

# Life as Interplay of Information and Matter

G rard Battail

*Retired from E.N.S.T., Paris.*

Received: March 14, 2015 / Accepted: April 11, 2015 / Published: July 25, 2016

**Abstract:** Any sequence can be transformed into an equivalent one by alphabet change and/or encoding. Thus an information cannot be identified to a single sequence, but should be defined as the equivalence class among sequences with respect to such transformations, an abstract entity. It may be represented by its ‘information message’ defined as the shortest binary sequence in this class, whose length quantitatively measures the information. Any sequence must be borne by a physical medium, so information bridges the abstract and the concrete. Perturbations in the physical world result in random symbol errors, while information may instruct the assembly of physical objects by the agency of the semantics it bears. Information theory tells that errorless communication is possible provided a long and redundant enough error-correcting code is used. Conserving genetic information over the ages needs error-correcting codes making the exact regeneration of genomes possible provided it is performed frequently enough. In engineering, such codes are conveniently defined by mathematical constraints which make the symbols mutually dependent, but constraints of any kind, defining ‘soft codes’, also do the job. The better conservation of old parts of the genome implies moreover that the genomic error-correcting codes are made of nested component codes. This scheme accounts for basic life features: need for successive generations, existence of discrete species with hierarchical taxonomy, trend of evolution towards increased complexity, etc. Transcription of the genes and their translation into polypeptidic chains, becoming proteins when properly folded, are controlled by enzymes, which as proteins are needed for their own synthesis. A feedback loop results, referred to as ‘semantic’ since it implements the genetic code which tells the meaning of the genes. Once closed, it keeps its structure, thereby conserving the genetic code. The onset of a new semantic feedback loop originates an organic code in Barbieri’s meaning, which itself induces a new soft component in the system of genomic nested error-correcting codes by imposing new constraints to the genome. Due to the feedback, constraints incurred by the proteins affect the genome. Besides the synthesis of proteins, the assembly of living structures, as instructed by the genome, involves interwoven semantic feedback loops. Although closed, they do not prevent the genome lengthening, say by horizontal genetic transfer, which increases both the information quantity,

---

**Corresponding author:**

G rard Battail, Retired from E.N.S.T., Paris. E-mail: [gb@gerard.battail.name](mailto:gb@gerard.battail.name).

hence its semantic specificity, and the redundancy, hence its error-correction ability. Thus, closing semantic feedback loops does not prevent evolution: new structures can be appended with improved resilience to errors. Regeneration failure is very infrequent and results in a widely different genome which originates a new species if the phenotype it specifies withstands the Darwinian selection. A possible scenario for the origin of semantic feedback loops (hence maybe of life) is proposed.

## 1. Introduction

This paper interprets life as resulting from the interaction of information and matter. Defining information according to the engineering practice shows that it is basically an abstract entity. Being borne by symbolic sequences inscribed on a physical medium, information bridges the abstract and the concrete.

The concrete world acts on information because physical perturbations permanently degrade information-bearing sequences. However, error-correcting codes enable their regeneration, which conserves information if it is performed as frequently as to almost always succeed, i.e., before the cumulated number of errors exceeds the correction ability of the code.

Information acts on matter by the agency of genomes which instruct the assembly of living structures, which especially include semantic feedback loops locked by enzymes which catalyse their own assembly, to be described in the sequel.

The successive establishment of semantic feedback loops induced constraints which endowed genomes with a system of nested error-correcting codes which protects them and conserves the information they bear. At variance with engineering where the coding constraints are defined by mathematical relationships, they are here physical-chemical (especially steric) or linguistic, and the codes they induce are referred to as *soft*. These codes successively appeared during the ages thus resulting in a *nested* system which is likened to Barbieri's organic codes [1, 2, 3].

A more comprehensive account of the research presented here can be found in [4].

## 2. The Physical World Acts on Information

### 1. Information is an abstract entity which dwells in the physical world

Only *discrete* information is relevant in biology. The extension to continuous information is possible but involves mathematical difficulties and will be left aside. A discrete information is assumed to be represented by a *sequence* of symbols belonging to some finite set referred to as the *alphabet*. The sequence is borne by some physical medium referred to as its *support*.

Defining information according to the engineering practice shows that it is basically non-physical (at variance with the opinion of Schrödinger [5], Brillouin [6], and of most contemporary physicists). Since any sequence can be transformed into an equivalent one by alphabet change and/or encoding, an information cannot be identified to a single sequence. It should indeed be defined as the *equivalence class* among sequences with respect to such transformations. It is thus an *abstract entity*.

An information may then be represented by its ‘information message’ defined as the shortest binary sequence<sup>1</sup> in this class, whose length in binary digits (bits) quantitatively measures the information. Its bits are mutually independent, equiprobable, and each of them is essential to the information’s integrity. No topology exists within informations: any change in an information message results in the representative of another information. An information is thus a *nominable entity* in Barbieri’s meaning.

Associating with each bit of an information message a dichotomic choice (answering a question or doing an action) endows the information it represents with a *semantic content* which possibly refers to the *concrete* world. An information then appears as a content for semantics, just like a shell contains a hermit crab. Thus, the length of the information message is a measure of the semantic specificity as well as that of the information quantity.

