Arbitrariness

In science, arbitrariness means independence from physical necessity. Codes and conventions, for example, are arbitrary because their rules are not dic- tated by the laws of physics. In the case of the genetic code, it is has been proven beyond doubt that any codon can be associated with any amino acids, just as any group of dots and dashes can be associated with any letter of the alphabet. The genetic code, in short, is a real code because it does have the essential feature that defines any code: the arbitrariness of the coding rules. It must be underlined, however, that this point has raised streams of objec- tions, all claiming that arbitrariness is a myth because there are all sorts of regularities in the genetic code. In reality, a few simple cases are enough to deflate this argument. In the Morse code, for example, the most frequent letters of the alphabet are associated with the simplest combinations of dots and dashes, but nobody would dream to conclude that the Morse code is not made of arbitrary rules because of that regularity. In any language there are countless regularities, and yet arbitrariness exists even in the number and the type of letters that make up an alphabet. Regularities, in short, are perfectly compatible with arbitrariness. What they are not compatible with is randomness, but arbitrariness should not be confused with randomness.

Biolinguistics

It is a research field dedicated to studying language as any other attribute of our species, and more specifically, as an organ of the mind/brain. In the editorial of the first issue of the journal Biolinguistics (2007) the editors Cedric Boeckx and Kleanthes Grohmann stated that there is both a weak and a strong sense to the term ‘biolinguistics’. The weak sense refers to the fact that linguists are seriously engaged in “discovering the properties of grammar, in effect carrying out the research program Chomsky initiated in Syntactic Structures (1957)". The strong sense refers to "attempts to pro- vide explicit answers to questions that require the combination of linguistic insights and insights from related disciplines (evolutionary biology, genetics, neurology, psychology, etc.)." They underlined that Eric Lenneberg's book, Biological Foundations of Language (1967), was an outstanding example of research in biolinguistics in the strong sense.

Biosemiotics

The synthesis of biology and semiotics that today we call 'biosemiotics' was developed independently in two fields that lie at the opposite ends of academia. The first origin took place in molecular biology as a result of the discovery of the genetic code (the name "molecular biosemiotics" was coined by Marcel Florkin in 1974 precisely to designate the study of semio- sis at the molecular level). The second origin took place in the humanities and was masterminded by Thomas Sebeok in two distinct stages. In 1963, Sebeok extended semiosis from human culture to all animals and founded the new research field of 'zoosemiotics' (Sebeok 1963). More than 20 years later, he made a second extension from animals to all living creatures and called it 'biosemiotics' (Anderson et al. 1984, Sebeok and Umiker-Sebeok 1992, Sebeok 2001). These two 'birthplaces' of biosemiotics have nurtured two different concepts of semiosis that still divide the field into two opposite schools. In biology, the existence of a real genetic code is proof enough that semiosis exists at the molecular level, and this implies that organic semiosis is defined by coding. In the humanities, the dominant view is the Peircean concept that semiosis is always an interpretive process, and this implies that Peircean semiosis is defined by interpretation. We have therefore two types of semiosis, one based on coding and one based on

interpretation, and each of them represents phenomena that undoubtedly exist in Nature. There is ample evidence that animals are capable of interpreting the world, and this clearly means that Peircean (or interpretive) semiosis is a reality. But it is also evident that the rules of the genetic code do not depend on interpretation because they have been the same in all living creatures and in all environ- ments ever since the origin of life. The division between the two schools of biosemiotics is precisely about this point. According to the 'biological' school, the two types of semiosis are both present in Nature and represent two distinct evolutionary developments. According to the Peircean school, instead, interpretive semiosis is the only type that has existed on Earth ever
since the origin of life.

Code
A code is a set of rules that create a correspondence between two independent worlds. 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 signals and driving behaviours. A language is a correspondence between words and objects. The genetic code is a correspondence between triplets of nucleotides, called codons, and amino acids. What is essential in all codes is that the coding rules are not dictated by the laws of physics. They are arbitrary in the sense that they are independent from physical necessity and this implies that they can be established only by natural or by cultural conventions.

