Journal of Biosemiotics, 1 (2005) 81-101

Life is "artifact-making"

Marcello Barbieri
Dipartimento di Morfologia ed Embriologia
Via Fossato di Mortara 64, 44100 Ferrara, Italy
brr@unife.it

ABSTRACT – Biological information and the genetic code are key concepts in Modern Biology, and yet an influential school of thought maintains that ultimately they are but metaphors because they cannot be expressed in terms of physical quantities. Here it is argued that this *physicalist* thesis would be true if genes and proteins were made by spontaneous assemblies, like all inorganic molecules, but they are not. They are *manufactured* by molecular machines that physically stick their subunits together in the order provided by an external template. Genes and proteins, in other words, are molecular *artifacts* because they are made by external agents on the basis of external instructions. This in turn implies that all biological objects are artifacts, and the physicalist thesis collapses because there is a fundamental difference between *natural objects* and *artifacts*. Natural objects can be completely accounted for by physical quantities whereas artifacts require additional entities like sequences and codes, or equivalent entities like organic information and organic meaning. It is shown furthermore that organic information and organic meaning are *fundamental* entities of Nature which are brought into existence by the molecular processes of copying and coding. This implies that, far from being metaphors, they are as real as the processes that produce them, and in fact they can be defined by operative procedures that make them as objective and reproducible as the physical quantities. The idea that *life is artifact-making* has also implications for the origin and the evolution of living systems. Since copying and coding are two different ways of producing biological artifacts, there are two different mechanisms of evolutionary change: evolution by natural selection (based on copying) and evolution by natural conventions (based on coding). A consequence of this idea is that many organic codes should exist in Nature in addition to the genetic code, and the experimental evidence does support this breathtaking conclusion.

KEYWORDS – Macroevolution, information, meaning, organic codes, natural selection, natural conventions, common descent.

Introduction

The history of life has been shaped by great events that today are referred to as *Major Transitions* (Maynard Smith and Szathmary, 1995), or *Steps of Macroevolution*, or, more simply, *Origins*. They are the origins of: (1) genes, (2) proteins, (3) first cells, (4) eukaryotes, (5) embryos, (6) mind, and (7) culture. The key feature that defines these extraordinary events is the appearance of new biological objects, and this inevitably raises the question *"what is a biological object?"*, or, with a more familiar expression, *"What is life?"*

Another problem raised by the Major Transitions is the boundary between physics and biology. If life comes from inanimate matter and yet is different from it, what is it that *joins* and what is it that *divides* life from non-life? This is *the boundary problem,* and here again we realize that our conclusions depend upon the answer that we give to the question *"What is life?"*

In this paper it is proposed that *"life is artifact-making"*, i.e. that *"all biological objects are artifacts"*, and it is shown that this new concept does throw a new light on the above problems. In the case of the Major Transitions, it will be argued that there have been two distinct mechanisms of macroevolution because there are two distinct ways of producing biological artifacts. In the case of the boundary problem, it will be shown

that the relationship between physics and biology is equivalent to the one that exists between natural objects and artifacts. The concept that *life is artifact-making*, in short, make us realize that the Major Transitions, the boundary between physics and biology and the definition of life can be regarded not so much as three separate problems but rather as three aspects of a larger problem, thus giving us a feeling of what people used to call the *unity of Nature*.

PART 1 – BIOLOGY AND PHYSICS

A new definition of life

Genes and proteins differ from inorganic molecules not because they have different structures and functions but because they are *produced* in a totally different way. All inorganic molecules are made by self-assembly and their structure is determined *from within*, i.e., by *internal* factors. Genes and proteins, instead, are produced by molecular machines which physically stick their subunits together in an order provided *from without*, by *external* templates. They are assembled by molecular robots on the basis of outside instructions, and this makes them as different from ordinary molecules as *artificial* objects are from *natural* ones. Indeed, if we agree that objects are natural when their structure is determined from within and artificial when it is determined from without, then we can truly say that genes and proteins are *artificial molecules*, that they are *artifacts made by molecular machines*. This in turn implies that all biological objects are artifacts, and we arrive at the general conclusion that the whole of life is artifact-making.

We may find it difficult to accept this idea, but let us not forget that it is based on the most basic experimental properties of genes and proteins. It is the direct consequence of the most surprising discovery of Molecular Biology, the discovery that there is a totally unexpected gulf between life and non-life. The great divide is not between organic and inorganic structures. It is between structures which are built from within and structures which are built from without, between molecules which are made by spontaneous assemblies and molecules which are manufactured by molecular machines.

The definition of life as *artifact-making* is centered on a single feature and this appears to go against a long tradition according to which life cannot be reduced to any one property. In reality no such contrast exists, because the new definition does not account for *all* aspects of life. It singles out the one feature that divides inanimate matter from the living world, and in so doing it tells us why living systems are fundamentally different from non-living ones, but that does not give us a complete description of life. What it does give us, however, is a new beginning. More precisely, a new theoretical framework for the study of the origin and the evolution of life.

The First Major Transition

The discovery that genes and proteins are *manufactured* molecules has direct implications for the origin of life, because it tells us that primitive molecular machines came into existence long before the origin of the first cells. The simplest molecular machines we can think of are molecules that could join other molecules together by chemical bonds, and for this reason we may call them *bondmakers*. Some could form bonds between amino acids, some between nucleotides, others between sugars, and so on. It has been shown, for example, that short pieces of ribosomal RNA have the ability to form peptide bonds, so it is possible that the first bondmakers were RNA molecules of small or medium-size molecular weights. Among the various types of bondmakers, furthermore, some developed the ability to join nucleotides together in the order provided by a *template*. Those bondmakers, in other words, were *making copies* of nucleic acids, so we can call them *copymakers*. And the appearance of the first copymakers was a real turning point, because it set in motion an extraordinary sequence of events.

