Biosemiotics (2014) 7:167–179
DOI 10.1007/s12304-014-9212-1

EDITORIAL

Introduction to Code Biology

Marcello Barbieri

Received: 11 May 2013 / Accepted: 21 June 2013 / Published online: 21 August 2014

The New World of the Organic Codes

The genetic code appeared on Earth at the origin of life, and the codes of culture arrived almost 4 billion years later, at the end of life's history. Today it is widely assumed that these are the only codes that exist in Nature, and if this were true we would have to conclude that codes are *extraordinary exceptions* because they appeared only at the beginning and at the end of evolution. In reality, various other organic codes (codes between organic molecules) have been discovered in the past few decades.

In 1975, the American biochemist Gordon Tomkins published a paper entitled *The Metabolic Code. Biological symbolism and the origin of intercellular communication* (Tomkins 1975). That was the very first announcement of a new organic code after the discovery of the genetic code, but tragically Tomkins died that very year and his new world of organic symbolism remained unexplored. Some 10 years later, Edward Trifonov started a life-long campaign in favour of the idea that genomes carry several overlapping codes simultaneously, not just the classic triplet code, and gave them the collective name of *sequence codes* (Trifonov 1987, 1989, 1999).

Finally, at the end of the 1990s and in the early 2000s, a whole set of new organic codes came to light. Among them: the *adhesive code* (Readies and Takeichi 1996; Shapiro and Colman 1999), the *splicing codes* (Barbieri 1998, 2003; Pertea et al. 2007; Barash et al. 2010; Dhir et al. 2010), the *signal transduction codes* (Barbieri 1998, 2003, 2008), the *sugar code* (Gabius 2000, 2009), the *histone code* (Strahl and Allis 2000; Turner 2000, 2002, 2007), the *cytoskeleton codes* and the *compartment codes* (Barbieri 2003, 2008), the *neural code* (Nicolelis and Ribeiro 2006; Osborne et al. (2008), the *tubulin code* (Verhey and Gaertig 2007), the *nuclear signalling code* (Maraldi 2008), and the *ubiquitin code* (Komander and Rape 2012).

It must be pointed out that various authors have defined codes in different ways, a problem that is not uncommon in biology, but in our case a solution does exist because there is an operative definition that can be applied to all organic codes. This paper will start with that definition and will use it to illustrate three outstanding examples of organic codes. After that, it will be shown that the experimental data allow us to

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

recognize the existence of organic codes in many other cases, and we will briefly examine four of them: the origin of the first cells, the origin of animals, the origin of mind and the origin of language.

An Operative Definition

An operative definition is one that allows us to make experimental tests that prove whether or not organic codes do exist in Nature. The starting point is the idea that a code is always *a set of rules that establish a correspondence between two independent worlds* (Barbieri 2003).

The Morse code, for example, is a correspondence (or a *mapping*) between the letters of the alphabet and groups of dots and dashes. The highway code is a correspondence between street signals and driving behaviours (a red light means 'stop', a green light means 'go', and so on). What is essential in all codes is that the coding rules are not dictated by the laws of physics and chemistry. In this sense they are *arbitrary*, and the number of arbitrary relationships between two independent worlds is potentially unlimited. In the Morse code, for example, any letter of the alphabet can be associated with countless combinations of dots and dashes, which means that a specific link between them can be realized only by selecting a small number of rules. And this is precisely what a code is: *a small set of arbitrary rules selected from a potentially unlimited number in order to ensure a specific correspondence between two independent worlds*.

In biology, organic codes are relationships between two worlds of organic molecules and are necessarily implemented by other molecules, called *adaptors*, that build a bridge between them. The adaptors are required because there is no necessary link between the two worlds, and a fixed set of adaptors is required in order to guarantee the specificity of the correspondence (Barbieri 2003). The adaptors, in short, are essential 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 discovery of the organic codes, and their existence in Nature is no longer a matter of speculation. It is, first and foremost, an experimental problem. More precisely, we can prove that an organic code exists, if we prove the existence of three entities: (1) two independents worlds of molecules, (2) a potentially unlimited number of arbitrary connections between them implemented by adaptors, and (3) a selection of the adaptors (a set of coding rules) that ensures a specific correspondence.