As an abstract entity, an information is represented by a sequence of symbols. Any sequence must be borne by a physical medium, so information actually dwells in the physical world. Information thus *bridges* the *abstract* and the *concrete*.

Information can be annihilated if its physical support is destroyed, but it can also be shared if it is written on several distinct supports.

Life is interpreted here as resulting from the interplay of information and matter: the physical world acts on information, but information too acts on the physical world. This interplay is highly dissymmetrical: perturbations in the physical world result in random symbol errors affecting information-bearing sequences, while information may instruct the assembly of physical objects by the agency of the semantics it bears.

## 2. Physical perturbations irreversibly degrade sequences

The physical world is seen since Ludwig Boltzmann as basically *chaotic*, especially at the molecular scale. ‘Thermal noise’ is just the macroscopic average result of random molecular movements. The second law of thermodynamics states that any physical system progressively incurs an irreversible degradation since its physical entropy can but increase.

---

<sup>1</sup>Its existence and uniqueness is stated by the fundamental theorem of source coding of information theory.

A physical medium which bears some sequence does not escape this degradation, which results in random errors affecting the sequence symbols. This is especially true of the DNA molecule. Indeed, genomic mutations occur during time intervals as short as a human life. The average number of erroneous symbols in the sequence is an increasing function of time, so the ability of the medium to keep its initial information, measured by its *capacity* (an upper bound of it is calculated in [4], Eq. (8.2), p. 166), approaches zero as time passes. This fact seems to preclude the long-term conservation of any sequence, hence of any information.

Genetics tells that genomes instruct the assembly of living things within the physical world. Then the information borne by genomes acts on material structures by the agency of its semantic content. It turns out that life, hence heredity, lasts for at least 3.5 billion years, and this undeniable fact seems in absolute contradiction with the above statement which however is no less undeniable, that any sequence-bearing medium suffers an inescapable degradation. We show how information theory succeeds in solving this blatant contradiction, thanks to error-correcting codes.

### 3. How Genomic Information Resists Physical Perturbations

#### 1. *Errorless communication is possible despite symbol errors*

The fundamental theorem of channel coding in information theory tells that *errorless* communication of a sequence is possible in the presence of *symbol errors*. More precisely, the probability of the erroneous recovery of a sequence tends to zero if it is optimally encoded into a sequence of length approaching infinity [7,8]. This paradoxical but highly useful result is obtained by channel encoding, which consists of replacing the given sequence by a longer but fully equivalent one, belonging to a set of sequences as different from each others as to enable identifying one of them even if a number of its symbols are in error. A set of sequences of finite length endowed with this property is referred to as an *error-correcting code*. Being necessarily longer than the original sequence, a sequence belonging to the code, or *codeword*, is *redundant*. The more redundant, the more efficient can be a code. Encoding an information message by means of such a code protects the information it represents.

Notice that ‘error-correcting’ is somewhat misleading when qualifying finite-length codes: correction is at best highly likely, but it cannot be absolutely guaranteed.

#### 2. *Genomic error-correcting codes are needed*

The existence of genomic error-correcting codes must be assumed in order to account for the conservation of genomes [9, 10, 11]. Physical perturbations result in symbol errors occurring with non-zero

probability, which entails that the average cumulated number of errors increases with time. Assuming it is encoded into an error-correcting code, a genome can be *exactly regenerated* if the cumulated number of symbol errors does not exceed the correction ability of the code. The genome is thus conserved with high probability provided it is regenerated frequently enough. The number of cumulated symbol errors within some time interval is random, but if this number is large enough in the average (hence for a code long enough), its variance is comparatively small as a consequence of the law of large numbers. Then the genome conservation is highly probable if the time interval between successive regenerations properly matches the code performance, given the average frequency of errors.

There is no other way to solve the contradiction between the conservation of genomes and the unavoidable degradation of sequences than assuming the existence of genomic error-correcting codes. We now list their needed properties:

- they should be (very) redundant;
- at variance with engineering codes, they are not necessarily defined by mathematical equalities, but may result as well from physical-chemical or linguistic *constraints*. In this case, we refer to them as *soft codes*;
- the conservation of very old parts of genomes like the *HOX* genes demands that they are made of several *nested component codes* which originated successively during the ages.

### 3. Genomes are redundant

With a 4-symbol alphabet, a sequence of 133 nucleotides suffices to count the atoms of the visible universe since the number of distinct sequences having this length,  $4^{133}$ , approximately equals  $10^{80}$ , the estimated number of these atoms. This moderate length of 133 nucleotides is to be compared with that of genomes, 1,000 or so for the simplest viruses,  $10^6$  at least for bacteria and much more for animals and plants, e.g., about  $3.2 \times 10^9$  for humans. There is thus room for an *immense* redundancy.

Stuffing cannot be responsible for the excess beyond the number of symbols which is strictly necessary because any symbol in a codeword contributes in the conservation of this codeword, hence in its own conservation. There are thus no ‘junk’ symbols. Else, it would be impossible to associate a genome with a species. The living world would be populated with chimeras, not with species members.