The copying of a template is the elementary act of gene duplication, the very first step toward the phenomenon of *heredity*. When a process of copying is repeated indefinitely, furthermore, another phenomenon comes into being. Copying mistakes become inevitable, and in a world of limited resources not all changes can be implemented, which means that a process of selection is bound to take place. Molecular copying, in short, leads to *heredity*, and the indefinite repetition of molecular copying leads to *natural*

selection. That is how natural selection came into existence. Molecular copying started it and molecular copying has perpetuated it ever since.

In the history of life, molecular copying came into being when the first copymakers appeared on the primitive Earth and started making copies of nucleic acids. This implies that *natural* nucleic acids had already been formed by spontaneous reactions on our planet, but that was no guarantee of evolution. Only the copying of genes could ensure their survival and have long-term effects, so it was really the arrival of copymaking that set in motion the extraordinary chain of processes that we call evolution. The first Major Transition is generally described as the origin of genes, but it does seem more accurate to say that it was the origin of molecular *copying*, or the origin of *copymakers*, the first molecular machines that started multiplying nucleic acids by making copies of them.

The Second Major Transition

The transition from natural to manufactured molecules was relatively simple for genes but much more complex for proteins, because genes can be copied while proteins cannot. Manufactured genes could be made simply by copying natural genes, and all that was required to that purpose were molecules which had a polymerase-like activity. Manufactured proteins, instead, could not be made by copying, and yet the information to make them had to come from molecules that can be copied, because only those molecules can be inherited. The information for manufacturing proteins, therefore, had to come from genes, so it was necessary to bring together a carrier of genetic information (a messenger RNA), a peptide-bondmaker (a piece of ribosomal RNA) and molecules that could carry both nucleotides and amino acids (the transfer RNAs). The first *protein makers*, in short, had to bring together three different types of molecules (messenger, ribosomal and transfer RNAs), and were therefore much more complex than copymakers.

The outstanding feature of the protein-makers, however, was not the number of components. It was the ability to ensure a one-to-one correspondence between genes and proteins, because without it there would be no biological specificity and no heredity. If the links between genes and proteins could have been determined by *stereochemistry*, as one of the earliest models suggested, a one-to-one correspondence would have been automatically ensured. Protein synthesis would have been but a form of *indirect copying*, and the problem of explaining its specificity would be relatively simple. The stereochemical hypothesis, however, turned out to be false. There simply is no chemical necessity in the links between codons and amino acids, and a one-to-one correspondence between them could only be the result of conventional rules. In short, only a genetic code could guarantee biological specificity, and this means that the evolution of the translation apparatus had to go hand in hand with the evolution of the genetic code. Protein synthesis arose from the integration of two different processes, and the final machine was a *"code-and-template-dependent-peptide-maker"*, or, more simply, a *codemaker*.

The second Major Transition of the history of life is generally described as the origin of proteins, but again it would be more accurate to say that it was the origin of *codemaking*, or the origin of *codemakers*, the first molecular machines that discovered molecular coding and started populating the Earth with codified proteins.

The physicalist thesis

Molecular biology is based on two key discoveries: one is that the linear sequence of nucleotides represents the *information* carried by a gene; the other is that the sequence of nucleotides in genes determines the sequence of amino acids in proteins. In both genes and proteins, therefore, *biological information* was identified with, and defined by, the specific sequence of their subunits.

This concept of information has solved the century old mystery of inheritance, and has turned the whole of biology from an energy-based into an information-based science. Despite all this, however, it has been pointed out that biological information is not a physical quantity, neither a fundamental nor a derived one, because a sequence cannot be measured. This is further underlined by the fact that there is another type of information which can be measured and which is a true physical quantity. It is the information of a message that Shannon defined in 1948 with an entropy-like formula, and which can be referred to as *physical information*.

The crucial point is that Shannon's information does not depend on the sequence of subunits, while biological information is defined precisely by that sequence. Physical information, in other words, has nothing to do with *specificity*, while biological information has everything to do with it. The two concepts are literally worlds apart, and this reinforces the conclusion that biological information is not, and cannot be, a physical quantity. So what is it?

According to an influential school of thought, biological information is a *metaphor*. It is a linguistic construction that we use in order to avoid long paraphrases when we talk about living systems, but no more than that. It is 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, 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; Griffiths and Knight, 1998; Griffith, 2001, Boniolo, 2003). It is probably one of the most deeply dividing issues of modern science. Many biologists are convinced that biological information is a real and fundamental component of life, but physicalists insist that it is real only in a very superficial sense and that there is nothing fundamental about it because it is not a physical quantity.

Organic information

Biological (or *organic*) information has been defined as *the specific sequence of a molecule*, but this is not entirely satisfactory because it gives the impression that information is a *property* of molecules, 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. This 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 during an act of copying, in other words, that information comes 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 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. There simply is no difference between saying that molecule B is a copy of molecule A, and saying that molecule B carries the same information as molecule A.

We realize in this way that organic information is as real as the copying process that generates it, but we still do not know if we can reduce it to simpler entities, and in particular to physical quantities. Luckily, this problem has a straightforward solution because the sequences of genes and proteins have two very special characteristics. One is that *a change to 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 it is impossible to predict the element (n+1)*. This is equivalent to saying that *a specific sequence cannot be described by anything simpler than the sequence itself*. Organic information, in short, cannot be reduced to anything else, and this makes of it an *irreducible* (or *fundamental*) entity of Nature. But what kind of entity is it? How does it fit into our schemes?

According to a long tradition, natural entities are divided into *quantities* and *qualities*. Quantities can be measured and are objective, while qualities are subjective and cannot be measured. In the case of organic information, however, this scheme breaks down. Organic information 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. 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 non-measurable)* we must recognize the existence in Nature of a third type of entities (*objective but not measurable*). Information is one of them, and we can also give it a suitable name. Since it can be described only by *naming* its sequence, we can say that organic

information is a *nominable* entity, or that it belongs to the class of the nominable entities of Nature (Barbieri, 2003b; 2004).

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 combinations of dots and dashes with the letters of the alphabet, and in the same way the genetic code is a correspondence between combinations 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 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 *word* "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 word. By the same token, the rule of the genetic code that 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 object which is related to another object by a code*, and that organic meaning exists wherever an organic code exists (Barbieri, 2003).

