Biosemiotics (2013) 6:33–59
DOI 10.1007/s12304-012-9149-1

ORIGINAL PAPER

The Paradigms of Biology

Marcello Barbieri

Received: 16 May 2011 / Accepted: 26 February 2012 / Published online: 24 April 2012

Abstract Today there are two major theoretical frameworks in biology. One is the 'chemical paradigm', the idea that life is an extremely complex form of chemistry. The other is the 'information paradigm', the view that life is not just 'chemistry' but 'chemistry-plus-information'. This implies the existence of a *fundamental* difference between information and chemistry, a conclusion that is strongly supported by the fact that information and information-based-processes like heredity and natural selection simply do not exist in the world of chemistry. Against this conclusion, the supporters of the chemical paradigm have pointed out that information processes are no different from chemical processes because they are both described by the same physical quantities. They may appear different, but this is only because they take place in extremely complex systems. According to the chemical paradigm, in other words, biological information is but a shortcut term that we use to avoid long descriptions of countless chemical reactions. It is intuitively appealing, but it does not represent a new *ontological* entity. It is merely a derived construct, a linguistic metaphor. The supporters of the information paradigm insist that information is a real and fundamental entity of Nature, but have not been able to *prove* this point. The result is that the chemical view has not been abandoned and the two paradigms are both coexisting today. Here it is shown that an alternative does exist and is a third theoretical framework that is referred to as the 'code paradigm'. The key point is that we need to introduce in biology not only the concept of information but also that of meaning because any code is based on meaning and a genetic code does exist in every cell. The third paradigm is the view that organic information and organic meaning exist in every living system because they are the inevitable results of the processes of copying and coding that produce genes and proteins. Their true nature has eluded us for a long time because they are *nominable* entities, i.e., objective and reproducible observables that can be described only by *naming* their components in their natural order. They have also eluded us because nominable entities exist only in artifacts and

M. Barbieri (✉)
Dipartimento di Morfologia ed Embriologia, Via Fossato di Mortara 64a, 44121 Ferrara, Italy
e-mail: brr@unife.it

biologists have not yet come to terms with the idea that *life is artifact making*. This is the idea that life arose from matter and yet it is fundamentally different from it because inanimate matter is made of spontaneous structures whereas life is made of manufactured objects. It will be shown, furthermore, that the existence of information and meaning in living systems is documented by the standard procedures of science. We do not have to abandon the scientific method in order to introduce meaning in biology. All we need is a science that becomes fully aware of the existence of organic codes in Nature.

Keywords Information · Meaning · Organic codes · Mechanism · Observables · Ontology

Introduction

From time immemorial it has been taken for granted that life is fundamentally different from matter, but in the last few centuries this belief has been seriously challenged by the view that '*life is chemistry*'. The idea that life had a natural origin on the primitive Earth suggests that the first cells came into being from previous chemical systems by spontaneous chemical reactions, and this is equivalent to saying that there is no *fundamental* divide between life and matter.

This 'chemical paradigm' is very popular, today, and is often considered in agreement with the Darwinian paradigm but this is not the case. The reason is that natural selection, the cornerstone of Darwinian evolution, does not exist in inanimate matter. In the 1950s and 60s, furthermore, molecular biology has uncovered two fundamental components of life – biological information and the genetic code – that are totally absent in the inorganic world, which means that chemistry alone is not enough, that '*life is chemistry+information*'. This is the 'information paradigm', the idea that information is unique to living systems and that a deep divide does exist between life and matter.

Ernst Mayr, one of the architects of the Modern Synthesis, has been one of the most outspoken supporters of the view that life is *fundamentally* different from inanimate matter. In *The Growth of Biological Thought* (1982), he made this point in no uncertain terms:

> "… The discovery of the genetic code was a breakthrough of the first order. It showed why organisms are fundamentally different from any kind of nonliving material. There is nothing in the inanimate world that has a genetic program which stores information with a history of three thousand million years!" (p. 124)

> "… Except for the twilight zone of the origin of life, the possession of a genetic program provides for an absolute difference between organisms and inanimate matter. " (p. 56)

The discoveries of molecular biology, in short, appear in contrast with the chemical paradigm, and this raises formidable problems. On the one hand it is an experimental fact that natural selection, biological information and the genetic code do not exist in inanimate matter. On the other hand, we seem unable to accept that life evolved from inanimate matter and yet it is fundamentally different from it. How can something give origin to something fundamentally different from itself? How could the physical world produce life if there is a discontinuity between them?

The aim of this paper is to show that a solution to these problems does exist, but it is not provided by the paradigms that are based respectively on chemistry and information. It is provided instead by a third approach that here is referred to as the 'code paradigm' because it is based on the organic codes of life. To this purpose the paper has been divided into two parts. The first is dedicated to the present paradigms of modern biology and the other to the new theoretical framework.

PART 1 Chemistry Versus Information

The Chemical Paradigm

Ever since the scientific revolution, physics has been the 'queen' science, and biologists have been split into opposite camps, one in favour and one against adopting its method, an approach which has become known as *mechanism*. In biology, the first version of mechanis was the Cartesian doctrine that "*the body is a machine*" and that the clock is its model: *"A healthy man is like a well functioning clock, and an ill man is like a clock that needs repairing"* (Descartes 1637).

The mechanical concept of nature spread very quickly in 17th century Europe, but not without conflict. Opposition came particularly from a new science that was slowly emerging from alchemy and that regarded the human body essentially as a seat of chemical reactions. The heirs of the alchemists were determined to leave magic behind but had no intention of accepting the 'mechanical' view of nature, and one of chemistry's founding fathers, Georg Ernst Stahl (1659–1731), launched an open challenge to mechanism. He claimed that organisms cannot be machines because what is taking place inside them are real transmutations of substances and not movements of wheels, belts and pulleys.

The arguments of the chemists did have an impact, and eventually forced mechanists to change their model. In the course of the 18th century, the view that organisms are *mechanical machines*, gradually turned into the idea that they are *chemical machines.* This change went hand in hand with the development of the steam engine, and that machine became the new model of biology. In the 19th century, furthermore, the study of the steam engine was pushed all the way up to the highest level of theoretical formalism, and culminated with the discovery of the first two laws of thermodynamics. The result is that any living system came to be seen as a *thermodynamic machine*, i.e., as a chemical machine that must be continuously active in order to obey the laws of thermodynamics.

The old opposition between physics and chemistry came to an end, and the two sciences together gave origin to a unified framework that is often referred to as the '*chemical paradigm*', the idea that life is an extremely complex form of chemistry. This is equivalent to saying that all biological processes are chemical transformations of *matter and energy*, and are completely described, in principle, by physical quantities.

The chemical paradigm has underlined time and again – against all forms of vitalism – that living systems are subject to the laws of thermodynamics, but it is by no means limited to this principle. It is a paradigm which has steadily grown by adding new arguments to its thesis. The non-equilibrium thermodynamics of Ilyia Prigogine, the phase-transitions of Stuart Kauffman, chaos theory and complexity theory, are all descriptions of natural processes that rightly belong to the framework of the chemical paradigm.

The same is true for the idea that life is shaped by physical forces and by mathematical principles, a recurrent theme in the history of science, from Goethe and D'Arcy Thomson to René Thom and Brian Goodwin and to the recent research field of Systems Biology. The chemical paradigm, in short, is the view that the laws of physics and chemistry and the principles of mathematics are all that we need to account for the presence of life in the universe.

The Information Paradigm

At the beginning of the 20th century, the rediscovery of the laws of Mendel led Wilhelm Johannsen to make a sharp distinction between the visible part of an organism (the *phenotype*) and the invisible part that carries its hereditary instructions (the *genotype*). Johannsen (1909) proposed that every living being is a dual entity, a synthesis of two complementary realities. This idea was largely ignored, at first, but a few decades later the computer made it immediately comprehensible. The *phenotype-genotype* duality was a *hardware-software* distinction, and became the prototype description of any organism. The model of the living system changed again and became the computer.

In 1953, James Watson and Francis Crick pointed out that the sequence of nucleotides represents the *information* carried by a gene. A few years later, the mechanism of protein synthesis was discovered and it was found that the sequence of nucleotides in genes determines the sequence of amino acids in proteins, with a process that amounts to a transfer of linear information from genes to proteins. This led to the idea that *biological information* is the specific sequence in which the subunits of a molecular polymer are arranged.