Code Biology
Code Biology is the study of all codes of life, from the genetic

code to the codes of culture. 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. According to official (textbook) science, these are the only codes that exist in Nature, and if this were true we would have to conclude that codes are extraordinary exceptions that 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 last 25 years, and it seems likely that many more will come to light in the future. The existence of many organic codes in Nature, however, is not only a new experimental fact. It is one of those facts that have extraordinary theoretical implications. The first is that all great events of macroevolution were associated with the appearance of new organic codes, and this gives us a completely new description of the history of life. The second great implication is about the mechanisms of evolution. The discovery that there are two fundamental molecular mechanisms at the basis of life, copying and coding, means that there are two distinct mechanisms of evolutionary change: evolution by natural selection, based on copying, and evolution by natural conventions, based on coding. The experimental discoveries and the theoretical implications of the organic codes make of Code Biology an entirely new field of research and an autonomous academic discipline, the real new frontier of biology.

Codepoiesis

Before the origin of the genetic code, the ancestors of the first cells were engaged in the process of evolving coding rules and contained therefore a code generating system. After the origin of the genetic code, however, the situation changed dramatically. No other modification in coding rules was tolerated and the system in question became a code conservation system. Another part of the system, however, maintained the potential to evolve other coding rules and behaved as a new code generating, or code exploring, system. 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 the two sides of a biological phenomenon that can be referred to as 'codepoiesis'. What is common to all living systems is either the generation or the conservation, or both, of organic codes, and this gives us an entirely new definition of the cell that can be expressed in this way: "the cell is a codepoietic system, i.e., a system that is capable of creating and conserving its own codes". This definition accounts for the two most important events of evolution. [1] The ability to create coding rules accounts for the origin of the genetic code and of all the other codes that followed. [2] 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 are conserved while everything else is changing.

Coding semiosis

Semiosis requires the existence of four distinct entities—signs, meanings, code and codemaker—because signs cannot exists without meanings and the re- lationship between them is necessarily based on the rules of a code, which in turn implies an agent that produces them, i.e., a codemaker. Semiosis, in other words, is necessarily based on coding, but this does not mean that it is based exclusively on coding. In the course of evolution, some animals have acquired the ability to interpret the world, and interpretation is cer- tainly a form of semiosis, since it makes use of signs and meanings, but it is not based on coding alone because it also requires memory, learning, mental representations and probably some form of 'abduction'. For the first three thousand million years of evolution, the Earth has been inhabited exclusively by single cells, and these are capable of coding and decoding the world but do not make representations and therefore cannot interpret them. In order to distinguish the semiosis of single cells from that of animals it is convenient to use terms that qualify them and to this purpose the two types are referred to as coding semiosis and interpretive semiosis.

Copying and Coding

Copying and coding are mechanisms that work at two distinct levels in every living system. Copying operates at the individual level of the molecules and coding at the collective level of the whole system. None of them is reducible to the other because they are complementary mechanisms. They evolved in parallel in the history of life just as individual words and rules of gram- mar evolved in parallel in the history of language. There are, furthermore, other two important differences between copying and coding. One is the fact that copying produces either exact copies or slightly different versions of the copied molecules, which means that natural selection produces new objects by gradually modifying preexisting ones. Natural selection, in other words, creates only relative novelties, not absolute ones. In the case of coding, in- stead, the situation is totally different. The rules of a code are not dictated by physical necessity, and this means that they can establish relationships that have never existed before in the Universe. Natural conventions, in short, have the potential to create absolute novelties. Another difference between copying and coding is that they involve two different entities. A variation in the copying of a gene changes the linear sequence, i.e., the information of that gene. A variation in a coding rule changes instead the meaning of that rule. The great difference that exists between copying and coding, and therefore between natural selection and natural conventions, comes from the difference that exists between 'information' and 'meaning'. There are, in short, three major differences between copying and coding: (1) copying acts on individ- ual molecules whereas coding acts at the collective level, (2) copying modifies existing objects whereas coding brings new objects into existence, and (3) copying is about biological information whereas coding is about biological meaning.