The existence of meaning in the organic world may seem strange, at first, but in reality it is not stranger than the existence of codes because they are the two sides of the same coin. Saying that a code establishes a correspondence between two objects is equivalent to saying that one object 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 and modern physics have readily accepted the concept of information but have carefully avoided the concept of meaning, and yet we have seen that 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 and coding are the processes; copymakers and codemakers are their agents; organic information and organic meaning are their results, or the kind of natural entity that they belong to. But the parallel goes even further than that. 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 general characteristics: they both are *objective-but-non-measurable* entities, they both are *irreducible* (or *fundamental*) entities of Nature, and since we can describe them only by naming their components, they both are *nominable* entities (Barbieri, 2003b; 2004). Finally, let us underline that they both are the 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 critical point was 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. Instead of 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 approach 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 object* is an objective and reproducible description of organic meaning.

While the physical quantities are evaluated *by measuring*, in other words, our biological entities 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. Alternatively, we can say that organic information and organic meaning are defined by the processes of copying and coding that bring them into existence, and this too amounts to an operative definition (Barbieri, 2003b; 2004).

We conclude that organic information and organic meaning can be defined by generalized operative definitions which are as reliable as the operative definitions of physics. This should ensure that they are no longer 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 if 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 concepts with the following statements:
(1) *Organic information is the specific sequence produced by a copying process*.
(2*) Organic meaning is the object which is related to another object by a code*.
(3) *Organic information and organic meaning are neither* quantities *nor* qualities. *They are a new kind of natural entities which are referred to as* nominable *entities.*
(4) *Organic information and organic meaning have the same scientific status as physical quantities because they are* objective *and* reproducible *entities which can be defined by operative procedures.*
(5) *Organic information and organic meaning have the same scientific status as* fundamental *physical quantities because they cannot be reduced to, or derived from, simpler entities*.

The problem of interpretation

Biologists have quickly accepted the idea of molecular information but not the idea of molecular meaning. Oddly enough, one of the most popular arguments against the existence of meaning at the *molecular* level has come from the study of animal communication, the very science that argued for the existence of language and meaning at the *animal* level. A signal that reaches an animal may set in motion an automatic response, and what we observe in these cases is the unfolding of a deterministic sequence of reactions. In other cases, however, there is an intermediate phase between signal and response, a phase where the signal is *interpreted* and the response changes accordingly. These are the cases where it is said that meaning appears, because interpretation is assumed to be the process that gives a meaning to a signal. Hence the conclusion that *meaning is always the result of an interpretation process*. No interpretation, no meaning.

In the case of protein synthesis, we know that codons are translated into amino acids according to the rules of the genetic code, but we also know that these reactions take place in a totally automatic way, with no room for interpretation. Hence the conclusion that there is no addition of meaning here, because protein synthesis is a deterministic chain of biochemical reactions. This argument is still popular and yet it is flawed because the determinism that we observe in protein synthesis is also observed in many cultural processes where we know that meaning does exist. The very act of speaking, for example, is based on the automatic application of prefixed rules. Language itself would not be possible if the meaning of its words had to be negotiated every time they are spoken. Once the basic rules have been fixed in the initial stage of learning, they are no longer changed and the code of a language becomes as deterministic as the genetic code.

The interpretation process that we observe in many animals can easily be understood as an evolution of their signal processing systems. It is likely that the most primitive reactions were heavily determined by genes, but the number of hard-wired responses could not increase indefinitely, and animals became more and more dependent on processes of learning in order to increase their behavioral repertoire. And learning how to respond to a signal means learning *how to interpret* that signal. Rather than memorizing an unlimited number of incoming stimuli, it was far more convenient to learn a few basic rules and let an interpretation phase decide the meaning of countless combinations of signals.

A process of learning, in other words, amounts to the construction of a behavioral code whose rules are *context-dependent*, and therefore *interpretation-dependent*. This gives the *impression* that the generation of

meaning comes from the interpretation process, but the truth is rather different. The interpretation process is necessary to make a choice among a plurality of rules, not to create the meaning of those rules. Meaning, in short, is always the result of a code, but sometimes the code is context-dependent, and in these cases it is associated with a process of interpretation. We conclude therefore that the definition of meaning as *an object which is related to another object by a code* is valid in all cases, with or without interpretation.

PART 2 – THE ORGANIC CODES

How many organic codes?

We have seen that codes exist in the organic world as well as in the mental world, and yet there seems to be a strange asymmetry between them. According to modern biology, there is only the genetic code in organic life, whereas the mental world is populated by a virtually unlimited number of cultural codes. We know, furthermore, that the genetic code came into being with the origin of life, while the cultural codes arrived almost four billion years later, with cultural evolution. This appears to suggest that organic evolution went on for four billion years, almost the entire history of life on Earth, without producing any other organic code after the first one, while cultural evolution produced an astonishing number of cultural codes in just a few thousand years. If this were true, there would indeed be an impressive asymmetry between the codes of the organic world and those of the mental world.

But is the asymmetry real? Are we sure that the genetic code is the only organic code of life? Luckily, this is a problem that we can deal with, because if other organic codes exist in Nature we should be able to find them by the classic experimental method of science, just as we have found the genetic code.

The very first step, in this enterprise, is to underline the difference that exists between *copying* and *coding*, a difference that is particularly evident in the seminal examples of those processes, i.e. in *transcription* and *translation*. In transcription, an RNA sequence is assembled from the linear information of a DNA sequence, and in this case a normal biological catalyst (an RNA polymerase) is sufficient, because each step requires a single recognition process. In translation, instead, two independent recognition processes must be performed at each step, and the system that performs the reactions (the ribosome) needs special molecules, first called *adaptors* and then *transfer RNAs*, in order to associate codons to amino acids according to the rules of the genetic code. Without a code, in fact, a codon could be associated with different amino acids and *biological specificity*, the most precious of life's properties, would be lost.

