HISTORY AND PHILOSOPHY OF THE LIFE SCIENCES
25(2003), pp. 257-275

Giovanni Boniolo
Dipartimento di Filosofia
Università di Padova
Piazza Capitaniato, 3
35100 – Padova - Italy
giovanni.boniolo@unipd.it

BIOLOGY WITHOUT INFORMATION

1. *Introduction*

Over these last few years once again the relationship between biology and information has been debated with great liveliness. The crucial points concern the meaning of the term 'information' and whether the so-called "information talk" is really necessary inside biology. In what follows, I will proceed by first commenting on some points of the debate (§ 2), then showing that a bio-physical account of the process from the nucleotide sequences to the correlated amino acid sequences is possible (§ 3). In this way, I will suggest that a satisfying account of that process can be offered without entering the quicksand of information.

With reference to the debate, it is worth emphasising that I place myself in the faction, just to quote some paradigmatic examples, of Sarkar (1996); Manher and Bunge (1997); Kitcher (2001); but also of Chargaff (1963), that is, in the faction of those who deny, even if starting from different presuppositions, any real philosophical and explicative value to the notion of information in contemporary molecular genetics. In particular, Sarkar (1996) points out that there is a sort of *aut-aut*. *Aut* we return to the old notion of specificity described in terms of new physics, i.e., Quantum Mechanics (but Sarkar (2000) seems to confess that he has no idea about what to do); *aut* we must look for a new account of information. In § 3, I will argue for the first horn of the disjunction and therefore against the second one. Meanwhile I will show how Quantum Mechanics is really relevant.

It should be noted I sustain neither that the notion of information had no relevant historical and sociological role in the genesis of molecular genetics, nor that it has no didactical value in teaching molecular genetics. I simply claim that *now* to speak in terms of information is no longer necessary, though everyone should be free to follows the path he prefers, especially if it is consistent, unambiguous, and well argued.

2. *Some weaknesses of the debate on biology and information*

2.1. *A note on philosophy of information and biology*

The philosophy of information might be considered as a new area of philosophical research "concerned with a) the critical investigation of the conceptual nature and basis principles of information, including its dynamics (especially computation and information flow), utilisation and sciences, and with b) the elaboration of information-theoretic and computational methodologies and their application to philosophical problems". In this way, Floridi (2002) defines such a domain of philosophical analysis. From his survey of the open problems in the philosophy of information, it appears clear that the main one is: "what is information?". In trying to circumscribe it, he writes

"Information can be viewed from three perspectives: information as reality (e.g., as patterns of physical signals which are neither true nor false), also known as *ecological* information; information about reality (semantic information, alethically qualifiable); and information for reality (instruction, like genetic information)" (Floridi, 2003a; cf. also 2003b).

Floridi indicates that there are at least six extensionalist tentative approaches to the definition of information. Of the six extensionalist approaches available on the market place, one is syntactical (i.e., Shannon, 1948), the other ones are semantical (Bar-Hillel and Carnap, 1953; Dretske, 1981; Barwise and Perry, 1983; Israel and Perry, 1990; Devlin, 1991; Floridi, 2003a).

It is quite irrelevant whether Floridi's survey is complete or not, but the conclusion he draws is not irrelevant: we have many semantic theories of information and none of them seems to be really good at explaining "information for reality", i.e., the kind of information concerning, for example, molecular genetics.

2.2. *Some remarks on the debate*

To begin with, let us note that if we try to answer our question - "Is information relevant for biology?" - without clarifying what we mean by the terms contained in it, we easily fall victims of our opponents. The latter can readily object, for example, that the problem is not well-posed and that we are making fallacies like ambiguity and amphiboly.

However it seems that by information we should mean exactly what Crick (1958) meant:

"By information I mean the specification of the amino acid sequence of the protein [from the nucleotide sequence of the DNA]" (p. 144)

If we glance at the contemporary debate, we realise that all the discussants agree with the fact that the syntactical approach to information is not able to grasp Crick's definition.[1] Perhaps the semantic ones could do it. But, none of the discussants has analysed with the necessary profundity the semantic theories on information now available. On the contrary, most of them (for example, Godfrey-Smith, 2000b; Sarkar, 2000; Sterenly, 2000) begin directly with their own notion of semantic information, without showing why they do not accept the other ones (to be honest, some of them quote, but very cursorily, 1981 Dretske's theory).

Another aspect must be noted. Showing the philosophical relevance of the problem does not mean, as also Griffiths (2001) points out, limiting ourselves to stressing that "The colloquial use of informational terms is all-pervasive in molecular biology. Transcription, translation, code, redundancy, synonymous, messenger, editing, proofreading, library – these are all technical terms in biology […] Molecular biologists, then, do make use of informational analogy in their daily work […]" (Maynard Smith, 2000a, p. 178). Affirming that something is philosophically relevant only because people speak about it (or in terms of it) means committing a fallacy (the so-called *argumentum ad populum*), as we know.