These discoveries gave origin to the '*information paradigm*', the second great theoretical framework of modern biology. It is the idea that living systems are *information-processing machines*, and that life is based not only on chemistry (energy and matter) but also, and above all, on *information* (Maynard-Smith 2000). In this framework, chemistry accounts for the hardware of living systems whereas information provides the software, and the view that 'life is chemistry' was replaced by the idea that 'life is chemistry+information'.

This, in turn, led to the concept of the 'genetic programme', the idea that the genome is for the cell what a programme is for a computer. The logical separation that exists between programme and machine implies that something similar exists between the genome and the cell, and such a biological separation has in fact been documented by an outstanding number of experimental results (Danchin 2009). Many genes, for example, have been transplanted from one organism to another and have turned out to be fully functional inside the new cells. Many bacteria now produce human proteins, and the very existence of viruses can be explained by the transmission of independent genetic strings, thus confirming that genes are separable from the cell machine. It has even been possible to transplant an entire genome from one species to another, thus proving that a genome does have a substantial degree of autonomy (Lartigue et al. 2007).

This informational view of life, has been fully accepted into the Modern Synthesis because the concept of information goes hand in hand with the processes of heredity and natural selection. Heredity is precisely the transmission of genetic information from one generation to the next, the short-term result of molecular copying. The long-term repetition of copying, on the other hand, is inevitably accompanied by errors, and in a world of limited resources not all copies can survive and a selection is bound to take place. That is how natural selection came into existence. It is the long-term result of molecular copying, and can exist only in a world of molecules that carry information.

Today, in other words, heredity and natural selection are both squarely based on information, and the information paradigm has become, to all effects, the core of the Modern Synthesis, a view of life which is in conflict with the chemical paradigm, because information, heredity and natural selection simply do not exist in the world of chemistry.

Sequences and Specificity

There is a clear similarity between the sequence of nucleotides in a gene and the sequence of letters in a sentence, and in both cases we say that they carry information: hereditary information in the genes and syntactic information in language. This concept, furthermore, can be generalized to many other types of sequences. A painting or a photograph, for example, can be digitized and represented by a matrix of pixels arranged in rows and columns, but the rows can also be arranged one after the other in a line and give origin to a one-dimensional string of pixels, i.e., to a sequence. The same can be done with sounds and music, first by digitizing them and then by arranging their elements in a linear order. Finally, we can represent letters, numbers and many other symbols with the bytes of computer language, and any sequence can be described as a sequence of bytes. More precisely, we can represent in computer language any configuration of abstract or concrete objects that is (1) linear, (2) digital and (3) finite.

We obtain in this way a first definition: a sequence is any collection of a finite number of digital objects that are arranged in a linear order. The fact that the objects (nucleotides, letters, pixels, musical notes etc.) are arranged on a line in a precise way, and not at random, means that a sequence necessarily has a unique, or *specific* order, and it is exactly this order that biologists call *information*. More precisely it is called *sequence information*, because all sequences have it, and often it is also called *specificity* to underline that its defining feature is the specific order of its components.

A sequence defined in this way has two outstanding characteristics. The first is that specificity, has nothing to do with meaning. The sequence information of the word 'ape', for example, is the same in all languages, but in English it means 'tailless monkey', whereas in Italian it means 'bee' and in French it has no meaning. Specificity, or sequence information, in other words, is a syntactic entity not a semantic one, and for this reason it is often referred to as *syntactic information*.

The second outstanding characteristics is that specificity *cannot be measured*. It can only be identified by *naming* its components in their natural order. This is something that biologists, and in particular geneticists, tend to ignore. The usual reaction I get when I say that genetic sequences cannot be measured is "What are you talking about? We measure them everyday – it's called genotyping".

In order to clarify this point we need to keep in mind that there are two features in a sequence that can be measured. The first is the *size*, or the *length*, of the sequence, a quantity that represents the total number of its characters, or the number of bytes that we need to store it in a computer's memory.

The second is the *relative distance* that exists between two sequences, a quantity that measures the degree of *compatibility*, or *relatedness*, that exists between them. Genotyping, for example, is the technique that measures the number of nucleotides that two genetic samples have in common. The same technique, applied to literature, allows us to measure the relative distance that exists between two books, or two languages, by counting the number of words that they have in common. The relative distance that exists, for example, between *Hamlet* and *Macbeth*, gives us a measure of their relatedness, but has nothing to do with the specific order of their words, i.e., with their specificity, and even less with their meanings.

Any sequence, in conclusion, is characterized by four distinct entities, two of which can be measured whereas the other two are not measurable: (1) the *length* of a sequence can be measured (in bits or bytes), (2) the relative *distance* of a sequence from another sequence can be measured, (3) the *specificity* of a sequence cannot be measured, and (4) the *meaning* of a sequence cannot be measured (and may not exist).

The fact that in *all* sequences – organic or linguistic – specificity can only be named, inevitably raises a question: if specificity, or sequence information, cannot be measured, what is it? What does it represent in Nature? This is an extremely important point, because, as we will see, the difference between the Chemical paradigm and the Information paradigm is precisely the fact that they give different answers to that question. The *nature* of sequence information, in other words, is nothing less than the key concept that divides the two present paradigms of biology.

Shannon's Information Theory

The concept of information has been introduced in science in two very different ways. In biology, as we have seen, the concept of genetic information was introduced by Watson and Crick in 1953 and is identified with the specific sequence of nucleotides. In engineering, the information quantity of a message is measured by an entropy-like formula introduced by Claude Shannon in 1948, and is referred to as *statistical information*.

Shannon was particularly interested in telephone transmissions and described any communication system as a combination of a source (that produces signals), a destination (that receives them) and a channel in between. He realized that it is practically impossible to remove the effects of noise in *analog* signals, and insisted that communication must use *digital* messages. More precisely, Shannon proved, in his famous Capacity Theorem, that the effects of noise can be reduced by as much as we want to when messages are digitized. This means that *reliable* communication is possible even through *unreliable* channels, a result which paved the way to the tremendous expansion and success of the communication technologies.

The goal of communication is the reliable transmission of *all* messages, whatever is their meaning, and this is why in engineering information has been sharply separated from meaning. In his seminal papers, Shannon expressed this concept in no uncertain terms:

> "The fundamental problem of communication is that of reproducing at one point, exactly or approximately, a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem."

It must be underlined that Shannon avoided the semantic aspects of communication, but not the syntactic ones. The messages that are transmitted from a sender to a receiver have characters arranged in a specific order, and it is precisely that specific order that must be reconstructed at the receiver's end. The transmission of messages, in short, is the transmission of *sequence information*, or *specificity*, but it must be underlined that Shannon information is *not* sequence information. This point is worth calling attention to, because it has been the source of much confusion. Let us not forget that sequence information cannot be measured whereas Shannon's information is measurable.

Shannon conceived information as an entity that is generated whenever uncertainty is reduced, so he measured information by measuring *changes in uncertainty* (Shannon 1948). To this purpose he proposed that the number of binary decisions that are necessary to take in order to completely identify a sequence represents a measure of that sequence. In reality it is a measure of the *length* of a sequence, and it is expressed in bits, but it is obtained from considerations of probability theory whereas other methods measure the length of a sequence in more deterministic ways. The fact that Shannon's measure is expressed by an entropy-like formula explains why it is has been called *Shannon information*, and this unfortunately has created much confusion in biology where information is associated with specificity, not with size.

Shannon was able to prove a number of theorems on the ability of any communication system to transmit messages and established in this way an entirely new field of research which has become known as 'Information Theory'. In engineering, this field has been extremely successful but in biology its impact has been much more limited. The reasons, however, are still far from clear, and it may well be that the biological potential of Shannon's theory has not yet come to light and remains a challenge for the future.

The Ontological Claim of the Information Paradigm

The discovery of biological information was the event that transformed biochemistry into molecular biology, and the paradigm that 'life is chemistry' into the new paradigm that 'life is chemistry-plus-information'. This idea implies that information is *ontologically* different from chemistry, but can we prove it? Ontology is the study of being and saying what an entity is amounts to defining it. Ontology, in short, is concerned with the definition of entities at the most basic level.

The ontological claim of the Information paradigm is that life is *fundamentally* different from matter because there is an *ontological* difference between information and chemistry. Unfortunately, however, the proponents of the Information paradigm have never proved this claim. Ernst Mayr, as we have seen, has repeatedly stated that the existence of the genetic code is enough to prove that life in fundamentally different from chemistry, but has not been able to say why.