These concepts can easily be generalized. We are used to think that biochemical processes are all *catalyzed* reactions, but in reality we must distinguish very sharply between *catalyzed* and *codified* reactions. The catalyzed reactions are processes (like transcription) that require only one recognition process at each step. The codified reactions, instead, require two independent recognition processes at each step and a set of coding rules. The catalyzed reactions, in other words, require *catalysts*, while the codified reactions require *adaptors*, i.e. catalysts plus a code.

Any organic code is a set of rules that establish a correspondence between two independent worlds, and this necessarily requires molecular structures that act like *adaptors*, i.e. that perform two independent recognition processes. The adaptors are required because the two worlds would no longer be independent if there were a necessary link between them, and a set of rules is required in order to guarantee the specificity of the correspondence. The adaptors, in other words, are necessary in all organic codes. They are the molecular *fingerprints* of the codes, and their presence in a biological process is a sure sign that that process is based on a code. This gives us an *objective criterion* for the search of organic codes, and their existence in Nature becomes therefore, first and foremost, an experimental problem.

The splicing codes

One of the greatest surprises of molecular biology was the discovery that the primary transcripts of the genes are often transformed into messenger RNAs by removing some RNA strings (called *introns*) and by joining together the remaining pieces (the *exons*). The result is a true assembly, because exons are assembled into messengers, and we need therefore to find out if it is a *catalyzed* assembly (like transcription) or a *codified*

assembly (like translation). In the first case the cutting-and-sealing operations, collectively known as *splicing*, would require only a *catalyst* (comparable to RNA-polymerase), whereas in the second case they would need a catalyst and a set of *adaptors* (comparable to ribosome and tRNAs).

This suggests immediately that splicing is a codified process because it is implemented by structures that are very similar to those of protein synthesis. The splicing systems, known as *spliceosomes*, are huge molecular machines like ribosomes, and employ small molecular structures, known as *snRNAs* or *snurps*, which are very much comparable to tRNAs. The similarity, however, goes much deeper than that, because the snRNAs have properties that fully qualify them as *adaptors*. They bring together, in a single molecule, two independent recognition processes, one for the beginning and one for the end of each intron, thus creating a specific correspondence between the world of the primary transcripts and the world of messengers.

The two recognition steps are independent not only because there is a physical distance between them, but above all because the first step could be associated with different types of the second one, as demonstrated by the cases of *alternative splicing*. The choice of the beginning and of the end of an intron, furthermore, is the operation that actually defines the introns and gives them a *meaning*. Without a complete set of such operations, primary transcripts could be transformed arbitrarily into messenger RNAs, and there would be no biological specificity whatsoever.

In conclusion, in RNA splicing we find the three basic characteristics of codes:

(1) a correspondence between two independent worlds,

(2) the presence of molecular adaptors,

(3) a set of rules that guarantee biological specificity.

We conclude therefore that the processing of RNA transcripts into messengers is truly a codified process based on adaptors, and takes place with rules that can rightly be given the name of *splicing codes* (Barbieri, 1998; 2003).

The signal transduction codes

Cells react to a wide variety of physical and chemical stimuli from the environment and in general their reactions consist in the expression of specific genes. We need therefore to understand how the environment interacts with the genes, and the turning point, in this field, came from the discovery that the external signals (known as *first messengers*) never reach the genes. They are invariably transformed into a different world of internal signals (called *second messengers*) and only these, or their derivatives, reach the genes. In most cases, the molecules of the external signals do not even enter the cell and are captured by specific receptors of the cell membrane, but even those that do enter (some hormones) must interact with intracellular receptors in order to influence the genes (Sutherland, 1972).

The transfer of information from environment to genes takes place therefore in two distinct steps: one from first to second messengers, which is called *signal transduction*, and a second path from second messengers to genes which is known as *signal integration*. What surprises about signal transduction is that there are hundreds of first messengers (hormones, growth factors, neurotransmitters, etc.) whereas the known second messengers are only four (cyclic AMP, calcium ions, inositol trisphosphate and diacylglycerol) (Alberts *et al*., 1994).

First and second messengers, in other words, belong to two very different worlds, and this suggests immediately that signal transduction may be based on organic codes. This is reinforced by the discovery that there is no necessary connection between first and second messengers, because it has been proved that the same first messengers can activate different types of second messengers, and that different first messengers can act on the same type of second messengers.

The experimental data, in brief, prove that external signals do not have any instructive effect. Cells use them to *interpret* the world, not to yield to it. This conclusion amounts to saying that signal transduction is based on organic codes, which is in fact the only plausible explanation of the data, but of course we would also like a direct proof. As we have seen, the signature of an organic code is the presence of adaptors, and the molecules of signal transduction have indeed the typical characteristics of the adaptors. The transduction system consists of at least three types of molecules: a *receptor* for the first messengers, an *amplifier* for the second messengers and a *mediator* in between (Berridge, 1985). The system performs two independent recognition processes, one for the first and the other for the second messenger, and the two steps are connected by the bridge of the mediator. The connection however could be implemented in countless different ways since any first messenger can be coupled with any second messenger, and this makes it

imperative to have a code in order to guarantee biological specificity.

In signal transduction, in short, we find all three characteristics of codes:

(1) a correspondence between two independent worlds,

(2) a system of adaptors which give meanings to molecular structures,

(3) a collective set of rules that guarantee biological specificity.

The effects that external signals have on cells, in conclusion, do not depend on the energy or the information that they carry, but only on the *meaning* that cells give them with rules that we can rightly refer to as *signal transduction codes* (Barbieri, 1998; 2003).