Three Outstanding Examples

The Genetic Code

In protein synthesis, a sequence of nucleotides is translated into a sequence of amino acids, but it has been shown that there is no necessary link between nucleotides and amino acids. These molecules belong to two independent worlds, and a bridge between them is realized by a third type of molecules, called *transfer-RNAs*, that act as adaptors and perform two distinct operations: at one site they recognize groups of three nucleotides, called *codons*, and at another site they receive amino acids by enzymes called *aminoacyl-synthetases*. The key point is that a binding between synthetases and transfer-RNAs can be realized in countless different ways, and this means that in principle any amino acid can be associated with any codon. The number of connections, in other words, is potentially unlimited, and only the selection of a small fixed set of adaptors can ensure a specific mapping. This is *the genetic code*: a fixed set of rules of correspondence between codons and amino acids that are implemented by adaptors. In protein synthesis, in conclusion, we find all the three essential components of a code: (1) two independents worlds of objects (nucleotides and amino acids), (2) a potentially unlimited number of arbitrary connections created by adaptors, and (3) a set of coding rules (a selection of the adaptors) that ensures the specificity of the correspondence.

The Signal Transduction Codes

Signal transduction is the process by which cells transform the signals from the environment, called *first messengers*, into internal signals, called *second messengers*. First and second messengers belong to two independent worlds because there are literally hundreds of first messengers (hormones, growth factors, neurotransmitters, etc.) but only four great families of second messengers (cyclic AMP, calcium ions, diacylglycerol and inositol trisphosphate) (Alberts et al. 2007). The crucial point is that the molecules that perform signal transduction are true adaptors. They consists of three subunits: a *receptor* for the first messengers, an *amplifier* for the second messengers, and a *mediator* in between (Berridge 1985). This allows the transduction complex to perform two independent recognition processes, one for the first messenger and the other for the second messenger. Laboratory experiments have proved that any first messenger can be associated with any second messenger (Alberts et al. 2007), which means that there is a potentially unlimited number of arbitrary connections between them. In signal transduction, in short, we find all the three essential components of a code: (1) two independents worlds of objects (first messengers and second messengers), (2) a potentially unlimited number of arbitrary connections produced by adaptors, and (3) a set of coding rules (a selection of the adaptors) that ensures the specificity of the correspondence.

The Splicing Codes

All DNA genes are copied into RNA molecules that are called *primary transcripts*, and all proteins are synthesized from RNA molecules that are called *messenger RNAs*. In bacteria, the primary transcripts are directly used as messenger RNAs, but in virtually all other creatures things are much more complicated. The primary transcripts are first cut into pieces and then some of them (called *introns*) are removed and the remaining pieces (called *exons*) are joined together to form the messenger RNAs. These cutting-and-sealing operations are collectively known as *splicing*, and we want to find out whether or not splicing is based, like protein synthesis, on a code. In this case, the existence of a code looks immediately likely because there are clear parallels between splicing and protein synthesis: (a) the splicing catalysts, known as *spliceosomes*, are huge molecular machines like ribosomes, (b) splicing employs small molecules, known as *snRNA*, that are comparable to transfer-RNAs, and (c) both are processes that assemble molecules: splicing assembles messenger-RNAs from exons whereas protein synthesis assembles proteins from amino acids. The crucial point is that the choice of the beginning of an intron is completely independent from the choice of its end, and this proves that the snRNAs are real adaptors because they perform two independent recognition processes at each step of the reaction. This conclusion is further reinforced by the fact that in some cases there is not just one, but various sets of splicing rules that are employed in different occasions, for example at different stages of embryonic development (*alternative splicing*). In splicing, in other words, we find all the three essential components of a code: (1) two independents worlds of objects (primary transcripts and messenger RNAs), (2) a potentially unlimited number of arbitrary connections produced by adaptors, and (3) a set of coding rules (a selection of the adaptors) that ensures the specificity of the correspondence.