Perhaps the strongest criticism of the chemical paradigm has come from the Information Theory camp, and in particular from Hubert Yockey, one of the organizers of the first congress dedicated to the introduction of Shannon's information in biology (Yockey et al. 1958). In a long series of articles and books, Yockey (1974, 1992, 2000, 2005) has underlined that heredity is transmitted by factors that are "segregated, linear and digital" whereas the compounds of chemistry are "blended, three-dimensional and analog".

> "Chemical reactions in non-living systems are not controlled by a message. If the genetic processes were purely chemical, the law of mass action and thermodynamics would govern the placement of amino acids in the protein sequences according to their concentrations … There is nothing in the physico-chemical world that remotely resembles reactions being determined by a sequence and codes between sequences" (Yockey 1992)

Yockey has tirelessly pointed out that no amount of chemical evolution can cross the barrier that divides the analog world of chemistry from the digital world of life, and concluded from this that the origin of life cannot have been the result of chemical evolution. This is therefore, according to Yockey, what divides life from matter: information is ontologically different from chemistry because linear and digital sequences cannot be generated by the analog reactions of chemistry.

At this point one would expect to hear from Yockey how did linear and digital sequences appear on Earth, but he did not face that problem. He claimed instead that the origin of life is *unknowable*, in the same sense that there are propositions of logic that are *undecidable*. The problem, with this argument, is that the existence of undecidable propositions has been *proven* in logic, whereas the conclusion that the origin of life is unknowable is just an assumption. It may be a legitimate assumption, in principle, but in no way it is comparable to Gödel's theorem and certainly it does not carry the same weight.

It is important however to recognize that Yockey's distinction between analog and digital entities cannot be ignored. He was absolutely right in saying that the spontaneous reactions of chemistry cannot produce molecules with linear and digital sequences, and this is indeed the crucial problem that must be faced by any scientific theory on the origin of life. The information paradigm has not solved this problem and it is for this reason that it has not been able to prove its ontological claim.

The Ontological Claim of the Chemical Paradigm

The view that *'life is chemistry'* was proposed for the first time by Jan Baptist van Helmont (1648), and has been re-proposed countless times ever since. One of the most recent formulations has been given by Günther Wächtershäuser (1997) in these terms "If we could ever trace the historic process backwards far enough in time, we would wind up with an origin of life in purely chemical processes".

He added that "The science of chemistry, however, is an ahistoric science striving for universal laws… so this is the challenge of the origin of life: to reduce the historic process of biological evolution to a universal chemical law of evolution". The difficulty of this task, he pointed out, is due to the fact that "Chemistry is mechanistic and history teleological, and the life sciences are the arena where mechanistic explanations and teleological understanding come into close encounter."

Wächtershäuser claimed that "information is a *teleological* concept", and gave a specific example of the conflict between mechanism and teleology: "On the level of nucleic acid sequences it is quite convenient to use the information metaphor … and apply teleological notions such as 'function' or 'information'… but in the course of the process of retrodiction the teleological notions, whence we started, fade away. And what remains is purely chemical mechanism". This amounts to saying that biological information, the most basic concept of molecular biology, does not *really* belong to science.

This is the ontological claim of the Chemical paradigm, the idea that all natural processes are completely described, in principle, by physical quantities. This view is also known as *physicalism*, and it is based on the fact that biological information, or biological *specificity*, is not a physical quantity. So, what is it? A similar problem arises with the genetic code. The rules of a code cannot be measured and cannot be reduced to physical quantities. So what are they?

According to physicalism, biological information and the genetic code are mere *metaphors*. They are linguistic expressions that we use as shortcuts in order to avoid repeating every time all the details of long chains of chemical reactions. But behind those terms there are only chemical reactions and nothing else. They are like those computer programs that allow us to write our instructions in English, thus saving us the trouble to write them with the binary digits of the machine language. Ultimately, however, there are only binary digits in the machine language of the computer, and in the same way, it is argued, there are only physical quantities at the most fundamental level of Nature.

This conclusion, known as *the physicalist thesis*, has been proposed in various ways by a number of scientists and philosophers (Chargaff 1963; Sarkar 1996, 2000; Mahner and Bunge 1997; Griffith and Knight 1998; Griffith 2001; Boniolo 2003), and it is equivalent to the thesis that '*life is chemistry*'.

This is one of the most deeply dividing issues of modern science. Many biologists are convinced that biological information and the genetic code are real and fundamental components of life, but physicalists insist that they are real only in a very superficial sense and that there is nothing fundamental about them because they must be reducible, in principle, to physical quantities.

The Idea That "Life is Artifact-Making"

According to the chemical paradigm, the first cells evolved from chemical systems by spontaneous chemical reactions that are all fully described, in principle, by physical quantities. No other entities are required to explain the origin of life by chemical evolution, and this is why physicalism concludes that biological information and the genetic code are purely metaphorical terms.

It must be underlined that the physicalist thesis would be absolutely correct if genes and proteins were spontaneous molecules because there is no doubt that all spontaneous reactions are completely accounted for by physical quantities. This, however, is precisely the point that molecular biology has proved wrong. Genes and proteins are *not* produced by spontaneous processes in living systems. They are produced by molecular machines that physically stick their subunits together according to sequences and codes and are therefore *manufactured molecules*, i.e., *molecular artifacts*. This in turn means that all biological structures are manufactured, and therefore that the whole of life is *artifact-making* (Barbieri 2004, 2006, 2008). This conclusion may appear paradoxical, at first, but let us take a closer look.

All chemical reactions are either spontaneous or catalyzed processes, and biochemistry has clearly shown that virtually all reactions that take place in living systems are catalyzed processes. What molecular biology has discovered is that the production of genes and proteins requires not only catalysts but also *templates*. The catalysts join the subunits together by chemical bonds, and the templates provide the *order* in which the subunits are assembled. It is precisely that order that determines biological specificity, the most important characteristic of life, and that order comes from a molecule that is *outside* the assembled molecule.

This is precisely the characteristic that divides spontaneous objects from artifacts. In spontaneous and in catalyzed processes, the order of the components comes *from within* the molecules, i.e., is determined by *internal* factors, whereas in genes and proteins it comes *from without*, from an *external* template.

The difference between spontaneous and manufactured objects, in short, does not exists only at the macroscopic level of culture. It exists also at the molecular level, because it is an *experimental fact* that genes and proteins are manufactured molecules. It is also an experimental fact that they are *template-dependent* molecules, and this means that they are molecular artifacts.

Let us now look at the difference between the processes that manufacture genes and proteins. They both require catalysts and templates, but in addition to that proteins also require a set of coding rules (in the form of molecular adaptors). This is because genes are nucleic acids that are formed by copying a template, whereas proteins cannot be copied. Their order must still come from nucleic acids (because only these molecules can be inherited) but a sequence of nucleic acid has to be translated into a sequence of amino acids and this is achieved, in protein synthesis, by the rules of the genetic code.

We realize in this way that there are two distinct processes at the basis of life: the *copying* of genes and the *coding* of proteins. Genes are manufactured by molecular machines that can be referred to as *copymakers* and proteins by molecular machines that can be called *codemakers*. Copying and coding, on the other hand, are both artifact-making processes and life as we know it requires both of them. We can truly say therefore that *life is artifact-making*, or, more precisely, that *life is artifact-making by copying and coding*.

This makes us realize that the physicalist thesis is wrong because it is only spontaneous processes, not *all* processes, that are completely described by physical quantities. Manufacturing processes require additional entities, like sequences and coding rules, that are not physical quantities, because they cannot be measured, but which are absolutely essential to the description of all living systems.

The Origin of Linear and Digital Sequences

The existence of linear and digital sequences in life is a fact, an experimental fact, and all biologists acknowledge it. It is equally a fact that linear and digital sequences that direct the synthesis of molecules do not exist in the inanimate world, so it is beyond dispute that a divide does exist between life and matter. It is the divide between the analog world of chemistry and the digital world of life, and it is not a fiction. The problem is the origin of that divide, not its existence.

Hubert Yockey has underlined that spontaneous reactions cannot produce a living cell, and that, let us repeat it, is formally correct. The real answer to Yockey is not a denial of this point, but the argument that it does not apply to living cells because spontaneous reactions simply do not exist in them. The evidence shows that genes and proteins are manufactured by molecular machines in all present cells, and the most logical conclusion we can draw is that this has been true also for all the cells of the past, including the first cells.