The compartment codes

Eukaryotic cells not only produce molecules of countless different types but manage to deliver them to different destinations with astonishing precision, and this gives us the problem of understanding how they manage to cope with such an immensely intricate traffic. The first step in the solution of this mystery came with the discovery that the Golgi apparatus is involved not only in the biochemical modification of innumerable molecules but also in the choice of their geographical destination. But the truly remarkable thing is that all this is achieved with an extremely simple mechanism. More precisely, the Golgi apparatus delivers an astonishing number of molecules to their destinations with only three types of vesicles. One type has labels for the transport of proteins outside the cell and another for their delivery to the cell interior, whereas the vesicles of the third type carry no destination label, and are programmed, *by default*, to reach the plasma membrane. As we can see, the solution is extraordinarily efficient. With a single mechanism and only two types of labels, the cell delivers a great amount of proteins to their destinations, and also manages to continually renew its plasma membrane.

The Golgi apparatus, however, is a place of transit for only a fraction of the cell proteins. The synthesis of all eukaryotic proteins begins in the soluble part of the cytoplasm (the *cytosol*) together with that of a signal that specifies their geographical destination. The piece of the amino acid chain that emerges first from the ribosome (the so-called *peptide leader*) can contain a sequence that the cell interprets as an *export signal to the endoplasmic reticulum.* If such a signal is present, the ribosome binds itself to the reticulum and delivers the protein into its *lumen*. If not, the synthesis continues on free ribosomes, and the proteins are shed into the cytosol. Of these, however, only a fraction remains there, because the amino acid chain can carry, in its interior, one or more signals which specify other destinations, such as the *nucleus*, the *mitochondria*, and other cell compartments. Proteins, in conclusion, carry with them the signals of their geographical destination, and even the absence of such signals has a meaning, because it implies that the protein is destined to remain in the cytosol.

The crucial point is that there is *no necessary correspondence* between protein signals and geographical destinations. The export-to-the-nucleus signals, for example, could have been used for other compartments, or could have been totally different. They and all the other geographical signals are purely conventional labels, like the names that we give to streets, to cities, to airports and to holiday resorts. The existence of eukaryotic compartments, in other words, is based on natural conventions, and to their rules of correspondence we can legitimately give the name of *compartment codes* (Barbieri, 2003).

The cytoskeleton codes

A cytoskeleton is absolutely essential for typical eukaryotic processes such as phagocytosis, mitosis, meiosis, ameboid movement, organelle assembly and three-dimensional organization of the cell, i.e., for all those features that make eukaryotic cells so radically different from bacteria. The actual cytoskeleton, in reality, is an integrated system of three different cytoskeletons made of filaments (*microfilaments, microtubules* and *intermediate filaments*) each of which gives a distinct contribution to the three-dimensional form of the cell and to its mobility.

The driving force of the cytosleton is a very unusual mechanism that biologists have decided to call *dynamic instability*. The cytoskeletal filaments, especially microtubules and microfilaments, are in a state of continuous flux where monomers are added to one end and taken away at the other, and the filament is growing or shortening according to which end is having the fastest run. But what is really most surprising is that all this requires *energy*, which means that the cell is investing enormous amounts of energy not in

building a structure but *in making it unstable*.

In order to understand the logic of dynamic instability, we need to keep in mind that cytoskeletal filaments are unstable only when their ends are not attached to particular molecules that have the ability to anchor them. Every microtubule, for example, starts from an organizing center (the *centrosome*), and the extremity which is attached to this structure is perfectly stable, whereas the other extremity can grow longer or shorter, and becomes stable only when it encouters an anchoring molecule in the cytoplasm. If such an anchor is not found, the whole microtubule is rapidly dismantled and another is launched in another direction, thus allowing the cytoskeleton to explore all cytoplasm's space in a short time.

Dynamic instability, in other words, is a mechanism that allows the cytoskeleton to build structures with an *exploratory strategy*, and the power of this strategy can be evaluated by considering how many different forms it can give rise to. The answer is astonishing: the number of different structures that cytoskeletons can create is *potentially unlimited*. It is the anchoring molecules (that strangely enough biologists call *accessory proteins*) that ultimately determine the three-dimensional forms of the cells and the movements that they can perform, and there could be endless varieties of anchoring molecules.The best proof of this enormous versatility is the fact that the cytoskeleton was invented by unicellular eukaryotes but was later exploited by metazoa to build completely new structures such as the axons of neurons, the myofibrils of muscles, the mobile mouths of macrophages, the tentacles of killer lymphocytes and countless other specializations.

In conclusion, dynamic instability is a means of creating an endless stream of cell types with only one common structure and with the choice of a few anchoring molecules. But this is possible only because there is *no necessary relationship* between the common structure of the cytoskeleton and the cellular structures that the cytoskeleton is working on. The anchoring molecules (or accessory proteins) are true *adaptors* that perform two independent recognition processes: microtubules on one side and different cellular structures on the other side. The resulting correspondence is based therefore on *arbitrary* rules, on true natural conventions that we can refer to as *the cytoskeleton codes* (Barbieri, 2003).

The sequence codes (Edward Trifonov)

In the 1980s and 90s, Edward Trifonov started a long campaign in favour of the idea that the nucleotide sequences of the genomes carry several messages simultaneously, and not just the message revealed by the classic triplet code. According to Trifonov, in other words, the genetic code is not alone since there are many other codes in the nucleotide sequences of living organisms. This conclusion rests upon Trifonov's definition that *"a code is any sequence pattern that can have a biological function"* or *"codes are messages carried by sequences"* or *"a code is any pattern in a sequence which corresponds to one or another specific biological function"* (Trifonov, 1989; 1996; 1999).

The plurality of codes described by Trifonov is a result of his particular definition of a code, but it is not necessarily limited by that, and could well be compatible with different definitions. The splicing code, for example, is a code not only according to his criterion, but also according to the operative definition that a code is a set of rules of correspondence implemented by adaptors. This suggests that Trifonov's conclusions may have a general validity, and at least some of his sequence codes may turn out to be true organic codes. For the time being, however, let us acknowledge the fact that according to Trifonov's definition there are at least eight sequence codes in the genomes of living creatures, in addition to the classic triplet code (Trifonov, 1996):

(1) The *transcription codes* include promoters and terminators, and are rather universal, though different in prokaryotes and in eukaryotes.