### 4. Genomes are protected by ‘soft’ error-correcting codes

We have seen in Sec. 1 that the effectiveness of an error-correcting code results from its words being as different as possible from each other. This is obtained in engineering by imposing mathematical constraints

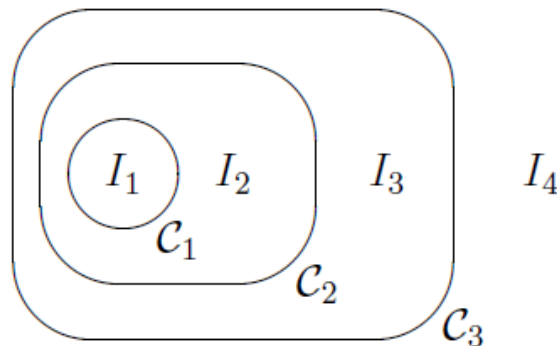
on codewords. The very nature of constraints is however unimportant. What matters is their *existence*, and the Shannon-McMillan theorem of information theory states under rather general conditions that constraints on a set of sequences entail its error-correcting ability [12]. Codes, which we refer to as ‘soft’, can thus be defined by non-mathematical constraints. It turns out that the DNA molecule is affected with constraints of several kind, especially steric. For instance its wrapping around histone octamers (in eukaryotic cells) induces constraints on the successive nucleotides. Moreover similar constraints affect proteins, hence the polypeptidic chains from which they originate, and induce constraints on the genes which instruct their assembly (by the agency of semantic feedback loops, to be introduced later). Some structures in the proteins like  $\alpha$ -helices and  $\beta$ -sheets actually prevent certain successions of amino-acids, hence of the corresponding codons. All these constraints generate soft codes as components of the overall genomic code.

Besides containing the genes which specify proteins, genomes instruct the assembly of larger-scale living structures, which demands some syntax, thus implying the existence of linguistic constraints, hence of other component codes in addition to the physical-chemical ones already mentioned.

##### *5. Genomic error-correcting codes are made of nested component codes*

In the absence of coding, the oldest parts of the genome would be the most degraded. It is the exact contrary which is true: some of the best conserved parts of the genomes, e.g., the *HOX* genes but also some parts not involved in the synthesis of proteins, are the oldest ones. This fact is easily explained by assuming that the genomic error-correcting code has progressively been established during the ages, by successive encodings resulting in *nested codes*. Some information message has once been encoded. Another information message has been later appended to the result of the first encoding, and the message thus obtained has been encoded again. The initial information is thus protected twice, by its first encoding and because the result of this encoding has been itself later encoded. This process can be repeated arbitrarily many times. The older an information message, the more numerous component codes protect it and the better it is conserved. A very redundant and heterogeneous code results if the component codes are many.

The fortress metaphor provides an intuitive illustration of nested component codes: a code is represented as a closed wall which protects what is inside it against outside attackers (see Fig. 1). Several concentric walls have been successively built to enclose information, so the content of the oldest, most central, wall is much better protected than the more recent and peripheral information. A multiplicity of walls is much safer than each of them separately.



**Figure 1.** A code  $C_j$ ,  $j=1,2,3$ , is represented as a closed wall which protects what is inside it.  $I_1$ ,  $I_2$ ,  $I_3$  and  $I_4$  are successive information messages.  $I_1$  is protected by 3 codes,  $I_2$  by 2 codes,  $I_3$  by a single code and  $I_4$  is left uncoded.

#### 6. The assumed existence of such codes explains basic features of the living world

The properties of a system of nested genomic error-correcting codes explain many basic features of the living world left unexplained by mainstream biology. To list a few of them:

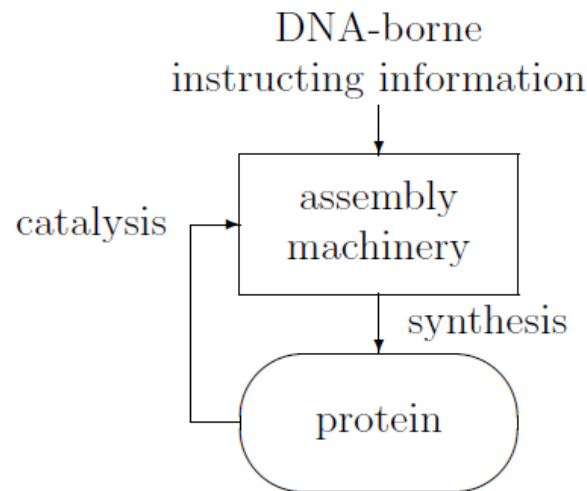
- Nature proceeds with successive generations (which actually imply regenerations).
- Living beings belong to discrete species which, moreover, can be ordered according to a hierarchical taxonomy as a consequence of the nested structure of the genomic error-correcting code.
- Evolution trends towards increasing complexity because longer codes can be more efficient in terms of error correction, hence have been favoured by the Darwinian selection.

## 4. How Information Acts on Matter

Up to now, we have seen how information can be conserved despite the effects of physical perturbations on genomic information-bearing sequences. We now turn on how information acts on matter.

### 1. Basic scheme of a semantic feedback loop

The genes bear the information which instructs the synthesis of proteins. Their transcription and their translation into polypeptidic chains, becoming proteins when properly folded, are controlled by enzymes. As proteins, these enzymes are needed for their own synthesis. We may represent this process as a *semantic feedback loop*. In its most basic form, such a loop is shown in Fig. 2.