Yockey's critique of chemical evolution is justified only if we assume that chemical evolution was but a sequence of *spontaneous* reactions, because linear, digital and specific properties do not exist in spontaneous processes. But they do exist in all manufacturing processes, including those that take place at the molecular level. The answer to Yockey's argument, in short, is that genes and proteins are molecular artifacts, that life itself is artifact-making (Barbieri 2003, 2008).

When a copymaker scans a nucleic acid and makes a copy of that molecule, what is happening is precisely an operation that brings into existence a linear and digital copy of a pre-existing molecule. It was molecular copying, the simplest form of artifact-making, that started manufacturing biological objects and set in motion the odyssey of life on the primitive Earth.

What is particularly important point, to our purposes, is that we now understand why it is possible that life evolved from inanimate matter and yet it is fundamentally different from it. The divide between life and matter is real because inanimate matter is made of spontaneous structures and life is made of manufactured objects. The idea that life is artifact-making, in short, is the only logical alternative to the chemical paradigm, and allows us to study the origin of life as a natural phenomenon that was brought into existence by the evolution of molecular machines.

A Useful Metaphor

We find it difficult to accept that life evolved from matter and, at the same time, that it is fundamentally different from it. How can something give origin to something fundamentally different from itself? The way out of this dilemma, as we have seen, is the idea that life is artifact-making, i.e., that the fundamental properties of life did not arise spontaneously from inanimate matter but were brought into existence by molecular machines. This idea, however, does not seem intuitively appealing, so it may be useful to illustrate it with a metaphor. It is a sort of cartoon, if you like, but if used consistently it is as rigorous as a technical argument.

The metaphor consists in saying that all spontaneous molecules are 'grey' (all shades of grey between white and black), whereas all manufactured molecules are 'coloured' (all colours of the rainbow). With this terminology, the concept that life is artifact-making amounts to saying that the world of life is coloured whereas the world of inanimate matter is grey, and this gives us a new way of formulating the problem of the origins. Earth was a lifeless planet, at the beginning, and all its molecules were grey, so how did coloured molecules appear out of grey matter?

Spontaneous genes and spontaneous proteins did appear on the primitive Earth but they did not evolve into the first cells, because spontaneous processes do not have biological specificity. They gave origin to *molecular machines* and it was these machines and their products that evolved into the first cells. The simplest molecular machines that could appear spontaneously on the primitive Earth were molecules that could stick monomers together at random (*bondmakers*) or in the order provided by a template (*copymakers*). These molecules started manufacturing polymers such as polypeptides, polynucleotides and polysaccharides, and had the potential to produce them indefinitely, thus increasing dramatically their presence on the primitive Earth. The unlimited repetition of copying, furthermore, is inevitably accompanied by errors, and in a world of limited resources a selection is bound to take place. That is how natural selection came into being, and that is why there is no natural selection in the spontaneous reactions of chemistry.

It must be underlined that the origin of molecular copying does require extremely improbable events. In a primitive environment where chemical evolution had already accumulated many varieties of organic molecules, the appearance of bondmakers and copymakers was as likely as that of any other average-size structure. The origin of proteins, on the other hand, was a much more complex affair, because proteins cannot be copied and their reproduction required the evolution of supramolecular systems that developed a *code* and which can therefore be referred to as *codemakers*. The evolution of the molecular machines, in short, started with bondmakers, went on to copymakers and finally gave rise to codemakers.

If we translate all this in the terminology of grey and coloured molecules, we can say that the first molecular machines were grey (because they appeared spontaneously) and that they started producing coloured molecules (because manufactured molecules are coloured). The first molecular machines were therefore a special type of grey molecules, and we may call them 'silver' molecules. The machines that came after them, however, could incorporate also coloured molecules, and eventually these replaced all grey elements in them. The silver molecular machines evolved into coloured machines and we can illustrate this transformation by saying that they became 'golden' molecular machines. At this stage, the divide between life and matter became complete, because all the components of life, molecules and molecular machines, were coloured, whereas all the components of inanimate matter were grey.

The Stalemate

The double helix and the genetic code have been two of the major scientific discoveries of all times and yet, surprisingly, the *old regime* has not been deposed. The majority view, today, is still the idea that life is an extremely complex form of chemistry that evolved spontaneously from primitive chemical systems. This view is based on the physicalist thesis that all biological processes are completely described, in principle, by physical quantities, which means that there is nothing fundamental in them. Entities like genetic information and coding rules are regarded as metaphorical terms that people use simply because they are intuitively appealing. In such a framework, the revolution of molecular biology amounts to little more than the introduction of fancy names into the solid body of biochemistry.

This is the great paradox of modern biology. On the one hand, genetic information and the genetic code have become the bread and butter of biological research, and on the other hand we are told that they are mere linguistic decorations. The paradox is due to the fact that the information paradigm has *claimed* that genetic information is a new fundamental entity, but has not been able to say *why*. The present stalemate between the two paradigms of modern biology, in other words, is due to the fact that the information paradigm has not offered a proper alternative to the chemical paradigm.

Here we have seen that such an alternative does exist, because the physicalist thesis is valid only in spontaneous systems whereas genes and proteins are never formed by spontaneous reactions. They are invariably manufactured by molecular machines, and all manufacturing processes do not require only physical quantities but also additional entities like sequences and codes. The alternative to the view that '*life is chemistry*', in short, is the view that '*life is artifact-making*'.

Unfortunately, modern biology has accepted the concept of information but not the concept of meaning, and this is equivalent to saying that genetic information is real but the genetic code is not. In the case of meaning, in other words, the information paradigm has accepted the physicalist thesis, and in so doing it has compromised the possibility of demonstrating its own thesis: if the genetic code is a metaphor, why should genetic information be different?

What we need, therefore, is a new paradigm that fully accepts the implications of the discovery that life is based on copying and coding, and that these processes necessarily require sequences and codes. We need a paradigm where biological sequences (organic information) and coding rules (organic meaning) are *real* and *fundamental* entities of Nature, as real and fundamental as the physical quantities.

PART 2 The Code Paradigm

Schrödinger's Prophecy

In 1944, Erwin Schrödinger wrote *"What is Life?"*, a little book that inspired generations of physicists and biologists and became a landmark in the history of molecular biology. There were two seminal ideas in that book: one was that the genetic material is like an *"aperiodic crystal"*, the other was that *"the chromosomes contain a code-script for the entire organism"*. The metaphor of the aperiodic crystal was used by Schrödinger to convey the idea that the atoms of the genetic material must be arranged in a unique pattern in every individual organism, an idea that later was referred to as *biological specificity.* The metaphor of the code-script was used to express the concept that there must be *a miniature code* in the hereditary substance, a code that Schrödinger compared to *"a Morse code with many characters"*, and that was supposed to carry *"the highly complicated plan of development of the entire organism"* (Schrödinger 1944). That was the very first time that the word 'code' was associated with a biological structure and was given a role in organic life.

The existence of specificity and code at the heart of life led Schrödinger to a third seminal conclusion, an idea that he expressed in the form of a prophecy: *"Living matter, while not eluding the 'laws of physics' as established up to date, is likely to involve hitherto unknown 'other laws of physics', which, however, once they have been revealed, will form just an integral part of this science as the former"*.

Schrödinger regarded this prophecy as his greatest contribution to biology, indeed he wrote that it was *"my only motive for writing this book"*, and yet that is the one idea that even according to his strongest supporters did not stand up to scrutiny. Some 30 years later, Stent and Calendar (1978) gave up the struggle and concluded that *"No 'other laws of physics' turned up along the way. Instead, the making and breaking of hydrogen bonds seems to be all there is to understanding the workings of the hereditary substance"*.

Schrödinger's prophecy of new laws of physics appears to have been shipwrecked in a sea of hydrogen bonds, but in reality that is true only in a superficial sense. The essence of the prophecy was the idea that the two basic features of life - specificity and the genetic code - require *new fundamental entities* of Nature that are *"hitherto unknown"*, and in that form it is still valid. The fact that Schrödinger invoked new *laws of physics* should not have obscured the substance of the prophecy, which can be expressed in this way: in order to understand life we need to discover something fundamentally new, something that is still not part of physical theory.