(2) The *gene splicing code* for the processing of nuclear pre-mRNA is largely undeciphered. Its main components are obligatory GU- and AG-ends of introns, as well as rather conserved consensus sequence features around the ends.

(3) The *translation pausing code*, for the regulation of translation, is encoded by clusters of rare triplets for which the aminoacyl-tRNAs are in limited supply.

(4) The *DNA structure code* , or *DNA shape code*, is a sequence-dependent local shape of DNA which is a crucial component of the protein-DNA recognition.

(5) The *chromatin code* describes those sequence features that direct the histone octamer's binding to DNA and the formation of nucleosomes.

(6) The *translation framing code* is overlapping with the triplet code (Trifonov, 1987), and ensures the correct reading frame during translation.

(7) The *modulation code* is about the repeating sequences and regulates the number of repeats as an adjustable variable to modulate expression of the nearby gene.

(8) The *genome segmentation code* is one of the emerging new codes, and is due to fact that the genomes appear to be built of rather standard size units.

A stream of codes

Most of the papers which have been published on biological codes do not make any reference to their definition of code. This is the case, for example, of the reports which have described and discussed a truly remarkable wealth of experimental data on the *Adhesive Code* (Redies and Takeichi, 1996; Shapiro and Colman, 1999), on the *Sugar Code* (Gabius, 2000; Gabius et at., 2002), and on the *Histone Code* (Strahl and Allis, 2000; Jenuwein and Allis, 2001; Turner, 2000; 2002; Gamble and Freedman, 2002; Richards and Elgin, 2002).

The practice of studying something without precisely defining it is fairly common in many sciences, and biology is no exception. The paradigmatic example is life itself, a phenomenon that we keep studying even if nobody seems to agree on its definition. Another instructive case is the concept of species, for which there is no definition that is universally valid and yet this does not prevent biologists from doing experiments, obtaining results and making sensible predictions on countless species of living creatures. Precise definitions, in short, are not always essential, but in some cases they are, and this is one of them. More precisely, we should be aware that an operative definition of organic codes in terms of adaptors would have provided a crucial guideline in at least two important cases.

(1) One is the research on new biological codes such as the Adhesive Code, the Sugar Code and the Histone Code. The problem here is that the experimental data *suggest* the existence of organic codes but do not *prove* it. And yet the results could have been conclusive because they are all compatible with the existence of true adaptors. On the face of the evidence, for example, it is most likely that lectins are the adaptors of the Sugar Code and that cadherins are the adaptors of the Adhesive Code. If that had been proved, there would be no doubt that we are in the presence of true organic codes. But people did not use a definition of codes based on adaptors, so they did not look for adaptors.

(2) The second case is that of the classical research on signal transduction. Here the amount of experimental data is so enormous to be beyond description, and yet there is a remarkable paradox in this field. The only logical explanation of the facts is that signal transduction is based on organic codes and yet the word "code" has never been mentioned, so people have never looked for coding rules. The evidence has actually proved that signal transducers have the experimental characteristics of true adaptors, and yet the word "adaptors" has never been mentioned with reference to a code. This habit could well go on indefinitely by inertia, and only a precise definition of organic codes can convince people that an alternative (and much more convincing) explanation of the facts already exists.

It has been the existence of adaptors which has proved the reality of the Genetic Code, and the same is going to be true for the Signal Transduction Codes, for the Adhesive Code, for the Sugar Code and for the Histone Code. An operative definition based on adaptors, furthermore, is the only scientific instrument that can allow us to prove the existence of other organic codes in Nature. And when we really start looking for them, we may well discover that so far we have only scratched the surface. That there is a long golden stream of organic codes out there.

PART 3 – THE SEMANTIC LOGIC OF LIFE

Artifact-making and Natural Selection

The first Major Transition gave origin to genes by the mechanism of *copying*, while the second Major Transition gave origin to proteins by the mechanism of *coding*. At the very beginning of the history of life we find two different mechanisms, copying and coding, but what about the rest of the history? Which mechanisms do we find in the other Major Transitions?

It is no secret that today most biologists regard natural selection as the sole mechanism of evolution, but

that is hardly surprising because natural selection is based on molecular copying. More precisely, on the indefinite repetition of a process of molecular copying in a world of limited resources. In these circumstances, copying mistakes are bound to happen and that inevitably leads to selection. It is molecular copying, in other words, that produces natural selection, and this means that *natural selection would be the sole mechanism of evolution if variations in molecular copying were the sole mechanism of biological change*.

As a matter of fact, this *could* have happened. If living systems were entirely made of RNA enzymes and RNA genes, only the copying of RNA molecules would be necessary, and natural selection would indeed be the sole mechanism of evolution. But that is not what happened. Long before the origin of the first cells, proteins were being made on the primitive Earth, and proteins, unlike genes, could not be made by copying. The manufacture of proteins required codemakers, not copymakers. It required two independent recognition processes, not one, and above all it required the rules of a code. In an RNA world, in short, molecular copying – and therefore natural selection – *could* have been enough, but in a world where proteins exist there must necessarily be natural conventions, and these cannot be reduced to natural selection because coding cannot be reduced to copying.

There is however another scenario where we could say that natural selection has *virtually* been the sole mechanism of evolution. If no other organic code had appeared on Earth after the genetic code, we would have to conclude that copying has been the sole mechanism of molecular change *for almost four billion years*, and natural selection could legitimately be regarded as the sole mechanism of evolution for almost the entire history of life. In this case, the origin of the genetic code at the beginning and the origin of the cultural codes at the end of the history of life could be regarded as two extraordinary exceptions, and natural selection would remain *in practice* the sole mechanism of evolutionary change.

But the genetic code is not the only code of life. There are many other organic codes in Nature, and this means that they had origins and histories, that they came into being during the course of evolution. This in turn means that copying and coding operated throughout the whole history of life, and gave origin to very different types of changes. The idea that life is artifact-making and that copying and coding are two different ways of producing biological artifacts makes us realize that there have been two distinct mechanisms in the history of life. Evolution was not produced only by natural selection but *by natural selection and by natural conventions* (Barbieri, 1985) - which in no way is a denial or a belittlement of natural selection. It is only an extension of it.