**Figure 2.** The genomic information instructs the assembly of a protein needed for this very assembly. As an enzyme, it controls its *own* synthesis, thus locking a loop. The arrows in the figure represent *irreversible* actions, so the whole loop is one-way.

The previous figure depicts the simplest example of semantic feedback loop, assuming that a single enzyme controls the synthesis of a protein. The molecular machinery performs the assembly of a protein according to semantic rules, as instructed by a symbolic information of genomic origin. This protein moreover controls the machinery operation since the instructed assembly can be performed only if it is present.

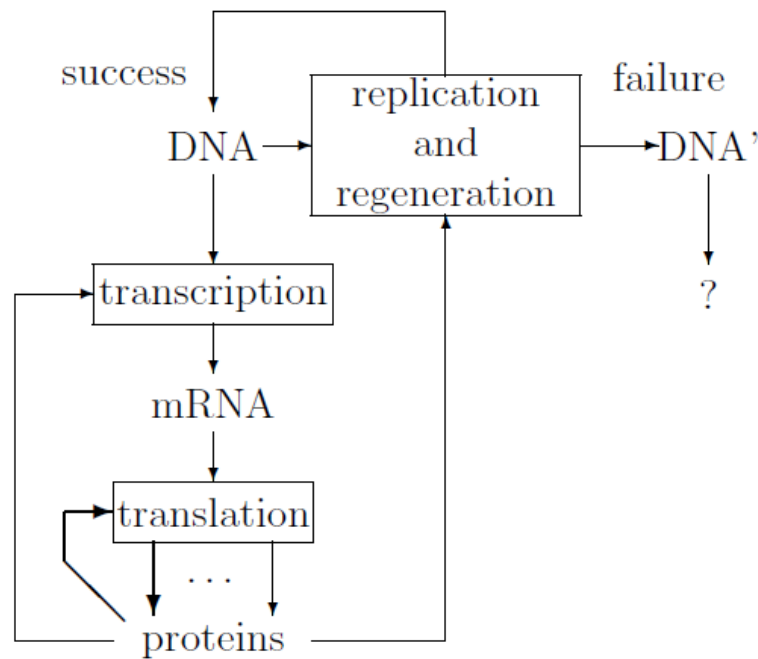
The synthesized protein is then an *enzyme* needed for executing the instructions, which should thus *match* the enzyme specificity. The enzyme just acts as a *key* which enables its *own* assembly. A semantic feedback loop acts as a *trap* since, once assembled, it keeps its structure and conserves the semantic rules it implements. We may think of this property of semantic feedback loops as performing what has been referred to by Crick as a ‘frozen event’. *Semantics* is then *implemented* by the ribosomic *machinery*.

## 2. Interwoven semantic feedback loops

Living structures actually involve a number of combined semantic feedback loops. These loops are *interwoven* in the sense that the assembly of a single protein demands that several functions are performed. Several elementary loops according to the initial basic scheme are thus combined, so that *all* the proteins which act as enzymes for these functions are needed for the synthesis of *any* of them. Instead of a single key as in the initial basic scheme, as many keys as elementary loops now lock the system, which is the more enduring, the more numerous the keys.



As a first example the following figure schematically represents the genetic communication in a prokaryote, which performs the functions of replication-regeneration of the genome and those of transcription and of translation of the genes. The more complicated example of eukaryotic cells will be examined later.



**Figure 3.** Modelling genetic communication in a prokaryote as a system of semantic feedback loops. The arrows originating in ‘proteins’ denote enzymatic actions. The semantic feedback loop which pertains to the genetic ‘code’ (mapping) is drawn in heavy lines (bottom left). DNA’ denotes a DNA string which differs from the original one.

### 3. Reproductive feedback loop

The upper part of the previous figure concerns DNA replication and regeneration. It pertains to each individual in a homogeneous population of cells which descend from a single ancestor and possess the same DNA.

We assumed that the function of regeneration, similarly to others, needs the agency of enzymes. Since the very existence of codes enabling genome regeneration is not recognized by mainstream biology, how regeneration is performed remains unknown and, in particular, the needed enzymes have not been identified. The feedback structure of this ‘reproductive loop’ entails that *constraints on proteins induce constraints on the genome which instructs their own assembly* because any element of the reproductive loop is located both upstream and downstream from any other one, including itself (the words ‘upstream’ and ‘downstream’ are meaningful because the loop is one-way). Any element of the loop acts on the one immediately upstream by

the agency of the downstream ones. This looks like teleology but causality is not violated. It is thus possible to associate with any semantic feedback loop a genomic soft code defined by the specific constraints affecting its enzymes. These constraints are superimposed to the others, and together they result in the genomic error-correcting system of nested codes.

Also notice that the reproductive feedback loop works regardless of the genome size, whose length is thus unlimited.

#### *4. Success or failure of regeneration*

Successful regeneration results in a genome strictly identical to the original one. It is why the replication and regeneration process has been drawn as a loop in the left part of Fig. 3, which is labelled 'success'. Then the genomic information is conserved.