Theoretical Implications

The genetic code, the signal transduction codes and the splicing codes are the most outstanding examples of organic codes, but many other cases have been described in the scientific literature, and it is likely that more will come to light in the future. The existence of many organic codes in Nature is therefore an *experimental* fact – let us never forget this – but also more than that. It is one of those facts that have extraordinary theoretical implications.

The first is the role that the organic codes had in the history of life. The genetic code was a precondition for the origin of the first cells, the signal transduction codes divided the first cells into three primary kingdoms (*Archaea*, *Bacteria* and *Eukarya*), the splicing codes were instrumental in the origin of the nucleus, the histone code provided new working rules of chromatin, and the cytoskeleton codes allowed the *Eukarya* to perform internal movements, including those of mitosis and meiosis (Barbieri 2003). It will be shown, furthermore, that organic codes had a key role in the major transitions of multicellar life, in particular in the origin of animals, the origin of mind and the origin of language. The great events of macroevolution, in short, were associated with the appearance of new organic codes, and this gives us a new description and a completely new understanding of the history of life.

The second theoretical implication of the organic codes comes from the fact that codes involve *meaning* and we need therefore to introduce in biology, with the standard methods of science, not only the concept of information but also the concept of meaning.

The third theoretical implication comes from the fact that the organic codes have been highly conserved in evolution. Before the origin of the genetic code, the ancestral precursor was engaged in evolving coding rules and was therefore a *code generating system*. After the origin of the code however no other modification in coding rules was accepted and the cell became a *code conservation system*. Another part of the ancestral cells, however, maintained the potential to evolve other coding rules and behaved as new *code generating*, or *code exploring, systems*. In the early *Eukarya*, for example, the cells had *a code conservation part* for the genetic code, but also a *code exploring part* for the splicing code. This tells us something important about life. The origin of the first cells was based on the ability of the ancestral systems to *generate* the rules of the genetic code, and the subsequent evolution of the cells was based on two complementary processes: one was the *generation* of new organic codes and the other was the *conservation* of the existing ones. Taken together, these two processes are referred to as *codepoiesis*, a phenomenon that accounts for the two most important events that took place in evolution. The ability to create coding rules accounts for the origin of the genetic code and of all the other codes that followed. The ability of the cell to conserve its own codes accounts for the fact that the organic codes are the great invariants of life, the entities that must be conserved while everything else is changing.

Origin and Evolution of the Cells

The data from molecular biology have revealed that all known cells belong to three distinct primary kingdoms, or domains, that Carl Woese called *Archaea*, *Bacteria* and *Eucarya* (Woese 1987, 2000). The fact that virtually all cells have the same genetic code means that this code appeared in primitive systems that are collectively known as the *common ancestor*. But how did the common ancestor give origin to the cells of the three primary kingdoms? A good clue comes from the fact that all cells have a context-dependent behaviour because they regulate protein synthesis according to the signals that come from the environment (Jacob and Monod 1961). This means that a signal transduction code was of paramount importance to the ancestral systems, which makes it very likely that they made various attempts to develop it.

It is an experimental fact, at any rate, that *Archaea*, *Bacteria* and *Eucarya* have three different types of membranes and three distinct signalling systems, and this suggests that the three domains came into being by combining the universal genetic code with three distinct signal-transduction codes. This amounts to saying that the genetic code was instrumental to the origin of the common ancestor and that the signal transduction codes were instrumental to the origin of the first cells.

In order to understand the evolution of the first cells we need to keep in mind that bacteria appeared very early on our planet and some of them have remained substantially the same ever since. This is dramatically illustrated by the fact that modern stromatolites built by cyanobacteria are virtually identical to the 3.4 and to the 1.8 billion year old stromatolites that have been found in the fossil record (Schopf 1999; Knoll 2003). Primitive bacteria, in other words, already had the main characteristics of their modern descendants, and this tells us something important about the early history of life. It tells us that the descendants of the common ancestor had two evolutionary strategies in front of them, one based on increasing simplification, or *streamlining*, and one based on increasing complexity.