Artifact-making and Common Descent

It has been said (and it is probably true) that Darwin's greatest idea was not the principle of natural selection but the theory of common descent, the idea that *"all the organic beings which have ever lived on this Earth may be descended from some one primordial form"* (Darwin, 1859). In fact, when Theodosius Dobzhansky (1973) wrote that *"Nothing in biology makes sense except in the light of evolution"*, it was common descent that he had in mind. The idea that all creatures of the present are linked to all creatures of the past, is indeed the greatest unifying theme in Biology, the concept that we use as an Ariadne's thread to reconstruct the history of life.

Common descent, however, is not a single theory. The process of *evolution from a common ancestor* is compatible with different mechanisms, and these give us different versions of common descent because they give us different reconstructions of the tree of life. In order to find out the truth about common descent, therefore, we need to know the actual mechanisms that gave origin to biological objects in the course of time. How did novelties appear on Earth? Did new objects arise only by the gradual modification of previous objects, i.e. *by natural selection*, or also by the origin of new organic codes, i.e. *by natural conventions*? Natural selection produces novelties by transforming existing systems into slightly different ones, whereas natural conventions bring *absolute novelties* into existence. And this gives us two very different theories of Common Descent.

If evolution had taken place only by natural selection, we would have to conclude that nothing similar to the origin of the genetic code could have happened again in the four billion years of life's history. But if many other organic codes have come into being, we would have to conclude that there have been many other *origins* in the history of life because any new organic code gives origin to unprecedented structures. Evolution by natural selection, in short, implies *Common Descent with a Single Origin*, whereas evolution by

natural selection and by natural conventions leads to *Common Descent with Multiple Origins*. (This is not the old theory that *cells* originated many times, because the multiple origins are referred to *codes* not to cells).

The idea that natural conventions bring absolute novelties into existence can perhaps be illustrated by the case of those particular mental objects that we call *numbers*. There is little doubt that numbers originated from the practice of counting, but there is also little doubt that the endless properties of the world of numbers were not all produced by the primitive act of counting. Numbers are absolute novelties because their properties did not exist before, did not arise by emergence and were not all produced by the original process of counting that brought numbers into existence. The properties of numbers are the result of mental conventions that have the extraordinary ability to produce other mental conventions in a sequence that apparently has no end.

It is true that numbers are not organic objects, but it is also true that mental objects and organic objects are both artifacts produced by codemakers, and both of them were absolute novelties that gave origin to entirely new worlds. The origin of mind, in other words, was not less of a novelty than the origin of proteins or the origin of the first cells (and in all these cases, absolute novelties does not imply *sudden* novelties). The theory of Common Descent with Multiple Origins makes us realize that absolute novelties appeared not only at the beginning, but throughout the entire history of life. And this is not a denial or a belittlement of the traditional theory of common descent form a single origin. It is only an extension of it.

Artifact-making and the Origin of Life

The rediscovery of Mendel's laws in 1900 suggested that heredity is governed by rules of its own, quite different from those of metabolism, and in 1909 Wilhelm Johannsen raised this difference to the highest theoretical level of a fundamental dichotomy of Nature. He proposed that every living creature is made of two distinct categories: a hereditary part that he called *genotype* and a physiological, or visible, part that he named *phenotype*. This was indeed the logical implication of Mendel's laws, but Johannsen's idea was rejected by most biologists because it was widely believed that heredity and metabolism were both accounted for by proteins. It took nearly 50 years to reverse that conclusion, and the change came in two stages. First with the discovery that heredity is based on nucleic acids, not on proteins, and then with the discovery that heredity is carried by linear sequences whereas metabolism is produced by three-dimensional structures. But perhaps the decisive factor was the appearance and the diffusion of the computer because the duality of software and hardware made the meaning of genotype and phenotype immediately clear. This is probably why biologists accepted so quickly a dualistic model of life that previously had been stubbornly rejected for decades.

One of the consequences of the genotype-phenotype duality is that even the most primitive cells had to be made of genes and proteins. Those molecules, however, could have produced an integrated system only after a long evolutionary process, and this implied that life on Earth started either with genes or with proteins. The origin-of-life theories were divided therefore into *heredity-first* and *metabolism-first* models, a scheme that had a strong intuitive appeal even because it was echoing the popular metaphor of *the-chicken-and-the egg*. Genes and proteins, genotype and phenotype, software and hardware, heredity and metabolism, nature and nurture: those were thought to be the fundamental categories of the cell and of life itself.

But genes and proteins are manufactured molecules, so there had to be molecular machines to bring them into existence. There had to be copymakers and codemakers, and these were a category of their own, the category of the producers, of the agents, of the makers. And since they were mainly made of ribonucleic acids, their category has been called *ribotype*, the system that looks like an intermediary between genotype and phenotype but which is, in reality, their ancestral maker. The cell, in other words, is not a duality, but a trinity of genotype, phenotype and ribotype (Barbieri, 1981). The cell, if we must use a metaphor, is like a city, where the proteins are the buildings, the genes are their projects, and the ribotype is the inhabitants. In this framework it doesn't even make sense to ask whether it was the buildings or the projects that came first, because we already know that it was the inhabitants – the makers – that had to be there before everything else.

We realize in this way that our theories on the Origin of Life depend upon the answer that we give to the question "What is Life?". The most basic evidence of molecular biology tells us that genes and proteins are manufactured molecules, and this means that *life is artifact –making*. It means that the cell is like a city, not like a computer. That the true model of the cell is not the duality of genotype and phenotype but the trinity

genotype, phenotype and ribotype. That life was born when the first molecular machines appeared on Earth and started populating it with molecular artifacts.

Schrodinger's prophecy

In 1944, Erwin Schrodinger 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 Schrodinger 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 Schrodinger compared to *a Morse code with many characters*, and that was supposed to carry *the highly complicated plan of development of the entire organism*. That was the very first time that the word "code" was associated with a biological structure and was given a biological function.