Let us turn therefore to this generalized version of Schrödinger's prophecy. He anticipated the concept of biological specificity (what today we call biological sequences, or biological information), and announced that there must be a 'code-script' in every living cell. Both ideas were truly prophetic, at the time, and both turned out to be true. That should be enough for us take a new look at the *essence* of his prophecy: is it true that we need something fundamentally new in order to explain biological information and the genetic code?

The 'Special Constraints' Solution

In the 1960s, Howard Pattee pointed out that the genetic code is fully compatible with the theory developed by John von Neumann on self-replicating machines. Von Neumann had shown that a self-replicating system capable of open-ended evolution must necessarily contain a description of itself, and such a description must be categorically different from the controlled system ("the map is not the territory"). The description of a system, on the other hand, is necessarily made of entities that represent, or ‘stand for’, its material components, and function therefore as signs or symbols. According to von Neumann, in short, an evolvable self-replicating system must be *a physical system controlled by symbols*, or, more precisely, by a programme, by the rules of a code (von Neumann 1951, 1958, 1966).

This was enough, according to Pattee, to prove that every living cell is controlled by a real code, and he set out to find out how physical theory can account for the existence of the genetic code without resorting to the Schrödinger solution of “*new laws of physics*”. To this purpose, Pattee focussed on the idea that physical theory does not consists only of physical laws, but of laws *plus* initial conditions and boundary conditions, both of which are often referred to as *constraints*.

This had been known since Newton’s time, of course, but physicists had consistently assumed that laws are fundamental whereas constraints have only an accessory role. The reality, however, turned out to be very different. Murray Gell-Mann (1994) has underlined that “the effective complexity of the universe receives only a small contribution from the fundamental laws. The rest comes from ‘frozen accidents’, which are precisely the result of constraints. All planets, for example, are formed according to universal physical laws, and yet they are all different. Their individual features are due to the particular constraints of their development, and the distinction between laws and constraints is so important that Eugene Wigner (1964) called it “Newton’s greatest discovery”.

In this novel theoretical framework where laws and constraints have equally fundamental roles, Pattee argued that information and codes are perfectly compatible with physical theory because they have precisely the defining features of constraints. The rules of a code, for example, are limitations that drastically reduce the number of possibilities and can be regarded therefore as true natural constraints. In a similar way, Claude Shannon underlined that information is obtained whenever uncertainty is reduced, and concluded from this that the notions of information and constraint are interchangeable (Shannon 1948).

The solution proposed by Pattee, in short, is that information and codes do not require new laws of physics, because they are *a special type* of constraints and constraints are an integral part of physical theory (Pattee 1968, 1972, 1980, 1995, 2001, 2008). This is the *‘special constraint’* solution to the problem of the genetic code, a solution that is developed in three logical steps: (1) life requires self-replication (a biological principle), (2) evolution requires symbolic control of self-replication (von Neumann), and (3) physics requires that symbols and codes are special types of constraints (Pattee).

Such a conclusion, however, is not entirely satisfactory. It is certainly true that sequences and codes have the defining characteristics of constraints, but not all constraints lead to life, far from it, and it is not enough to say that they must be ‘special’ constraints. What is it that makes them special? What is it that distinguish the special constraints of information from the special constraints of the genetic code, and what is it that distinguish both of them from the countless constraints of inanimate matter?

The New Observables

Howard Pattee has pointed out that biology does not need new laws of physics because physical theory is based on laws and constraints, and entities like symbols and codes can be regarded as special types of constraints. This is undoubtedly true, but it is not the whole truth. Physical theory starts with the definition of fundamental entities, or *observables* (time, space, mass etc), and then looks for relationships between them which are referred to as laws and constraints. The basic components of physical theory, in short, are not two but three: laws, constraints, and observables.

The important point here is that the history of physics has not been made only by the discovery of new laws and new constraints, but also by the discovery of new observables. In Newton's physics, for example, the fundamental observables were time, space and mass, but then electricity required the addition of electric charge and thermodynamics required the addition of temperature.

If we assume *a priori* that life does not need new observables, we can limit ourselves to laws and constraints, but this is precisely the point that we cannot take for granted. Life is based on the copying of genes and on the coding of proteins and these processes require entities, like biological sequences and the rules of a code, that have all the defining characteristics of *new observables*. This is because the role of observables is to allow us to describe the world and we simply cannot describe living systems without sequences and codes. But what kind of entities are these new observables?

A biological sequence is a linear chain of units that represents *organic information*, and a biological code is a set of rules that associate an *organic meaning* to each unit of information. Sequences and codes, in short, are carriers respectively of organic information and organic meaning, and our problem is to understand the nature of these entities.

According to a long tradition, natural entities are divided into *quantities* and *qualities*. Quantities can be measured and are objective, whereas qualities are subjective and cannot be measured. In the case of organic information and organic meaning, however, this scheme breaks down. Organic information, for example, is not a quantity because a specific sequence cannot be measured. But it is not a quality either, because linear specificity is a feature that we find in organic molecules, and is therefore an objective feature of the world, not a subjective one. The same is true for organic meaning. This too cannot be measured, so it is not a quantity, but it is not a quality either because the rules of the genetic code are the same for all observers in all living systems.

A scheme based on quantities and qualities alone, in short, is not enough to describe the world. In addition to quantities (*objective and measurable*) and qualities (*subjective and not-measurable)* we must recognize the existence in Nature of a third type of entities (*objective but not-measurable*).

Organic information and organic meaning belong precisely to that new type of entities, and we can also give them a suitable name. Since organic information and organic meaning can be described only by *naming* their components, we can say that they are *nominable* entities, or that they belongs to the class of the nominable entities of Nature (Barbieri 2004, 2006, 2008).

It must be underlined that the existence of new observables in living systems is perfectly compatible with physics, because observables are an integral part of physical theory and the discovery of new observable has gone on throughout the history of science. Let us take therefore a closer look at these new natural entities and see if we can learn something more about them.

Names and 'Nominable' Entities

Physical theory consists of laws, constraints and observables, but in addition to these three components there is also a fourth one that should be taken into account, and that is *names*. Science is always expressed in words and we need therefore to give names to the objects and the processes that we observe in Nature. Names (including those that we call 'numbers') are necessarily a fourth essential component of physical theory, but are different from the first three because they change from one language to another. Laws, constraints and observables, in other words, do not depend upon the language that is employed to express them, whereas names are totally language-dependent. This is because names (or *nominal entities*, to use a classical term) in general have nothing to do with the intrinsic features of the named objects, and are therefore mere labels that we attach to them.

The deep divide that exists between 'names' and 'objects' has been at the centre of many controversies in the past, in particular of the celebrated medieval dispute over 'nominal entities' and 'real entities'. It has also had a long history in the philosophy of mathematics, where some have argued that numbers are 'invented' by the human mind, and others that they are 'discovered', a conclusion which implies that they have an existence of their own in some abstract Platonic world.

The relationship between names and objects is also a crucial issue in science, but here it has taken on a new form. Let us underline that all names are sequences of characters (alphabetic, numerical or alpha-numerical) and that each sequence is unique. Names, in other words, have *specificity.* In general, the specificity of a name has nothing to do with the characteristics of the named object, and in these cases we can truly say that names are mere labels. Science, however, has invented a new type of names where the sequence of characters does represent an order that is objectively present in the named objects.

The chemical formula of a molecule, for example, describes an objective sequence of atoms, and any atom can be described by the objective sequence of its quantum numbers. In these cases, the names are no longer arbitrary labels but true '*observables' because they describe characteristics that we observe in Nature. This shows that there are two distinct types of names in science: labels and observables.*

In the case of the observables, furthermore, there is another distinction that must be considered. When a molecule is formed spontaneously, its final sequence is due to the interactions between its own components, and in most cases it is completely determined by them. In the case of a protein, however, all its different amino acids interact by the same peptide bonds and a spontaneous assembly would produce a completely random order (which is incompatible with life). In this case, a specific sequence can be obtained only if the amino acids are put together by a molecular machine according to the order provided by a template that is *external* to the protein itself. We need therefore to distinguish between two different types of observables.