The cells that adopted a streamlining strategy got rid of all unnecessary components, lost the ability to evolve new organic codes and have remained substantially the same ever since. Other cells conserved their primitive features, including the potential to evolve new organic codes, and have became increasingly complex. This tells us that codes lie at the very heart of the evolutionary mechanism. The cells that did not evolve new organic codes became *bacteria* and have never changed their fundamental structure. The cells that evolved new codes, such as splicing codes, cytoskeleton codes, compartment codes, histone code and so on, became *eukarya* and have generated increasingly complex cellular organizations.

We realize in this way that there is a close association between the great events of macroevolution and the appearance of new organic codes, and we can also understand why. It is because a new code brings into existence an absolute novelty, something that has never existed before, because the adaptors of a code create associations that are not determined by physical necessity. Any new code, in conclusion, creates a genuine increase in complexity, to the point that the best measure of the complexity of a living system is probably the number of its codes.

The Origin of Animals

In the evolution of embryonic development, Nature made three main experiments, respectively with one, two and three types of cells called *germ layers*. One germ layer makes bodies which have no symmetry (the sponges); two germ layers build bodies with one axis of symmetry (the diploblasts or *radiata*, i.e., hydra, corals and medusae), and three germ layers produce bodies with three axes of symmetry (bodies which have a left and a right, a top and a bottom, a back and a front). These are the triploblasts or *bilateria*, and comprise the greatest majority of animals, vertebrates and invertebrates (Tudge 2000).

In principle, the number of three-dimensional patterns that cells could form in space was unlimited, so it was imperative to make choices. These choices, or constraints, turned out to be rules that specify a body-plan. More precisely the cells are instructed that their position is anterior or posterior, dorsal or ventral and proximal or distal *in respect to the surrounding cells*. These rules are implemented by molecules that are referred to as the *molecular determinants* of the body axes (Gilbert 2006). The crucial point is that there are countless types of molecular determinants and yet all triploblastic animals have the same axes (top-to-bottom, back-to-front and left-to-right). This shows that there is no necessary link between molecular determinants and body axes, and that in turns means that the links that we find in Nature are based on conventional rules, i.e., on the rules of organic codes that can be referred to as the *codes of the body-axes*.

It must be underlined that the body axes are relationships between *cells*, and this means that they do not determine only the axes of the body as a whole, but also those of all its constituent parts. In the hand, for example, the proximo-distal axis is the direction from wrist to fingers, the antero-posterior axis is from thumb to little finger, and the dorsal-ventral axis is from the outer surface to the palm of the hand. Right and left hands have different symmetries because their axes are one the mirror image of the other. There is therefore a multitude of axes in the animal body, and it turns out that many of them have the same molecular determinants. The products of the gene *Sonic hedgehog* (*Shh*), for example, determine the dorso-ventral axis in the forebrain and the antero-posterior axis in the hand, which again shows that molecular determinants are mere labels and represent the conventional rules of a code.

The antero-posterior axis of the body (the head-to-tail direction), is determined by two small depressions that are formed very early on the outer surface of the embryo and represent the signposts of mouth and anus. Between those two points, a third depression is produced by the movements of a colony of migrating cells that invade the space between the first two germ layers (ectoderm and endoderm) to form the middle germ layer (the mesoderm). The invagination point (the blastopore) can be set either near the mouth-signpost (the *stomodeum*) or near the anus-signpost (the *proctodeum*) and that choice determines the future organization of all organs in the body. The animals where the blastopore is formed near the signpost of the mouth (*stoma*) are invertebrates (technically *protostomes*): they have an outside skeleton, a dorsal heart and a ventral nervous system. The animals where the blastopore is formed away from the mouth signpost are vertebrates (more precisely *deuterostomes*): they have an inside skeleton, a ventral heart and a dorsal nervous system.

The whole organization of the body, in other words, is a consequence of a few parameters that determine the migrations of the mesoderm in respect to the body axes. The crucial point is that these migrations (the *gastrulation* movements) take place in countless different ways in both vertebrates and invertebrates, and this shows that they are not due to physical necessity but to the conventional rules of a *gastrulation code*. We realize in this way that the three-dimensional organization the animal body is determined by a variety of organic codes that together can be referred to as the *codes of the body-plan*.