Copying semiosis

In protein synthesis, a sequence of nucleotides is used to produce a sequence of amino acids according to the rules of the genetic code. In that case, there is no necessary connections between the components of the two molecules and the codons of nucleotides are used as conventional organic signs, i.e., as organic symbols. A sequence of nucleotides, however, can also be used by a copymaker (a polymerase) to produce a complementary copy of itself, and in that case the relationship between the two sequences is no longer established by adaptors but by direct physical interactions between complementary regions. These interactions, however, occur between very small regions of the molecules, and that means that the first sequence provides only a limited number of physical determinants for the second. The first sequence, in other words, does have a physical relationship with the second, but such relationship is undetermined and represents therefore only a 'cue', i.e., a natural sign, for the second. This means that the distinction between natural and conventional signs

exists also at the molecular level, and represents in fact a divide between two very different types of processes. Sequences of nucleotides are used as conventional signs in coding and as natural signs in copying. Molecular coding, in short, is a form of coding semiosis whereas molecular copying is a process of copying semiosis. The translation of genes into proteins, in other words, is based on coding semiosis whereas the replication and the transcription of genes are based on copying semiosis.
[Back to Contents]

]

Cultural Semiosis
(see Popper's Three Worlds and)

Information
(see Organic information)

Interpretation
Memory allows a system to compare a phenomenon with previous records of similar phenomena, and it is from such comparisons that a system can 'learn' from experience. Memory is clearly a prerequisite for learning, but what does learning achieve? What is the point of storing mental representations and comparing them? So far, the best answer to this problem is probably the idea that memories and learning allows animals to interpret the world. Interpretation, on the other hand, is a form of semiosis—because it is based on signs—but it is a new form because it also requires memory and learning. What is interpreted, furthermore, is not the world but representations of the world, and only multicellular systems can build them. Single cells decode the signals from the

environment but do not have the physical means to build internal representations of them and therefore cannot interpret them. They are sensitive to light, but do not 'see'; they react to sounds but do not 'hear'; they detect hormones but do not 'smell' and do not 'taste' them. It takes many cells which have undertaken specific processes of differentiation to allow a system to see, hear, smell and taste, so it is only multicellular creatures that have these experiences. The evolution from single cells to animals was a true macroevolution because it created absolute novelties such as feelings and instincts. Later on, another macroevolution gave to many animals the ability to interpret the world, and we can actually prove that this ability evolved in stages. The origin of interpretation provided animals with a new means of obtaining information about the world—a second modelling system—and gave origin to a new type of semiosis that can be referred to as interpretive, or Peircean, semiosis.

Interpretive semiosis
(see Coding semiosis and Biosemiotics)

Meaning
(see organic meaning)

Neural semiosis
(see Popper's Three Worlds)

Organic Codes
Are codes between organic molecules. Any organic code is a set

of rules that establish a correspondence between two independent organic worlds by means of molecular structures, called adaptors, that perform two independent recognition processes at each step. In the genetic code, for example, the adaptors are the transfer-RNAs. 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 are the key molecules 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. In addition to the genetic code, the existence of many other organic codes has been reported so far. Among them: the sequence codes (Trifonov, 1987, 1989; 1999), the adhesive code (Redies and Takeichi, 1996; Shapiro and Colman, 1999), the splicing codes (Barbieri, 1998, 2003; Pertea et al., 2007; Barash et al. 2010; Dihr et al., 2010), the signal transduction codes (Barbieri, 1998, 2003), the sugar code (Gabius, 2000, 2009), the histone code (Strahl and Allis, 2000; Turner, 2000; 2002), the cytoskeleton codes and the compartment codes (Barbieri, 2003, 2008), the tubulin code (Verhey and Gertig, 2007), a nuclear signalling code (Maraldi, 2008), and the ubiquitin code (Komander and Rape, 2012).

Organic information

In genes and proteins, biological, or organic, information has been defined as the specific sequence of their subunits. This definition however is not entirely satisfactory because it gives the impression that information is 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 a sequence provides a guide- line to a copymaker in a process of copying. It is only an act of copying, in other words, that brings information into existence. This tells us that organic information is not just the specific sequence of a molecule, but the specific sequence produced by a copying process. This definition underlines the fact that organic information is not a thing or a property, but the result of a process. It is, more precisely, an 'operative' definition, because information is defined by the process that brings it into existence. It must also be underlined that organic information is neither a quantity (because a specific sequence cannot be measured), nor a quality (because it is an objective feature of all copied molecules), and belongs instead to a third class of objects that have been referred to as nominable entities.