The existence of specificity and codes at the heart of life led Schrodinger 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"*. Schrodinger 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, Gunther Stent (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"*.

Schrodinger's prophecy of new laws of physics seems to have been shipwrecked in a sea of hydrogen bonds, but in reality that is true only in a very superficial sense. The essence of the prophecy was about the existence of *hitherto unknown* fundamental entities, and that turned out to be true. As we have seen, life is based on organic information and organic meaning, and these are indeed new fundamental entities of Nature. Schrodinger invoked the existence of new *laws* rather than of new *entities*, but that was only a minor imperfection, and should not have been allowed to obscure the substance of the prophecy.

There is however one thing that Schrodinger might not have approved in the answer that here has been given to the question *"What is Life?"*. Together with many other physicists, he believed that scientific truths must have *beauty*, and the answer *"Life is artifact-making"* might not be elegant enough to meet his criterion of truth. Luckily, there is a simple way out of this impasse, because the word *artifact-making* maintains its meaning even when we drop all its letters but the first three. In this way, the idea that *"Life is artifact-making"* would become *"Life is art"*, and this is a conclusion that even Schrodinger might have appreciated.

REFERENCES

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., & Watson, J.D. (1994). *Molecular Biology of the Cell.3rd edition.* Garland, New York.

Barbieri M. (1981). The Ribotype Theory on the Origin of Life. *Journal of Theoretical Biology*, 91, 545-601.

Barbieri M. (1985). *The Semantic Theory of Evolution*. Harwood, London and New York.

Barbieri M. (1998). The Organic Codes. The basic mechanism of macroevolution. *Rivista di Biologia-Biology Forum*, 91, 481-514.

Barbieri M. (2003). *The Organic Codes. An Introduction to Semantic Biology*. Cambridge University Press, Cambridge, UK.

Barbieri M. (2003b). Biology with Information and Meaning. *History and Philosophy of the Life Sciences*, 25, 243-254.

Barbieri M. (2004). The Definitions of Information and Meaning. Two possible boundaries between physics and biology. *Rivista di Biologia-Biology Forum*, 97, 91-110.

Berridge, M. (1985). The molecular basis of communication within the cell. *Scientific American*, 253, 142-152.

Boniolo G. (2003). Biology without Information. *History and Philosophy of the Life Sciences*, 25, 255-273.

Chargaff, E. (1963). *Essays on Nucleic Acids*. Elsevier, Amsterdam.
Darwin C. (1859). *On the Origin of Species by Means of Natural Selection*. Murray, London.
Dobzhansky T. (1973). Nothing in biology makes sense except in the light of evolution. *American Biology Teacher*, 35, 125-129.
Gabius H.-J. (2000). Biological Information Transfer Beyond the Genetic Code: The Sugar Code. *Naturwissenschaften*, 87, 108-121.
Gabius H.-J, André S., Kaltner H., and Siebert H.-C. (2002). The sugar code: functional lectinomics. *Biochimica et Biophysica Acta*, 1572, 165-177.
Gamble M.J. and Freedman L.P. (2002). A coactivator code for transcription. *Trends in Biochemical Sciences*, 27 (4), 165-167.
Griffith, P.E. (2001). Genetic Information: A Metaphor in Search of a Theory. *Philosophy of Science*, 68, 394-412.
Griffith, P.E. and Knight, R.D. (1998). What is the developmental challenge? *Philosophy of Science*, 65, 276-288.
Jenuwein T. and Allis D. (2001). Translating the Histone Code. *Science*, 293, 1074-1080.
Johannsen, W. (1909). *Elemente der exacten Erblichkeitslehre.* Gustav Fisher, Jena.
Mahner, M. and Bunge, M. (1997). *Foundations of Biophilosophy*. Springer Verlag, Berlin.
Maynard Smith, J. & Szathmary E. 1995. *The Major Transitions in Evolution.* Oxford University Press, Oxford.
Readies C. and Takeichi M. (1996). Cadherine in the developing central nervous system: an adhesive code for segmental and functional subdivisions. *Developmental Biology*, 180, 413-423.
Richards E.J. and Elgin S.C.R. (2002). Epigenetic Codes for Heterochromatin Formation and Silencing: Rounding up the Usual Suspects. *Cell*. 108, 489-500.
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*. Kluwer Academic Publishers, Dordrecht, 187-231.
Sarkar, S. (2000). Information in Genetics and Developmental Biology. *Philosophy of Science*, 67, 208-213.
Schrodinger E. (1944). *What is Life?* Cambridge University Press, Cambridge, UK.
Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, 27, 379-424, 623-656.
Shapiro L. and Colman D.R. (1999). The Diversity of Cadherins and Implications for a Synaptic Adhesive Code in the CNS. *Neuron*, 23, 427-430.
Stent G.S. and Calendar R. (1978). *Molecular Genetics*. W.H. Freeman, San Francisco.
Strahl B.D. and Allis D. (2000). The language of covalent histone modifications. *Nature*, 403, 41-45.
Sutherland, E.W. (1972). Studies on the mechanism of hormone action. *Science*, 177,401-408.
Trifonov E.N. (1987). Translation framing code and frame-monitoring mechanism as suggested by the analysis of mRNA and 16s rRNA nucleotide sequence. *Journal of Molecular Biology*, 194, 643-652.
Trifonov E.N. (1989). The multiple codes of nucleotide sequences. *Bulletin of Mathematical Biology*, 51, 417-432.
Trifonov E.N. (1996). Interfering contexts of regulatory sequence elements. *Cabios*, 12, 423-429.
Trifonov E.N. (1999). Elucidating Sequence Codes: Three Codes for Evolution. *Annals of the New York Academy of Sciences*, 870, 330-338.
Turner, B.M. (2000). Histone acetylation and an epigenetic code. *BioEssay*, 22, 836-845.
Turner, B.M. (2002). Cellular memory and the Histone Code. *Cell*, 111, 285-291.