The Origin of Mind

Mind is associated with feelings and sensations, and these are produced in the brain starting from electrical signals that come from the sense organs. Mechanical stimuli, for example, are detected at the surface of the body the by *pressure-receptor*s, and are transformed into *tactile sensations* in the brain. Rats have *mechano-receptors* on the tip of their whiskers while we have them on the tip of our fingers, and there is no doubt that we and rats explore the world in different ways, but do we use different transformation mechanisms? The experimental evidence is that we don't. The physiological processes that transform *mechanical stimuli* into *tactile sensations* seem to be the same in all animals (Nicolelis and Ribeiro 2006).

What is most important is that this is true also for all other sense organs. The experiments on animal brains show that all transformations of *sense stimuli* in *neural sensations* take place according to universal mechanisms. All of which suggests that there has been a universal *neural code* at the origin of mind just as there has been as a universal genetic code at the origin of life.

In the origin of life, the key event was the appearance of *proteins* and the genetic code played a crucial part because it was instrumental to protein synthesis. In the origin of mind, the key event was the appearance of *feelings* and it seems that a *neural code* was as instrumental to the production of feelings as the genetic code was to the production of proteins (Barbieri 2011). The parallel, therefore, is between feelings and proteins, and this immediately tells us that there are both similarities and differences between the two cases.

Proteins are *space-objects*, in the sense that they act in virtue of their three-dimensional structure, whereas feelings are *time-objects* because they are *processes*, entities that consist of flowing sequences of states. The same is true for their components. Proteins are assembled from smaller space-objects like amino acids, and feeling are assembled from lower level brain processes such as neuron firings and chemical signalling. But can we really say that a (nearly) universal neural code exists in all animals as a (nearly) universal genetic code exist in all cells?

We do know that the starting point of all neural processing is the electrical signals produced by sense organs, but we also know that the sense organs arise from the basic histological tissues of the body, and that these tissues (*epithelial*, *connective*, *muscular* and *nervous* tissues) are the same in all triploblastic animals. All signals that are sent to the brain, in other words, come from organs produced by a limited number of universal tissues, and that does make it plausible that they represent a limited number of universal inputs. But do we also have a limited number of universal outputs?

The neural correlates of the sense organs (feelings and perceptions) can be recognized by the *actions* that they produce, and there is ample evidence that all triploblastic animals have the same basic *instincts*. They all have the imperative to *survive* and to *reproduce*. They all seem to experience hunger and thirst, fear and aggression, and they are all capable of reacting to stimuli such as light, sound, smell and mechanical forces. The neural correlates of the basic histological tissues, in short, are associated with the basic animal instincts and these appear to be virtually the same in all triploblastic animals.

What we observe, in conclusion, is a universal set of basic histological tissues on one side, a universal set of basic animal instincts on the other side, and a set of neural transformation processes in between. The most parsimonious explanation is that the neural processes in between are also a universal set of operations. And since there is no necessary physical link between sense organs and feelings, we conclude that the bridge between them can only be the result of a virtually universal *neural code*.

The Codes of Language

We have inherited two modelling (or cognitive) systems from our animal ancestors. The first is the ability to *perceive* an inner and an outer world (a *Innenwelt* and a *Umwelt*), the second is the ability to *interpret* what goes on in the world with processes of extrapolation based on icons and indexes.

When a human baby is born, however, he is thrown into a situation that no animal has ever experienced. He is expelled from the uterus when his foetal development is barely half-way through, and the two animal modelling systems cannot help him because they are not yet fully in place (Barbieri 2010). In that situation a human baby has no choice but resorting to something else, and that is why he starts building a *third* modelling system from the very moment he is born. All he can do, on the other hand, is emitting sounds, and that is precisely what the third modelling system (language) is about: the use of sounds to attract attention.

At first sight it may seem that the enterprise is doomed to fail. A human baby is like a pilot who is ejected from his spaceship before reaching his destination, and in that case the changes of survival would be next to nil. In our case, however, the situation is different, because the extrauterine phase of human development was very short at the beginning of our history, and probably it took a few million years to reach the present value (it would be interesting to find out how long was that phase in the *Australopithecines* and in *Homo erectus*, for example).