Organic meaning

The Morse code is a correspondence between the letters of the alphabet and groups of dots and dashes and in the same way the genetic code is a correspondence between groups of nucleotides and amino acids. Let us notice now that establishing a correspondence between, say, object 1 and object 2, is equivalent to saying that object 2 is the meaning of object 1. In the Morse code, for example, the rule that 'dot-dash' corresponds to the letter 'A', is equivalent to saying that letter 'A' is the meaning of 'dot-dash'. By the same token, the rule of the genetic code that a group of three nucleotides (a codon) corresponds to an amino acid is equivalent to saying that that amino acid is the organic meaning of that codon. Anywhere there is a code, be it in the mental or in the organic world, there is meaning. We can say, therefore, that meaning is an entity which is related to another entity by a code or a convention, and that organic meaning exists whenever an organic code exists. 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. It must also be underlined that organic meaning—like organic information—is neither a quantity (because a coding rule cannot be measured), nor a quality (because the organic codes are objective features o life), and belongs instead to a third class of objects that have been referred to as nominable entities.

Organic semiosis
(see Signs and Biosemiotics)

Popper's Three Worlds
In the 1970s, Karl Popper argued that the unity of Nature is realized by the coexistence of three distinct domains, or 'Worlds'. The first (World 1) is the domain of all material objects, physical and biological, i.e., atoms, galaxies and bodies. The second (World 2) is the domain of the mind, the subjec- tive world of mental states, feelings, emotions and consciousness. The third (World 3) is the domain of all human artifacts and cultural products. The three worlds could hardly be more different, and yet they do have something in common. At the heart of all of them there are codes. The genetic code and other organic codes in World 1, neural codes in World 2 and countless cultural codes in World 3. The three worlds of Popper correspond therefore to three major types of semiosis that are referred to as organic, neural and cultural semiosis.

Representations
(see Interpretation)

Semantics

(see Syntax and semantics)

Semiosis

Semiosis is the production of signs, and semiotics is usually referred to as the study of signs (from the Greek semeion=sign) but these definitions are too restrictive because signs are always associated with other entities. A sign, to start with, is always linked to a meaning, which implies that sign and meaning cannot be taken apart because they are the two sides of the same coin. Semiotics, therefore, is the study of signs and meanings together, and a system of signs, i.e., a semiotic system, is always made of at least two distinct worlds: a world of entities that we call signs and a world of entities that represent their meanings. The link between sign and meaning, in turn, calls attention to a third entity, i.e., to their relationship. A sign is a sign only when it stands for something that is other than itself, and this otherness implies at least some degree of independence. It means that there is no deterministic relationship between signs and meanings. A semiotic system, therefore, is not any combination of two distinct worlds. It is a combination of two worlds between which there is no necessary link, and this implies that a bridge between the two worlds can be established only by conventional rules, i.e., by the rules of a code. This is what makes semiosis different from everything else: semiosis requires a system made of two independent worlds that are connected by the conventional rules of a code. A semiotic system, in other words, is necessarily made of at least three distinct entities: signs, meanings and code. Signs, meanings and code, however, do not come into existence of their own. There is always an 'agent' that produces them, and that agent can be referred to as a codemaker. In the case of culture, for example, the codemaker is the human brain; in the case of the cell, the codemaker is the ribonucleoprotein system that makes proteins according to the rules of the genetic code. We come in this way to a general conclusion that can be expressed in this way: a semiotic system consists of signs, meanings and code that are all produced by the same agent, i.e., by the same codemaker.

Semiotic dynamics

Semiotic dynamics is the study of the self-organizing and evolutionary dy- namics that lead to the evolution of conventional meaning and codes, includ- ing language. The term was first introduced in a 1999 paper by Luc Steels and Frederic Kaplan [1] which reported on a study of the conventionalization dynamics leading to the emergence of a shared lexicon in a group of au- tonomous distributed agents situated and grounded in an open environment. Other publications along this line of research include [2, 3, 4, 5]. The topic was later picked up by researchers in statistical physics (see, among others, [6, 7, 8, 9]) and is related to models of the evolution of cooperation and conventions in the field of evolutionary game theory, specifically signaling games and pre-play signaling [10, 11]. Some of the main theoretical results are that conventionalization is bound to occur in a population of agents if the agents are predisposed or at least have an incentive to cooperate and there is amplifying individual learning [3, 10] or if there is learning at the popu- lation level [11, 12]; and that the dynamics of conventionalization are like the dynamics of phase transitions [6]. These results are compatible with the hypothesis that the mechanism of evolution by natural conventionalization is responsible for the Major Transitions in macro evolution.