The genome denoted by DNA', obtained in the highly infrequent case of unsuccessful regeneration (indicated by 'failure' in the right part of the figure), markedly differs from the original one as a consequence of the error correction ability of the genomic code. It can replicate itself only if it instructs the assembly of the machineries which are needed to this end despite this difference. Else the whole process is aborted. That the process continues or not is indicated in the figure by the interrogation mark. If it continues, the genome bears a different information and may originate a new species.

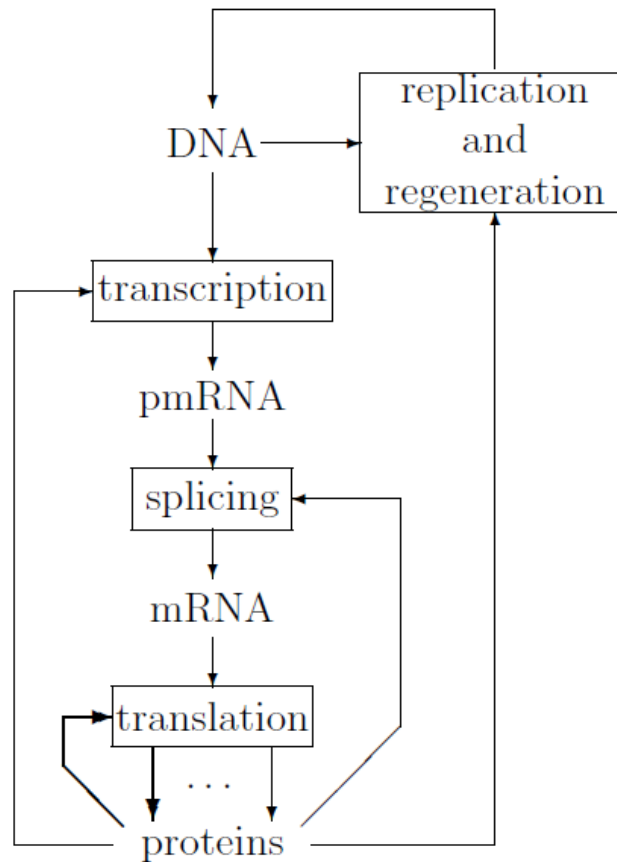
#### *5. New information, lengthened genomes*

The error-correction ability of the genomic code maintains the species integrity. In the very infrequent case of regeneration failure, new information arises. It originates a new species if it still specifies a phenotype and if this phenotype withstands the Darwinian selection. In contrast, the specificity of the enzymes which control the operations of replication, regeneration, transcription, and translation results in the whole process being merely possible. The failure of any of them would abort this process.

Although the system of semantic feedbacks as depicted in the figure is locked, nothing prevents the genome lengthening, e.g., according to 'horizontal genetic transfer' mechanisms. Genome lengthening then results at the same time in increasing its redundancy *and* the information quantity it bears. The former improves the error-correction ability of the genome, and the latter specifies phenotypic features perhaps advantageous with respect to the Darwinian selection.

Lengthening any symbolic sequence can indeed have two effects: increasing the length of its information message hence the information quantity it bears, and increasing its redundancy. One may think of such a sequence of length  $n$  as actually made of the  $k$  symbols of its information message,  $k < n$ , the

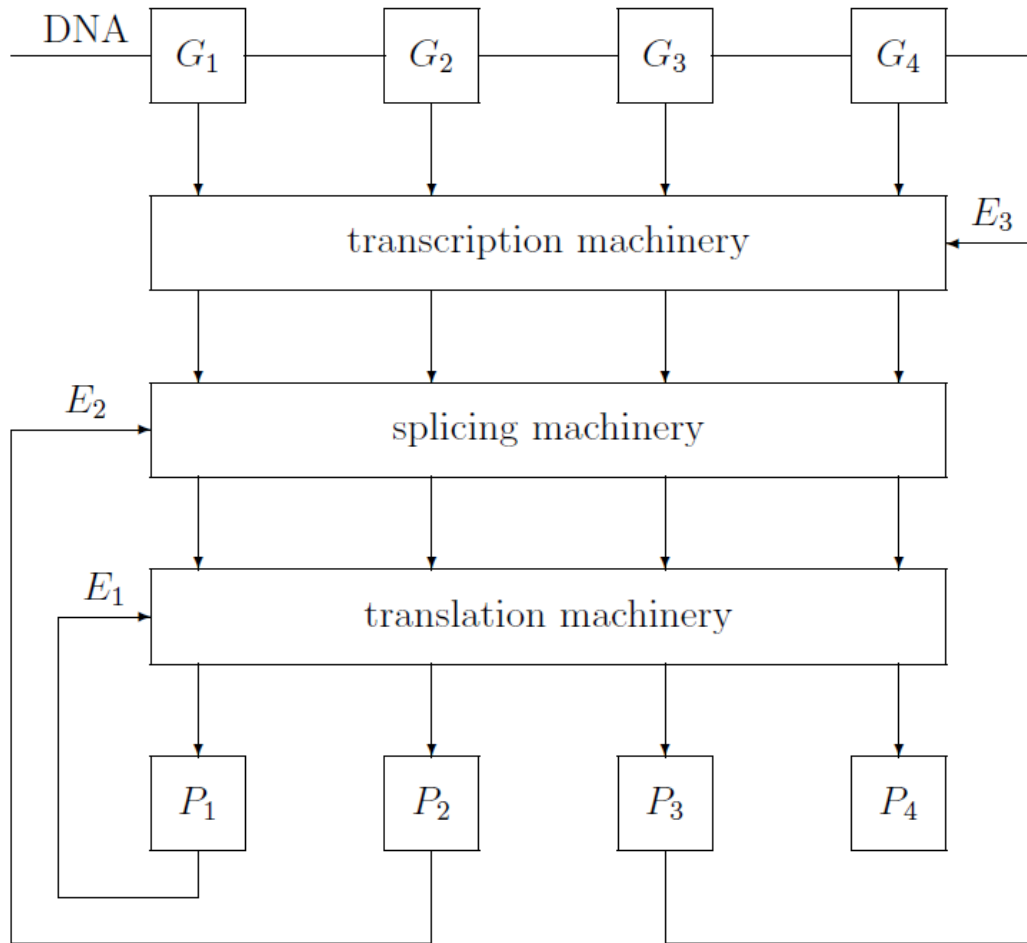
remaining  $n - k$  redundancy symbols being computed in terms of the information message symbols in order to make the whole sequence resilient to errors. Then  $k$  measures the information quantity borne by the sequence, hence its *semantic specificity*. The *resilience* of the sequence to casual errors, which is needed for its conservation, increases as  $n - k$  increases if an adequate encoding is employed. This remains true when the information message is not explicitly present in the sequence.