The sequence of quantum numbers in an atom, or the sequence of atoms in inorganic molecules, is determined *from within*, by internal factors, whereas the sequence of amino acids in a protein is determined *from without,* by external templates. In the first case the sequence is a *physically computable* entity, in the sense that it is the automatic result of physical forces, whereas in the second case it can only be described by 'naming' its components, and is therefore a *nominable* entity (this term should not be confused with the classical concept of *nominal* entity, which applies to all names). A *nominable* entity is not a label but an observable, and more precisely a *non-computable* observable.

All names, in conclusion, are specific sequences of characters, and in science they can be divided into two great classes: labels and observables. The observables, in turn, can be divided into *computable* entities and *nominable* entities. The important point is that physics and chemistry deal exclusively with computable entities (physical quantities), whereas nominable entities (information and coding rules) exist only in living systems. We need therefore to pay a special attention to these new observables, and make sure that they truly are fundamental entities of Nature.

Organic Information

In genes and proteins, biological, or organic, information has been defined as the specific sequence of their subunits. This definition however is not entirely satisfactory because it gives the impression that information is a *static* property, something that molecules have simply because they have a sequence. In reality, there are countless molecules which have a sequence but only in a few cases this becomes information. That happens only when copymakers use it as a guideline for copying. Even copymakers, however, do not account, by themselves, for information. Copymakers can stick subunits together and produce sequences, but without a template they would produce only *random* sequences, not specific ones. Sequences alone or copymakers alone, in other words, have nothing to do with information. It is only when a sequence provides a guideline to a copymaker that it becomes information for it. It is only an act of copying, in other words, that brings organic information into existence.

This tells us that organic information is not just the specific sequence of a molecule, but *the specific sequence produced by a copying process*. This definition underlines the fact that organic information is not a thing or a property, but the result of a process. It is, more precisely, an 'operative' definition, because information is defined by the process that brings it into existence. We realize in this way that organic information is as real as the copying process that generates it.

We have also seen that organic information is neither a quantity (because a specific sequence cannot be measured), nor a quality (because it is an objective feature of all copied molecules), and belongs instead to a third class of objects that have been referred to as *nominable* entities (Barbieri 2004, 2006, 2008).

We conclude that organic information is a new type of objects, and that it is essential to describe the organic molecules of Nature. To this purpose, in fact, it is no less essential than the physical quantities, and this means that organic information *has the same scientific 'status' as a physical quantity*. They both belong to the class of objective and reproducible entities that allow us to describe the world.

This conclusion, however, raises immediately a new problem, because there are two distinct groups of physical quantities: a small group of *fundamental* quantities (space, time, mass, charge and temperature) and a much larger group of *derived* quantities. That distinction applies to all objective entities, so we need to find out whether organic information belongs to the first or to the second group.

Luckily, this problem has a straightforward solution because the sequences of genes and proteins have two very special characteristics. One is that *a change in a* *single component of a biological sequence may produce a sequence which has entirely new properties*. This means that although a biological sequence can be said to have 'components', it is at the same time a single indivisible whole. The second outstanding feature is that *from the knowledge of n elements of a biological sequence we cannot predict the element (n+1)*. This is equivalent to saying that *a specific sequence cannot be described by anything simpler than itself*, so it cannot be a derived entity.

We conclude that organic information has the same scientific status as the physical quantities, because it is an objective and reproducible entity. But we also conclude that it does not have the status of a derived physical quantity because it cannot be expressed by anything simpler than itself. This means that organic information has the same scientific status as the fundamental quantities of physics, and is therefore a new irreducible entity of Nature, i.e., a new fundamental observable.

Organic Meaning

A code is a set of rules which establish a correspondence between the objects of two independent worlds. The Morse code, for example, is a correspondence between groups of dots and dashes with the letters of the alphabet, and in the same way the genetic code is a correspondence between groups of nucleotides and amino acids. Let us notice now that establishing a correspondence between, say, object 1 and object 2, is equivalent to saying that object 2 is the meaning of object 1. In the Morse code, for example, the rule that 'dot-dash' corresponds to the letter 'A', is equivalent to saying that letter 'A' is the meaning of 'dot-dash'. In the code of the English language, the mental object of the sound 'apple' is associated to the mental object of the fruit 'apple', and this is equivalent to saying that that fruit is the meaning of that sound.

By the same token, the rule of the genetic code that a group of three nucleotides (a codon) corresponds to an amino acid is equivalent to saying that that amino acid is the *organic meaning* of that codon. Anywhere there is a code, be it in the mental or in the organic world, there is meaning. We can say, therefore, that *meaning is an entity which is related to another entity by a code*, and that organic meaning exists whenever an organic code exists (Barbieri 2003, 2008).[1]

The existence of meaning in the organic world may seem strange, at first, but in reality it is no more strange than the existence of a code because they are the two sides of the same coin. To say that a code establishes a correspondence between two entities is equivalent to saying that one entity is the meaning of the other, so we cannot have codes without meaning or meaning without codes. All we need to keep in mind is that *meaning is a mental entity when the code is between mental objects, but it is an organic entity when the code is between organic molecules*.

Modern biology has readily accepted the concept of information but has carefully avoided the concept of meaning, and yet organic information and organic meaning are both the result of natural processes. Just as it is an act of copying that creates organic information, so it is an act of coding that creates organic meaning. Copying

[1] The definition of meaning and semiosis in terms of coding has been discussed in depth by Stefan Artmann (2007, 2009).

and coding are the processes; copymakers and codemakers are their agents; organic information and organic meaning are their results.

But the parallel goes even further. We have seen that organic information cannot be measured, and the same is true for organic meaning. We have seen that organic information is an objective entity, because it is defined by the same sequence for any number of observers, and that is also true for organic meaning, which is defined by coding rules that are the same for all observers. Finally, we have seen that organic information is an irreducible entity, because it cannot be described by anything simpler than its sequence, and the same is true for organic meaning, which cannot be defined by anything simpler than its coding rules.

Organic information and organic meaning, in short, belong to the same class of entities because they have the same defining characteristics: they both are *objective-but-not-measurable* entities, they both are *fundamental* entities because they cannot be reduced to anything simpler, and they both are *nominable* entities because we can describe them only by naming their components (Barbieri 2004, 2008).

Finally, let us underline that they are the twin pillars of life because organic information comes from the copying process that produces genes, while organic meaning comes from the coding process that generates proteins.

Operative Definitions

Physical quantities have three fundamental properties: (1) they are objective, (2) they are reproducible, and (3) they are defined by operative procedures. This last property is particularly important because it has provided the solution to one of the most controversial issues of physics. The controversy was about the theoretical possibility that the entity which is measured may not be the same entity which has been defined. This led to the idea that there should be no difference between what is measured and what is defined, i.e., to the concept of operative (or operational) definition: *a physical quantity is defined by the operations that are carried out in order to measure it*.

It was this operational approach that solved the definition problem in physics, and it is worth noticing that we can easily generalize it. Rather than saying that a natural entity is defined by the operations that measure it, we can say that a *natural entity is defined by the operations that evaluate it in an objective and reproducible way*. The advantage of this generalized formulation is that it applies to *all* objective entities, so it can be used not only in physics, but in biology as well. To this purpose, we only need to notice that *a* measurement is an objective and reproducible description of a physical quantity, just as the naming of a specific sequence is an objective and reproducible description of organic information, and just as the naming of a coded entity is an objective and reproducible description of organic meaning.

Whereas the physical quantities are evaluated *by measuring*, sequences and codes are evaluated *by naming their components*, but in both cases the entities in question are defined by the operations that evaluate them, and this is the essence of the operative approach. We may add that organic information and organic meaning can also be defined by the processes of copying and coding that bring them into existence, and that too amounts to an operative definition (Barbieri 2003, 2008).

We conclude that organic information and organic meaning can be defined by generalized operative procedures that are as reliable as the operative procedures of physics. This means that the definitions of information and meaning should no longer be at the mercy of endless debates on terminology as they have been in the past. The operative definitions are scientific tools which are justified by their own prescriptions, so there is no point in asking whether they are right or wrong. All we can ask of them is whether they contribute or not to our description and to our understanding of Nature.

At this point, we can summarize all the above arguments with the following concepts:

(1) The sequence used by a copymaker during a copying process is *organic information*.
(2) The sequence produced by a codemaker during a coding process is an *organic meaning*.
(3) Organic information and organic meaning are neither quantities nor qualities. They are a new kind of natural entities that are referred to as *nominable* entities.
(4) Organic information and organic meanings have the same scientific status as the quantities of physics because they are *objective* and *reproducible* entities that can be defined by operative procedures.
(5) Organic information and organic meanings have the same scientific status as the *fundamental* quantities of physics because they cannot be reduced to, or derived from, simpler entities.