The faculty of language, in other words, evolved in stages during the history of our species, and it is developed in stages during the ontogeny of every human being. It is built by a complex sequence of brain-wiring operations, and it is the rules of these operations that we need to discover if we want to understand what language really is.

These rules can be referred to as the *codes of language*, and although we know very little about them, we do have a general idea of what they are about.

One of the first codes of language, for example, is the set of rules that commit our two cortical hemispheres to very different linguistic functions (Sperry 1968, 1982). A second code is the rules by which the two hemispheres interact with each other. The sequence of the other codes of language is less clear, but we do know where it leads to, and this allows us to say when it is that a child has fully acquired the faculty of language. This target is reached when a child becomes capable of *storytelling*, the ability to invent stories and to share them with other people (Lord 1960, 1991).

Language is also used to verbalize concepts that come from our animal ancestors – in particular the concepts that arise from the ability to interpret the world – and has become in this way an extremely composite faculty, but it has not lost its original nature. The codes of language are a unique set that goes from the codes of our cortical hemispheres to the codes of storytelling. What made us human, in short, was not the ability to count, to think rationally, to predict the future and to solve problems. It was a faculty that started with the ability to attract attention by emitting sounds and evolved into the ability to tell stories about imaginary worlds and events that do not exist.

What is particularly inspiring about this idea is that it brings the origin of language in line with the other great events of macroevolution like the origin of life and the origin of mind. The common factor is that they were all associated with the appearance of new biological codes (Barbieri 2003).

The Unexpected Results of Coding

The organic codes may give the impression of being deterministic rules that turn living systems into biological robots, but this far from the truth. They are, in fact, the tools that bring creativity into life. It is the rules of grammar, for example, that allow us to create those endless permutations of words that generate the universe of literature. The key feature of the organic codes is the fact that they bring absolute novelties into existence and in so doing they produce objects that have totally *unexpected* properties. This is a crucial point, and in order to illustrate it let us start from the case of those particular human artifacts that we call *numbers*.

There is little doubt that numbers were generated by counting and that counting was favoured because it had practical advantages. The process of counting, however, produces exclusively *natural* numbers, but then we have discovered *prime* numbers, *fractional* numbers, *rational* and *irrational* numbers, *real* and *imaginary* numbers, and in so doing we have brought to light an endless stream of mathematical theorems. All these *additional* entities were not produced by counting, and this is why some mathematicians say that natural numbers were *invented* by man but all other rules of mathematics had to be *discovered*.

The world of mathematics was generated by the *genetic* rule of counting and then it developed into an increasingly complex world full of additional, or *epigenetic*, properties. A world of codified objects, in short, is a world of *artifacts*, and it is only partially determined by the coding rules that generate the artifacts. In general, it turns out to have unexpected '*rules of its own*', rules that we call epigenetic because they were not present at the beginning and are brought to light only by processes of exploration.

This is what we actually find in living systems. In the world of proteins, for example, there is a universal mechanism in every cell that produces linear polypeptides from linear sequences of genes, but then the polypeptides fold themselves up into three-dimensional structures and take up forms that were not written in the genes. That generates a whole new world of objects, and living cells appear to engage in a veritable exploration of the potentialities of the protein universe.

Another outstanding example is the body-plan of animals. It is based on instructions that specify only three essential relationships between the cells of the body (up and down, back and front, left and right) and yet the number of morphological designs that can be built with them is virtually unlimited.

Language, mathematics, proteins and animals are very different entities but deep down there is something in common between them. They all have (1) a 'genetic' algorithm that produces the objects of a potentially unlimited new world of artifacts (words, numbers, proteins and bodies) and (2) an exploratory procedure that brings into existence additional or 'epigenetic' properties of the new world that were not written in the coding rules and were not present at the beginning.

The organic codes, in conclusion, do not explain *everything*, far from it. They just account for coding. They code for objects that are absolute novelties and which have unpredictable properties. Far from being deterministic rules, the organic codes are the quintessential instruments of creativity and the higher their number the greater is the creative potential of a system. But they account only for the generative rules of life, not for the flesh and blood of history.