**Figure 4.** Further including the function of splicing in the system of semantic feedback loops of the previous figure; pmRNA denotes pre-messenger RNA.

### 6. Further inserting the splicing function

Figure 4 shows how the function of splicing, typical of the eukaryotes, has been inserted in the system of semantic feedback loops depicted by Fig. 3 as another loop mutually locked to those already present. Figure 5 below is a scheme of the same loop where its lower part, although still very simplified, is represented with more details. It has been assumed that each of the mentioned functions is controlled by a single enzyme. That several enzymes are actually needed for each function still increases the mutual locking of the semantic feedback loops.



**Figure 5.** System of semantic feedback loops in eukaryotic cells represented with more details, although still very simplified.  $G_1, \dots, G_4$  denote genes which instruct the assembly of proteins  $P_1, \dots, P_4$ .  $E_1, \dots, E_3$  denote the enzymatic actions of the proteins  $P_1, P_2$  and  $P_3$  which enable the functions of translation, splicing and transcription, respectively. The loops are clearly interwoven.

We may distinguish among proteins: (1) the enzymes which are necessary catalysts for the functions implied in the operation of the semantic feedback loops; and (2) other proteins, i.e., enzymes having other functions or mere building blocks. In Fig. 5, emphasis was laid on the enzymes  $P_1, P_2$  and  $P_3$  as belonging to the first category.  $P_4$  was the single representative of the second one. This category actually contains more elements than the first one, namely all the proteins which are used for building the remainder of the phenotype. Not being critically needed, proteins of the second category may incur regeneration errors while their assembly still results in viable phenotypes (later filtered by the Darwinian selection).

### *7. Origin of a new semantic feedback loop*

A new semantic feedback loop results from the insertion in the genome of a gene which instructs the assembly of a protein acting as an enzyme for controlling this very assembly. Such a new gene may result from the erroneous regeneration of a gene already belonging to the genome, or from genetic material being appended to the initial genome by horizontal genetic transfer. In any way, the insertion of a gene which specifies a protein controlling its own assembly is as infrequent as to make the arising of a new semantic feedback loop an exceptional event.

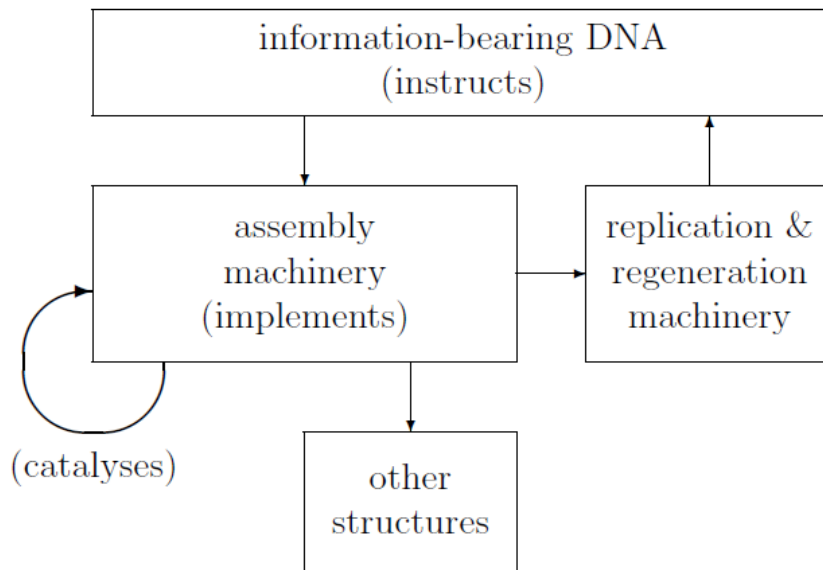
The onset of a new semantic feedback loop may also be interpreted as creating a new organic code in Barbieri's meaning [1,2,3] since it establishes a dependence between two sequences of entirely foreign kinds, one of them being the genome itself. This dependence creates a new constraint in the genome. We may thus identify the onset of an organic code to the introduction of a new component in the genomic error-correcting system of nested codes, which results itself from the onset of a new semantic feedback loop.