Are Operative Definitions out of fashion?

The idea that a physical quantity is defined by the operations that measure it is the core of *operationalism,* a school of thought that became popular in the 1920s and 30s, in particular after the publication of *The Logic of Modern Physics* by Percy W. Bridgman (1927).

Bridgman underlined that the operational method allows us overcome the age-old impasse that we find ourselves in when we wonder whether what we measure is actually what we have defined. But there is also much more than that. Bridgman pointed out that Einstein discovered the special theory of relativity precisely by analyzing the consequences of the way in which we measure space and time in conditions where the speed of light is finite. In that case, the operational method adopted by Einstein did not solve only a definition problem but was instrumental in fostering one of the greatest discoveries of all times. Another scientific revolution where operationalism had a crucial role was quantum physics. In that case the breakthrough came from analyzing the consequences that the measuring procedures have on physical quantities in conditions where energy comes in discrete packets.

Having solved those problems, however, physicists moved on to more general principles, such as conservation laws and symmetry breaking, and operationalism went out of fashion in their field. At the same time, on the other hand, operationalism was rediscovered in the social sciences and used for purposes that profoundly changed its original *raison d'etre*. It was adopted as a method that could give a scientific appearance to the formulation of social problems and psychological concepts. It became a set of procedures and questionnaires for defining things like "consumer satisfaction", "levels of frustration", "emotional involvement" and the like.

That probably explains why operationalism has lost much of its original appeal in science, and has not been applied to the definition of difficult and deep scientific concepts such as information and codes. And yet these concepts are as fundamental as those of space and time at the dawn of modern physics, and we should not forget that it was precisely the operational approach that paved the way to their understanding. The operational definitions of organic information and organic meaning, in short, are not mere definitions. They are a description of the process by which digital sequences are manufactured in the living world. They tell us how abstract entities like sequences and coding rules actually came into existence at the origin of life and have been brought into existence ever since.

At this point one could suggest that operationalism requires measurements and this makes it very similar to physicalism, the doctrine that observables are real only if they can be reduced, in principle, to physical quantities. In reality, physicalism is endorsed mostly by chemists and philosophers whereas physicists rarely support it. The introduction of quarks with "flavors" and "colors", for example, shows that non-quantitative observables can be accommodated in physics, and this may well be true also for the "nominable" entities of life. All we need, to that purpose, are operative procedures that show how these entities come into existence.

The Code Paradigm

The discoveries of the double helix and of the genetic code are the two pillars of modern biology, but there is a strange discrepancy between them. The first brought biological information to light and that concept was fully accepted into modern biology. The genetic code revealed the existence of biological *meaning* – because any code is a correspondence between signs and meanings – but that concept has been completely ignored by modern biology.

It is often said that the concept of meaning has also been kept out of Information Theory, but that is not the case. Information theory heavily relies on two types of codes that are known as source coding and channel coding. In both cases, one sequence is transformed into another, and apparently all that takes place is a replacement of *syntactic information* (or *specificity*). In reality, this operation is carried out by a codemaker that creates a correspondence between one syntactic sequence that represents a sign and a second syntactic sequence that represents its meaning. Any code, in other words, necessarily involves meaning, and the codes of Information theory are no exception.

Information theory, in short, is *not* independent from meaning. On the contrary, the mobile telephone, to name just one example, would not even exist without the introduction of error-correcting codes (Battail 2007, 2008), and all applications of Information Theory are heavily dependent on such codes. Information theory, in other words, does deal with codes and coding rules, but keeps meaning sharply distinct from information.

In biology, instead, no such clear distinction has been made, and meaning has been regarded not as an entity in its own right, but as a 'qualification' of information. Rather than talking of information and meaning, many biologists are talking of "meaningful information", "semantic information", "functional information" and the like.

In a recent review entitled *Information in Biological Systems* John Collier (2008) has listed at least seven different types of information that apparently form a nested hierarchy: (1) physical information (or "It from bit" information), (2) statistical information (or "negentropy"), (3) expressed information, (4) functional information, (5) meaningful information, (6) intentional information, and (7) social information.

Similar proposals have been made by many other authors with different terminologies, and there seem to be no end in sight to the proliferation of the information categories. But why does this happen? Why do we keep multiplying the types of information in order to account for properties that belong to the category of meaning? It is high time to acknowledge that in biology too we must face the issue of meaning, and to this purpose we should treasure the example of the communication sciences. We should accept that information and meaning are two distinct entities and stop trying to reduce one to the other.

The important point, at any rate, is that a genetic code exists in every cell, a fact which tells us that there are two distinct fundamental processes at the basis of life. The coding of proteins is as essential as the copying of genes and this implies that biological meaning is as necessary as biological information in living systems. This conclusion is nothing less than a new theoretical framework, and we have, therefore, three distinct paradigms in modern biology.

In addition to the idea that 'life is chemistry', and to the idea that 'life is chemistry-plus-information', we have a third paradigm which states that 'life is chemistry-plus-information-plus-codes'.

This is the *Code paradigm*, the idea that life is based on copying and coding, that we need to introduce in biology not only the concept of biological information but also the concept of biological meaning.

The Discovery of New Worlds

The history of physics tells us that scientific discoveries require three logical steps. First we look at the world and choose a certain number of entities to describe it, entities that are called *observables* (space, time, mass, etc.) precisely because they represent what we observe.. Then we look for relationships between observables and obtain models of the observed phenomena (regularities, equations, laws, etc.). Finally we use our models to make predictions that test them (we predict, for example, the nest eclipse of the moon etc.).

The choice of the observables is the first step in the procedure and the most critical. The movements of planets and stars, for example, can be described with only two observables - space and time - and in that case we get either a Ptolemaic model or a Copernican system. By introducing a third observable - mass - we obtain the laws of motion, universal gravitation and the Newton model of the world.

The three basic observables of classical physics can be combined together in different ways and produce many other derived observables (velocity, acceleration, force, energy, power, momentum, etc.), but what defines the whole system is the initial number of fundamental observables. The actual identity of these observables can be changed (space and time, for example, can be replaced by velocity and time, and in that case space becomes a derived entity), but the minimum number of fundamental observables does not change. That number defines a whole world of phenomena, and we can discover new worlds, i.e., new aspects of reality, only if we discover new fundamental observables. The world of electricity and magnetism, for example, required precisely the introduction of new fundamental observables, and so did the world of thermodynamics, the world of nuclear forces, and the world of elementary particles. All of which takes us to a question: do we need new observables in the world of life or not? This point is crucial, and the different paradigms of biology are nothing less than different ways to answer it.

The chemical paradigm states *a priori* that we do not need new observables to describe living systems, i.e., that life is completely described, in principle, by the quantities of physics. The information paradigm claims that information is a *fundamental* entity that exists only in living systems, but it has not been able to contrast the physicalist charge that there is nothing fundamental in it.

We can prove that this charge is wrong only by showing that information is a new *observable* and this can be done only by showing that information is the result of a manufacturing process by molecular copying. But as soon as we accept the reality of molecular copying we must also accept the reality of molecular coding, and therefore of another fundamental observable. This is the third paradigm of modern biology, the Code view of life, the idea that life is artifact-making by copying and coding.

The crucial point is that the existence of two new observables in living systems is not a hypothesis. It is an *experimental* fact. We can prove that biological sequences (organic information) and the rules of a code (organic meaning) are fundamental observables with the same procedures that we have used in the case of space, time, mass, temperature, etc. The only difference is that sequences and coding rules are *non-computable* observables, but there is no doubt that observables they are (we do observe them in living systems) and that they are *fundamental* observables (because we cannot describe living systems without them and because we cannot reduce them to anything else).

The discovery of classical physics, the discovery of thermodynamics, the discoveries of electromagnetism and of elementary particles, were all based on the discoveries of new fundamental observables, and now we realize that this is true also in biology. Life is indeed a new world, a new dimension of reality, because it is the result of copying and coding processes that bring two new fundamental observables into existence.

Conclusion

The idea that life is an extremely complex form of chemistry is still very popular, today, and is based on the physicalist thesis that all biological processes can be reduced, in principle, to physical quantities. According to this view, genetic information and the genetic code are not fundamental observables because they are not physical quantities. They are regarded instead as metaphorical and teleological terms that we use only because they are intuitively appealing.