The Three Worlds of Life

The genetic code was a precondition for the origin of life, the signal transduction codes divided the first cells into three primary kingdoms, the splicing codes were essential to the origin of the nucleus, the histone code provided the working rules of chromatin, and the cytoskeleton codes allowed the *Eukarya* to perform internal movements, including those of mitosis and meiosis (Barbieri 2003). All great events of cellular evolution, in short, were associated with the appearance of new organic codes, and cellular evolution went on for the first three billion years of the history of life. Eventually however, multicellular creatures did appear, and once again we find that their origins were associated with the appearance of new codes. In the course of multicellular evolution, however, something else did come into existence.

The organic codes have been the sole form of semiosis that existed on Earth in the first three billion years of evolution, but eventually two higher types of semiosis did appear. One evolved in nervous systems and gave animals the ability to *interpret* the world (*animal semiosis*). Interpretation is essentially what Peirce (1906) called an 'abduction', a process that is neither induction nor deduction but *a generalization from limited data whose result can vary according to context, memory and experience*.

More precisely, animals learned to interpret the world by using the two types of signs that Peirce called *icons* and *indexes*. They did not, however, exploit the third type of sign, the *symbols*. Only our species evolved that ability and developed a third type of semiosis that is based on language (*human semiosis*).

The evolution of life was characterized therefore by three great innovations: (1) the origin of organic semiosis, (2) the origin of animal semiosis, and (3) the origin of human semiosis. This fits nicely with the idea of the 'Three Worlds' proposed by Karl Popper (1972, 1979) because these worlds corresponds to three distinct types of semiosis, as illustrated in the Table below. This scheme shows that codes are fundamental components in all three worlds, but do not account for all types of semiosis. More precisely, it shows that coding is the sole mechanism of semiosis in the organic world, whereas animal semiosis is based on coding and interpretation, and human semiosis requires coding, interpretation and language.

The existence of many organic codes in Nature, in conclusion, is not only a major experimental fact. It is one of those facts that have extraordinary theoretical implications. The first, as we have seen, is that the great events of macroevolution were associated with the appearance of new codes. The second comes from the fact that codes create *meaning* and we need therefore to introduce in biology, again with the standard methods of science, not only the concept of information but also the concept of biological meaning. The third major theoretical implication comes from the fact that only the organic codes have been highly conserved in evolution, which means that they are the great invariants of life, the entities that must be perpetuated while everything else is changing. Code Biology, in short, is truly a new field of research because it is bringing to light new basic processes in evolution and new fundamental concepts in biology.

The Three Worlds of Life

Popper's Worlds	*Type of Semiosis*	*Mechanisms*	*Codes*
WORLD 1	Organic Semiosis	Coding	Organic codes
WORLD 2	Animal Semiosis	Coding Interpretation	Organic codes Neural codes
WORLD 3	Human Semiosis	Coding Interpretation Language	Organic codes Neural codes Language codes

Acknowledgments This paper has been written while on a Fellowship at the Stellenbosch Institute for Advanced Study (STIAS) in Stellenbosch, South Africa, as part of the project directed to establishing Code Biology as an autonomous field of research and a new academic discipline. I am deeply grateful to Jan-Hendrik Hofmeyr and Hendrik Geyer for providing the facilities and the environment that made it possible to realize the project, to build a Code Biology website and to write the Constitution of the International Society of Code Biology (ISCB).

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2007). *Molecular biology of the cell* (5th ed.). New York: Garland.

Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., et al. (2010). Deciphering the splicing code. *Nature, 465*, 53–59.

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: Cambridge University Press.

Barbieri, M. (2008). Biosemiotics: A new understanding of life. *Naturwissenschaften, 95*, 577–599.

Barbieri, M. (2010). On the origin of language. *Biosemiotics, 3*, 201–223.

Barbieri, M. (2011). Origin and evolution of the brain. *Biosemiotics, 4*, 369–399.

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

Dhir, A., Buratti, E., van Santen, M. A., Lührmann, R., & Baralle, F. E. (2010). The intronic splicing code: Multiple factors involved in ATM pseudoexon definition. *EMBO Journal, 29*, 749–760.