## **5. Extended semantic feedback loops, origin and evolution**

### *1. Extended semantic feedback loops*

Up to now, we have seen how semantic feedback loops control the assembly of proteins, as instructed by the genes. But a living thing is not an unorganized cluster of proteins. The molecular machineries of the phenotype which implement the functions of transcription, translation and replication-regeneration, as well as all others, must be assembled as instructed by the genome. Not all details of this process are known, at variance with the fairly well understood scheme which represents the assembly of proteins. Instructing the assembly of phenotypic machineries demands that the whole genome be endowed with some syntax, the constraints of which define some more layers in the system of genomic nested soft codes. The next figure schematically represents a living thing as a system of extended semantic feedback loops.

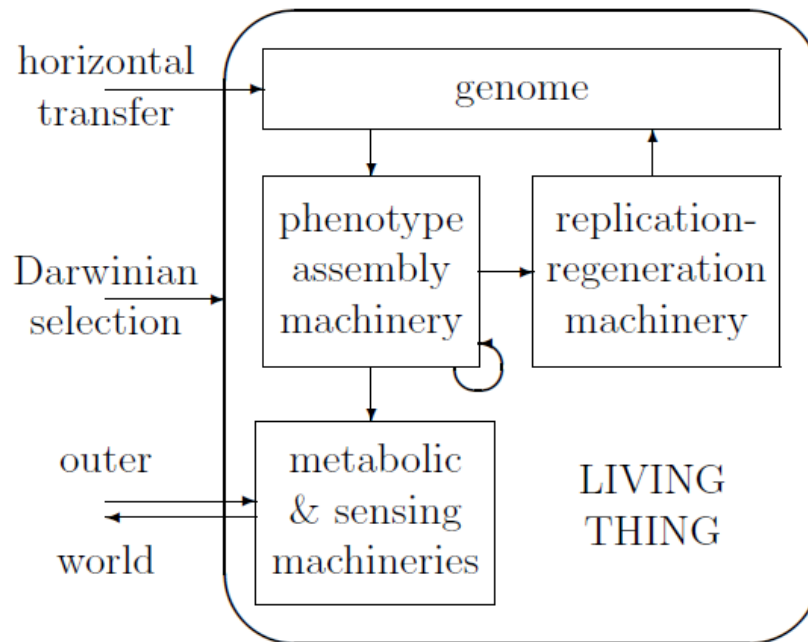
It may reasonably be conjectured that the semantic feedback loop scheme, which obviously operates as regards the synthesis of proteins, also works at higher hierarchical levels. One may thus think that schemes similar to that of Figs 3 and 4 are also valid in devices which implement the assembly of objects which were assembled in previous steps and are more complex than amino-acids. The concept of semantic feedback loop can thus be extended by assuming that it results from the genome specifying not only machineries which implement their own assembly, but also enzymatic keys without which they would not work, according to Fig. 6. These machineries are the more complex, the more complex the objects they assemble. This would explain the permanence of living structures.



**Figure 6.** Extended semantic feedback loops. The assembly machinery involves multiple feedbacks, schematically represented by the curved arrow at left.

## 2. Schematic representation of a living thing

A scheme representing a whole living thing results from appending blocks intended to represent its other functions to the scheme of the previous figure, as shown in Fig. 7. Besides the system of interwoven extended semantic feedback loops, this figure shows the connections of a living thing with its environment. Important parts of it are devoted to matter and energy exchanges with the outer world (metabolism), and the acquisition of information from it (sensing). The horizontal genetic transfer which sporadically occurs tends to increase the genome length. Of course, the living thing as a whole is submitted to the Darwinian selection.



**Figure 7.** A living thing, modelled as interwoven extended semantic feedback loops, within its environment. The information borne by the genome instructs the assembly of all phenotypic machineries, one of which in turn performs its replication and regeneration. Horizontal genetic transfer may increase the information quantity it bears. The phenotype assembly machinery controls the assembly of all phenotypic machineries, including itself. The living thing as a whole is subjected to the Darwinian selection.

### 3. Possible origin and evolution of semantic feedback loops

How the instructions borne by the genome are implemented depends on the assembly machinery in its present state. Can we understand how the semantic feedback loops came into existence?

We may think that, within a mixture of molecules, some of which having memory, i.e., able to bear symbolic sequences like modern DNA or RNA, others with enzymatic properties like modern proteins (and maybe certain having both abilities), a rudimentary semantic feedback loop, as described above and depicted in Fig. 2, has been assembled by chance. Then, once assembled it remains locked and thus conserves itself, at variance with fleeting structures which appear and disappear at random. Further lengthening of the memory may result in specifying other loops interwoven with the first one.

The probability that an initial system of semantic feedback loops is assembled may be extremely small. This event can however occur sooner or later if its probability is not strictly zero.

Maybe a rudimentary semantic specificity and a rudimentary error correction ability were enough for initiating the whole evolution process provided the genome length increases since it can then entail both an increase of information quantity, hence of semantic specificity, and of redundancy. At this early stage, of

course, the genomic error correction system was much less efficient than it is in the present, and similarly the enzymatic specificity was much less. The Darwinian selection can progressively improve both if it operates on increasingly long genomes.