Signs

Signs have been traditionally defined as "something that stands for something else", and in antiquity were divided into two great categories—conventional signs and natural signs—for two different reasons. One is because they derive either from nature (signa ex natura) or from culture (signa ex cultura). The other is because they are either symbols (signa symbolica) or symptoms (signa symptomatica). If we put together both characteristics, signs are defined in the following way:
1. the conventional signs are signa symbolica ex cultura, and 2. the natural signs are signa symptomatica ex natura. The discovery of the genetic code came as a bolt from the blue precisely because it revealed the existence of a third category of signs that all thinkers of the past had not predicted: the existence of symbols that come from nature, not from culture. In addition to the two classical categories, therefore, we now have a third one:
3. the organic signs are signa symbolica ex natura. This is the immense novelty of the genetic code. It brought to light a third type of semiosis that exists in the organic world and for this has been called organic semiosis.

Syntax and Semantics

In biolinguistics syntax refers to the study of the formal rules and principles that govern when sentences are well formed, without taking into account their meaning. In logics syntax refers to the formal rules that determine when logical statements are well formed without taking into account their meaning and, by extension, their truth-value. In computer science syntax refers to the rules that determine when statements in a computer program are well formed without taking into account (the result of) the actual computations denoted by the statements. Thus in general syntax is defined as the study of the properties of well formed statements in a symbol system without taking into account what the meaning or semantics is of the symbols and statements. There are some subtleties in this definition however. Firstly, symbols do not exist in or by themselves but only as the result of being qualified as such by some system, just as codons only exist within the context of a system consisting of a ribonucleoprotein and tRNA's. Secondly, the definition does not specify what it means to be 'well formed'. Again, deciding whether a statement is 'well formed' is an act of qualification. Since qualification is a matter of semantics, these issues indicate that syntax and semantics are in fact inseparable and intrinsically semiotic notions.

Zoosemiotics

It is the study of semiotics processes in animals, a field that Thomas Sebeok started with a research paper in 1963 and to which he gave its present name in the book Perspectives in Zoosemiotics (1972).

References

[1] L. Steels, F. Kaplan, Collective learning and semiotic dynamics, in: Flo- reano, Nicoud, Mondada (Eds.), Advances in artificial life, Proceedings of ECAL'99, Vol. 1674 of Lecture Notes in Computer Science, 1999.

[2] J. De Beule, B. De Vylder, T. Belpaeme, A cross-situational learning algorithm for damping homonymy in the guessing game, in: L. M. R. et al. (Ed.), Artificial Life X, MIT Press, 2006, pp. 466–472.

[3] B. De Vylder, K. Tuyls, How to reach linguistic consensus: A proof of convergence for the naming game, Journal of Theoretical Biology 242 (4) (2006) 818–831.

[4] B. De Vylder, The evolution of conventions in multi-agent systems, Ph.D. thesis, VUB Artificial Intelligence Lab (2007).
[5] P. Wellens, M. Loetzsch, L. Steels, Flexible word meaning in embodied agents, Connection Science 20 (2) (2008) 173–19.
[6] A. Baronchelli, M. Felici, E. Caglioti, V. Loreto, L. Steels, Sharp transi- tion towards shared vocabularies in multi-agent systems, J. Stat. Mech. P06014.
[7] C. Cattuto, V. Loreto, L. Pietronero, Semiotic dynamics and collabora- tive tagging, Proceedings of the National Acadamy of Sciences 104 (5) (2007) 1461–1464.
[8] A. Puglisi, A. Baronchelli, V. Loreto, Cultural route to the emergence of linguistic categories, PNAS 105 (23) (2008) 7936–7940.
[9] V. Loreto, A. Mukherjeeb, T. F., On the origin of the hierarchy of color names, PNAS 109 (18) (2012) 2403–2407.
[10] B. Skyrms, Signals, Evolution, Learning and Information, Oxford Uni- versity Press, New York, 2010.
[11] F. Santos, J. Pacheco, S. B., Co-evolution of pre-play signaling and cooperation, Journal of Theoretical Biology 274 (1) (2011) 30–5.
[12] J. De Beule, Overcoming the tragedy of the commune in the hawk-dove game through conventional coding, in: In Proceedings of BeneLearn and
PMLS, Ghent, Belgium, 2012.