the ideas of the previous chapters and presents a first outline of the framework of semantic biology. And in order to underline the logical structure of this framework, Chapter 9 makes a brief summary of it in eight propositions (four principles and four models).

The chapters of this book are arranged in a sequential order, but they are also largely autonomous and one can read them in any order. Everything in biology is linked to everything else, and it doesn't really matter where one starts from. What does matter is that whichever way we look at life today we realise that organic codes are there, that they have always been there, from the very beginning, and that it is about time we start taking notice.

1

THE MICROSCOPE AND THE CELL

The cell theory and the theory of evolution are the two pillars of modern biology, but only the latter seems to be the object of ongoing research and debates. The cell theory is generally regarded as a closed chapter, a glorious but settled issue in the history of science. The emphasis today is on cell experiments, not on cell theory, and there is no doubt that one of our greatest challenges is the experimental unravelling of the extraordinary complexity that has turned out to exist in the cellular world. At various stages of this book, however, we will see that the experimental results suggest new ideas, and at the end of the book it will be possible to combine them in a new model of the cell. This is because cells are not only what we see in a biological specimen through the lenses of a microscope, but also what we see through the lenses of a theory. The cell, after all, is a system, and understanding the logic of a system requires some theorising about it. And since this theorising has a long history behind it, let us begin by retracing the main steps of that intellectual journey. This chapter shows that the concept of the cell had to be imposed on us by the microscope because it was unthinkable in the world-view of classic philosophy. And after that intrusion, the concept has gradually changed and in so doing it has changed our entire approach to the problems of generation and embryonic development. But this historical journey is not without surprises, because it will take us toward an idea that all definitions of life of the last 200 years have consistently missed. The idea that epigenesis does not exist only in embryos but in every single cell. That the phenotype is always more complex that the genotype. That epigenesis is a defining characteristic of life.

The cell theory

The idea that all living creatures are made of cells has changed more than anything else our concept of life, and is still the foundation of modern biology. This great generalization was made possible by the invention of the microscope, but did not come suddenly. It has been the culmination of a collective research which lasted more than two hundred years, and in order to understand it we must be aware of the main problems that had to be solved.

Let us start with the microscope. Why do we need it? Why can't we see the cells with the naked eye? The answer is that the eye's retina itself is made of cells. Two objects can be seen apart only if their light rays fall on different cells of the retina, because if they strike the same cell the brain receives only one signal. More precisely, the brain can tell two objects apart only when their images on the retina have a distance between them of at least 150 μm (thousandths of a millimetre). The cells have average dimensions (10 μm) far smaller than that limit, and, even if an organism is stared at from a very close distance, their images overlap and they remain invisible. It is therefore necessary to enlarge those images *in order to increase their distance on the retina*, and that is where the microscope comes in.

Enlargements of 5 or 10 times can be obtained with a single lens (the so-called simple microscope) but are not enough for seeing the cells. Substantially greater enlargements require a two-lens system (a compound microscope) and the turning-point came in fact with the invention of that instrument. The first two-lens optical systems were the telescopes, and the idea of a compound microscope came essentially from them. In 1610 Galileo made one of the first compound microscopes with the two lenses of a telescope, and in 1611 Kepler worked out the first rules of the new instrument.

The invention of the microscope brought about an immense revolution in science. It led to the discovery of an entirely new world of living creatures that are invisible to the naked eye, the so-called *micro-organisms.* The microscopists of the seventeenth century were the first men who saw bacteria, protozoa, blood cells, spermatozoa and a thousand other *animalcula*, and gradually realised that the large

creatures of the visible world are actually a minority in nature. The micro-organisms make up the true major continent of life, and their discovery changed our perception of nature to the very core.

Unfortunately, the microscopes of the seventeenth and eighteenth centuries had a basic structural defect. Lenses that are made of a single piece of glass cannot focus in one point all the light rays that cross them, and their images are inevitably affected by aberrations. The rays that traverse the periphery of the lens, for example, do not converge with those that cross the central part, thus producing a *spherical aberration.* Likewise, the rays which have different colours (or frequencies) converge at different distances from the lens giving origin to *chromatic aberrations.* Because of these distortions, people could see only isolated cells, such as bacteria and protozoa, or plant cells, which are separated by thick cellulose walls, but could not see cells in animal tissues. It is true therefore that in those centuries people saw many types of cells, but the microscope was showing that the smallest units of plants (the compartments that in 1665 Robert Hooke called *"cells"*) are not seen in animals, and it was impossible therefore to think of a common structure.

The discovery that cells exist in all organisms required a new type of microscope, and this came only in the nineteenth century, when the aberration obstacle was overcome by the introduction of *achromatic lenses.* These are made of two or more pieces whose geometrical forms and refraction indices are such that the aberrations of one piece are precisely compensated by those of the other. The first achromatic microscope was build by Giovanni Battista Amici in 1810, and with this new instrument came a systematic revision of all that the microscope had revealed in previous centuries. In 1831 Robert Brown discovered that plant cells contain a roundish refracting mass that he called the *nucleus*, and inside the nucleus it was often possible to see an even more refracting structure that later became known as the *nucleolus.* In 1839 Matthias Schleiden and Theodor Schwann compared plant embryos (which do not have the thick cellulose walls of adult tissues) with animal embryos, and discovered that their microscopic structures are strikingly alike. They are both made of nucleated cells, hence the conclusion that the cell is a universal unit