Only the genomes which are the most effectively improved as regards both the semantic specificity and the error correction ability will survive the Darwinian selection. Then the high error correction ability and the sharp enzymatic specificity of modern genomes result from evolution.

#### 4. *Objects cannot be conserved, information can be*

It should be realized that, contrary to our intuition, *conservation* of an object is not the rule but the exception. This is especially true at the geological time scale. As stated by the second law of thermodynamics, the rule is indeed the object's *degradation*. Conserving an object actually needs an *active* approach involving information. We met such an approach when explaining the genome conservation over time intervals at the geological scale by means of an error-correcting code. Even during the much shorter life time of an individual member of a species, the stability of living structures is ensured only by means of semantic feedback loops. In any way, von Neumann has shown by logical arguments that an object can reproduce itself only if it contains its own symbolic description [13,14].

Strictly speaking, then, it is not a particular physical object which is conserved by such means. As regards living things, what can indeed be conserved is an *information*, i.e., a *physically inscribed abstract entity* which represents its composition and enables its re-assembly as frequently as needed. The book by Antoine Danchin describes the underlying biological mechanisms by means of the illuminating metaphor of the Delphic boat [15]. A boat is made of an assembly of planks. Rotten planks need to be changed. The Delphic oracle asked: is the boat the same after all planks have been removed? The answer is yes because what makes the boat to be a boat is how the planks are assembled, irrespective of the planks themselves.

## 6. Conclusion

The information borne by the genome instructs the assembly of all phenotypic structures by the agency of semantic rules. Its conservation is ensured by a system of nested error-correcting soft codes which makes it resilient to symbol errors, provided the regenerations it enables are frequent enough.

As specified by the genome, the phenotypic structures involve interwoven semantic feedback loops which, once assembled, remain locked thus ensuring their own conservation, which entails that of the semantic rules they implement. The locking of these structures does not prevent the genome lengthening which may thus specify new structures and bring more redundancy, hence can result in further evolution.



Any new semantic feedback loop originates a new component in the genomic nested soft code, hence a new organic code in Barbieri's meaning.

Thus, the genomic error-correcting code and the semantic feedback loops, acting together, make life resilient to the intrinsic trend of the physical world towards disorder.

## Acknowledgement

The influence of Marcello Barbieri is gratefully acknowledged. The convergence between my views and his is all the more interesting since our experience and backgrounds are quite different.

## References

- [1]. Barbieri, M., 'The organic codes: the basic mechanism of macroevolution', *Rivista di biologia – Biology forum*, Vol. 80, pp. 101–126, 1998.
- [2]. Barbieri, M., *The organic codes. An introduction to semantic biology*, London: Cambridge University Press, 2003.
- [3]. Barbieri, M., *Code biology*, Dordrecht: Springer, 2015.
- [4]. Battail, G. *Information and life*, Dordrecht: Springer, 2014.
- [5]. Schrödinger, E., *What is life?*, 1943. Reprinted together with *Mind and matter*, London: Cambridge University Press, 1967.
- [6]. Brillouin, L., 'Vie, matière et observation', Paris: Albin Michel, 1959.
- [7]. Shannon, C.E., 'A mathematical theory of communication', *Bell System Technical Journal*, Vol. 27, pp. 370–467 and 623–657, July and Oct. 1948. Reprinted in *Key papers in the development of information theory*, D. Slepian Editor, pp. 5–29, New York: IEEE Press, 1974, and in *Claude Elwood Shannon Collected Papers*, N.J.A. Sloane & A.D. Wyner, Editors, pp. 5–83, New York: IEEE Press, 1993.
- [8]. Gallager, R.G., *Information theory and reliable communication*, New York: Wiley, 1968.
- [9]. Battail, G., "Does information theory explain biological evolution?", *Europhysics Letters*, Vol. 40, No. 3, pp. 343–348, Nov. 1st 1997.
- [10]. Battail, G., 'An engineer's view on genetic information and biological evolution', *Biosystems*, Vol 76, pp. 279–290, 2004.
- [11]. Battail, G., 'Heredity as an encoded communication process', *IEEE Trans. Inf. Theory*, Vol. 56, No. 2, Feb. 2010, pp. 678–687. DOI 10.1109/TIT.2009.2037044

- [12]. McMillan, B., 'The basic theorems of information theory', *Ann. Math. Stat.*, Vol. 24, pp. 196–217, June 1953.  
Reprinted in *Key papers in the development of information theory*, D. Slepian Editor, pp. 57–80, New York: IEEE Press, 1974.
- [13]. von Neumann, J., *Theory of self-reproducing automata*, edited and completed by A.W. Burks, Urbana: University of Illinois Press, 1966.
- [14]. Pattee, H.H., & Raczaszek-Leonardi, J., *Laws, language and life. Howard Pattee's classic papers on the physics of symbols with contemporary commentary*, Dordrecht: Springer 2012.
- [15]. Danchin, A., *La barque de Delphes*, Paris: Odile Jacob, 1998. *The Delphic boat*, English translation by Alison Quayle, Harvard University Press 2001.