Gabius, H.-J. (2000). Biological information transfer beyond the genetic code: The sugar code. *Naturwissenschaften, 87*, 108–121.

Gabius, H.-J. (2009). *The sugar code. Fundamentals of Glycosciences*. New York: Wiley-Blackwell.

Gilbert, S. F. (2006). *Developmental biology* (8th ed.). Sunderland: Sinauer.

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology, 3*, 318–356.

Knoll, A. H. (2003). *Life on a young planet. The first three billion years of evolution on Earth*. Princeton: Princeton University Press.

Komander, D., & Rape, M. (2012). The ubiquitin code. *Annual Review of Biochemistry, 81*, 203–229.

Lord, A. B. (1960). *The singer of tales*. Cambridge: Harvard University Press.

Lord, A. B. (1991). *Epic singers and oral tradition*. Ithaca: Cornell University Press.

Maraldi, N. M. (2008). A lipid-based code in nuclear signalling. In M. Barbieri (Ed.), *The codes of life: The rules of macroevolution* (pp. 207–221). Dordrecht: Springer.

Nicolelis, M., & Ribeiro, S. (2006). Seeking the neural code. *Scientific American, 295*, 70–77.

Osborne, L. C., Palmer, S. E., Lisberger, S. G., & Bialek, W. (2008). The neural basis for combinatorial coding in a cortical population response. *Journal of Neuroscience, 28*(50), 13522–13531.

Peirce, C. S. (1906). The basis of pragmaticism. In C. Hartshorne & P. Weiss (Eds.), *The collected papers of charles sanders peirce, vols I-VI* (pp. 1931–1935). Cambridge: Harvard University Press.

Pertea, M., Mount, S. M., & Salzberg, S. L. (2007). A computational survey of candidate exonic splicing enhancer motifs in the model plant *Arabidopsis thaliana. BMC Bioinformatics, 8*, 159.

Popper, K. (1972). *Objective Knowledge*. Oxford: Clarendon.

Popper K (1979) Three worlds. *Michigan quarterly review*, 18 (1): 1–23. Reprint of The Tanner Lecture on Human Values delivered by Karl Popper at The University of Michigan on April 7, 1978. http://tannerlectures.utah.edu/lectures/documents/popper80.pdf.

Readies, C., & Takeichi, M. (1996). Cadherine in the developing central nervous system: An adhesive code for segmental and functional subdivisions. *Developmental Biology, 180*, 413–423.

Schopf, J. W. (1999). *Cradle of life: The discovery of Earth's earliest fossils*. Princeton: Princeton University Press.

Shapiro, L., & Colman, D. R. (1999). The diversity of cadherins and implications for a synaptic adhesive code in the CNS. *Neuron, 23*, 427–430.

Sperry, R. W. (1968). Mental unity following surgical disconnection of the cerebral hemispheres. *Harvey Lectures, 62*, 293–323.

Sperry, R. W. (1982). Some effects of disconnecting the cerebral hemispheres. *Science, 217*, 1223–1226.

Strahl, B. D., & Allis, D. (2000). The language of covalent histone modifications. *Nature, 403*, 41–45.

Tomkins, M. G. (1975). The metabolic code. *Science, 189*, 760–763.

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. (1999). Elucidating sequence codes: Three codes for evolution. *Annls of the New York Academy of Sciences, 870*, 330–338.

Tudge, C. (2000). *The variety of life. A survey and a celebration of all the creatures that have ever lived*. Oxford: Oxford University Press.
Turner, B. M. (2000). Histone acetylation and an epigenetic code. *Bioessays, 22*, 836–845.
Turner, B. M. (2002). Cellular memory and the histone code. *Cell, 111*, 285–291.
Turner, B. M. (2007). Defining an epigenetic code. *Nature Cell Biology, 9*, 2–6.
Verhey, K. J., & Gaertig, J. (2007). The tubulin code. *Cell Cycle, 6*(17), 2152–2160.
Woese, C. R. (1987). Bacterial evolution. *Microbiological Reviews, 51*, 221–271.
Woese, C. R. (2000). Interpreting the universal phylogenetic tree. *Proceedings of the National Academy of Sciences of the United States of America, 97*, 8392–8396.