We have seen however that the physicalist thesis is valid only in spontaneous systems, whereas genes and proteins are never formed spontaneously in real life. They are invariably manufactured by molecular machines, and all manufacturing processes do not require only physical quantities but also additional entities like sequences and coding rules.

The charge that information is a teleological concept is simply false, notwithstanding the fact that it is repeated fairly often. The truth is precisely the other way round. Information has all the defining features of a scientific concept because it has been defined in two different ways and in both cases there is nothing teleological about it.

(1) When it is defined by Shannon's approach, information is actually expressed by a formula, like any standard physical quantity.
(2) When it is defined as a sequence, information is no longer measurable, but it still is a fundamental observable because it is absolutely necessary to the *description* of living systems.

We simply cannot describe the transmission of genes or the synthesis of proteins without their sequences, and we cannot replace these sequences with anything else, which means that using information to describe living systems is perfectly equivalent to using space, time, mass and energy to describe physical systems. The truth, in other words, is that there is no more teleology in information and in the genetic code than there is in the quantities of physics and chemistry. Sequences (biological information) and coding rules (biological meaning) are *descriptive* entities and are absolutely essential to the scientific study of life.

The information paradigm, on the other hand, has claimed that information is distinct from chemistry but has not been able to say why. On top of that, it has accepted the concept of information but not the concept of meaning, which is equivalent to saying that genetic information is real but the genetic code is not, again without being able to say why.

We conclude therefore that we need a new paradigm that fully accepts the implications of the discovery of the genetic code. The implication that life is based on copying *and* coding, that 'life is artifact-making'. This is the code paradigm, the theoretical framework where biological sequences (organic information) and biological coding rules (organic meaning) are *real* and *fundamental* observables that are as essential to life as the fundamental quantities of physics.

Acknowledgments I wish to thank Gérard Battail, Cliff Joaslyn and Joachim De Beule for calling my attention to some critical points in the original draft of this paper and for suggesting very useful amendments.

References

Artmann, S. (2007). Computing codes versus interpreting life. Two alternative ways of synthesizing biological knowledge through semiotics. In M. Barbieri (Ed.), *Introduction to Biosemiotics. The new biological synthesis* (pp. 209–233). Dordrecht: Springer.

Artmann, S. (2009). Basic Semiosis as code-based control. *Biosemiotics, 2*(1), 31–38.

Barbieri, M. (2003). *The organic codes. An introduction to semantic biology.* Cambridge: Cambridge University Press.

Barbieri, M. (2004). The definitions of information and meaning. Two possible boundaries between physics and biology. *Rivista di Biologia-Biology Forum, 97*, 91–110.

Barbieri, M. (2006). Life and Semiosis: the real nature of information and meaning. *Semiotica, 158*(1/4), 233–254.

Barbieri, M. (2008). Biosemiotics: a new understanding of life. *Naturwissenschaften, 95*, 577–599.
Battail, G. (2007). Information theory and error-correcting codes in genetics and biological evolution. In M. Barbieri (Ed.), *Introduction to Biosemiotics* (pp. 299–345). Dordrecht: Springer.
Battail, G. (2008). *An outline of informational genetics*. Morgan & Claypool Publishers.
Boniolo, G. (2003). Biology without Information. *History and Philosophy of the Life Sciences, 25*, 255–273.
Bridgman, P. W. (1927). *The logic of modern physics*. Beaufort Books.
Chargaff, E. (1963). *Essays on nucleic acids*. Amsterdam: Elsevier.
Collier, J. (2008). Information in Biological Systems. In P. Adriaans & J. van Benthem (Eds.), *Handbook of philosophy of science, volume 8, philosophy of information*. Elsevier.
Danchin, A. (2009). Bacteria as computers making computers. *FEMS Microbiology Reviews, 33*, 3–26.
Descartes, R. (1637). *Discours de la Méthode*. Leiden, The Netherlands.
Gell-Mann, M. (1994). *The quark and the jaguar* (p. 134). New York: W. H. Freeman.
Griffith, P. E. (2001). Genetic information: a metaphor in search of a theory. *Philosophy of Science, 68*, 394–412.
Griffith, P. E., & Knight, R. D. (1998). What is the developmental challenge? *Philosophy of Science, 65*, 276–288.
Johannsen, W. (1909). *Elemente der exacten Erblichkeitslehre*. Jena: Gustav Fischer.
Lartigue, C., Glass, J. I., Alperovich, N., Pieper, R., Parmar, P. P., Hutchison, C. A., III, Smith, H. O., & Venter, J. C. (2007). Genome transplantation in bacteria: changing one species to another. *Science, 317*, 632–638.
Mahner, M., & Bunge, M. (1997). *Foundations of biophilosophy*. Berlin: Springer Verlag.
Maynard-Smith, J. (2000). The concept of information in biology. *Philosophy of Science, 67*, 177–194.
Mayr, E. (1982). *The growth of biological thought*. Cambridge: The Belknap Press of Harward University Press.
Pattee, H. H. (1968). The physical basis of coding and reliability in biological evolution. In C. H. Waddington (Ed.), *Toward a theoretical biology* (Vol. 1, pp. 67–93). Edinburgh University Press.
Pattee, H. H. (1972). Laws and constraints, symbols and languages. In C. H. Waddington (Ed.), *Towards a theoretical biology* (Vol. 4, pp. 248–258). Edinburgh: Edinburgh University Press.
Pattee, H. H. (1980). Clues from molecular symbol systems. In U. Bellugi & M. Studdart-Kennedy (Eds.), *Signed and spoken language: Biological constraints on linguistic form* (pp. 261–274). Dahlem Konferenzen: Verlag-Chemie.
Pattee, H. H. (1995). Evolving self-reference: matter, symbols, and semantic closure. *Communication and Cognition - Artificial Intelligence, 12*(1–2), 9–27. Special Issue Self-Reference in Biological and Cognitive Systems, Luis Rocha (ed.).
Pattee, H. H. (2001). The physics of symbols: bridging the epistemic cut. *BioSystems, 60*, 5–21.
Pattee, H. H. (2008). Physical and functional conditions for symbols, codes and languages. *Biosemiotics, 1* (2), 147–168.
Sarkar, S. (1996). Biological Information. A skeptical look at some central dogmas of molecular biology. In S. Sarkar (Ed.), *The philosophy and history of biology* (pp. 187–231). Dordrecht: Kluwer Academic.
Sarkar, S. (2000). Information in genetics and developmental biology. *Philosophy of Science, 67*, 208–213.
Schrödinger, E. (1944). *What is life?* Cambridge: Cambridge University Press.
Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal, 27*, 379–424. 623–656.
Stent, G. S., & Calendar, R. (1978). *Molecular genetics*. San Francisco: W.H. Freeman.
van Helmont, J.B. (1648). *Ortus Medicinae*. Amsterdam
von Neumann, J. (1951). General and logical theory of automata. In L. A. Jeffress (Ed.), *Cerebral mechanisms of behavior, the hixon symposium, vol. 5, No. 9* (pp. 316–318). New York: Wiley.
von Neumann, J. (1958). *The computer and the brain*. New Haven: Yale University Press.
von Neumann, J. (1966). *The theory of self-reproducing automata.* Edited and completed by A. Burks, Urbana, IL: University of Illinois Press, Fifth Lecture, pp.74–87.
Wächtershäuser, G. (1997). The origin of life and its methodological challenge. *Journal of Theoretical Biology, 187*, 483–494.
Wigner, E. (1964). Events, laws, and invariance principles. Wigner's Nobel Lecture, Stockholm, l0 Dec. 1963. Reprinted in *Science* 145, 995–999.
Yockey, H. P. (1974). An application of information theory to the Central Dogma and the sequence hypothesis. *Journal of Theoretical Biology, 46*, 369–406.
Yockey, H. P. (1992). *Information theory and molecular biology.* Cambridge: Cambridge University Press.

Yockey, H. P. (2000). Origin of life on earth and Shannon's theory of communication. *Computers and Chemistry, 24*, 105–123.

Yockey, H. P. (2005). *Information theory, evolution, and the origin of life*. Cambridge: Cambridge University Press.

Yockey, H. P., Platzman, R. L., & Quastler, H. (Eds.). (1958). *Symposium on information theory in biology.* New York and London: Pergamon.