Let us move to another point. Maynard Smith (2000a, p. 193) writes that "genes carry information". Unfortunately genes "do not carry anything". At most *we interpret them as carriers of information*. In the contemporary philosophy of science the idea that any datum and any observation are theory-laden is widely accepted (cf. Boniolo and Vidali, 1999). Therefore locution such as "genes carry information" must be intended only as a popular and metaphorical way of saying that we, knowing subjects, interpret genes as carriers of information.[2]

Nevertheless if we accept that the genes are laden by the theory of information we have chosen, we should concede that they could be laden also by different theories of information or by different approaches. For example, by a semiotic approach. In this case, we might follow Ch. S. Peirce's proposal and consider a gene as a sign (that is, as an icon, as an index and as a symbol) with a given interpretant made by a given interpreter. Or we might prefer a symbolic approach *à la* E. Cassirer. Or, if we were continentally-minded philosophers, we might suggest a hermeneutical *à la* H.G. Gadamer interpretation of the genes, that is, as a text to be read in a context. Everything depends on our philosophical taste and on our ability to justify an approach instead of another one. But it also depends on the cultural fashion. And now the information approach seems more up-to-date than the Peircean approach, the Cassirean approach, the Gadamerian approach (cf. Griffith, 2000, p. 395).

What has just been observed leads us to another note. If, for example, we took the content of Jablonka (2002), which is written in the "information language", we might easily and *mutatis mutandis* "translate" it into the "Peircean semiotic language", or into the "Cassirean symbolic language", or into the "Gadamerian hermeneutical language". Let us try to do it. It is a good sceptical lesson as to the "language" by means of which we write our papers. Or, at least, it is a good exercise for our students who, in this way, learn how and to what extent some languages are interchangeable.

Let us move to a last aspect of the contemporary debate: the one concerning the use of analogies in science. As is well known, this is a topic that began to be analysed in the 1960's and continued to be studied through the years intertwined with that of models (cf. Boniolo, 2003). After Hesse (1966), it was extremely clear that analogies may be useful both in constructing theories and in explaining aspects of the theories. Two different uses that we find in the debate on biology and information, even if without the old terminology. For example, Sarkar (2000, p. 209) speaks about the heuristic and the substantial uses of the concept of information.[3] Nevertheless anytime we work with analogies both in the discovery context and in the explanatory context, we must pay attention to the fact that some are positive and some are negative. Only the former have a heuristic or explanatory relevance. If we do not take into account this fact we run the risk of falling into the fallacy of the false analogy, as some discussants emphasised (for example, Quastler, 1958; Maynard Smith, 2000, p. 179). However such attention is worthless if not accompanied by a profound analysis of the notion of information. How can I individuate the positive and the negative analogies, and therefore how can I avoid falling into the fallacy of the false analogy, if I do not have any precise and unambiguous understanding of the notion of information?

3. *Biology without information*

After the above remarks on the debate, it is the time to ask ourselves: "Is an informational approach to the gene expression really necessary?".

As said, before answering I should disambiguate the notion of information. Actually I will not do so. If I am capable of giving an account of the gene expression without the notion of information, it follows that information is not necessary (and therefore I do not need to define it precisely and unambiguously). Of course, this should also be a blow to the supporters of the information talk in biology.

To reach this aim I will argue for the following thesis:

"*To give a correct and comprehensive analysis of the process which leads to the synthesis of the amino acid sequences a purely bio-physical approach is sufficien*t".

To justify such a thesis I will proceed along the following lines:

1) *first*, I will recall some well known facts about the structure of the genotype and the gene expression;
2) *second*, I will consider a probabilistic causal approach of the intracellular molecular interactions.[4]

In this way, I will show that it is possible to offer a purely probabilistic causal account, based on bio-physics, of the process which starts from the nucleotide sequences and arrives at the synthesis of the amino acid sequences.

3.1. *Complexity of the genotype structure and gene expression*

The content of this section should not even be considered, since any biologist and any philosopher of biology is more than familiar with it. It is, indeed, sufficient to have studied one good textbook (for example, Alberts *et al.*, 1994 and Stryer, 1995) to have a clear idea of the complexity of the genotype structure (introns, exons, satellite DNA, interspersed repeated DNA, transposable DNA elements, etc.; structural genes, regulatory genes, mutator genes, etc.) and of the

complexity of the gene expression. Two complexities which were very far from being comprehended by Crick and the coeval colleagues in the Fifties of the last century and that we ourselves have not yet totally mastered.

However, for the sake of self-consistence, let us consider the gene expression. There are at least six steps to arrive at the amino acid sequence of a protein:

1) the pre-transcription with the formation of a particular chromatin structure at the level of a particular DNA sequence;
2) the transcription of this DNA sequence to obtain the primary RNA transcript;
3) the formation of mRNA thanks to the splicing of the introns (if we are considering eukaryotes);
4) the transport of the mRNA out of the nucleus;
5) the translation of the mRNA at the level of the ribosomes and the consequent synthesis of the correlated amino acid sequence;
6) the definitive maturation of the amino acid sequence.

We know that each of these steps is controlled in very sophisticated way, but if only one is badly regulated, and this may probabilistically happen by taking into account the inevitable metabolic errors, the protein does not have the correct sequence and its function may be also drastically altered, as to the function of the protein with the correct sequence.

It is not necessary here to dwell upon each step, but something is worth recalling:

1) The gene expression is regulated by certain proteins which, depending on the type of the living being, in different times, in different tissues, and at different stages of the development, allow the beginning of the transcription. Of course the correct expression occurs only if the promoter sequence of the gene has not suffered any limiting alteration, if the transcription factors (which, in the case of eukaryotes, assemble with the polymerase and together bind to the gene control region) have the correct structure, if the gene regulatory proteins, which bind to the various regulatory sequences, have the correct structure; that is, if both the transcription factors and the regulatory proteins are the final result of a correct synthesis of non-modified genes.
2) The real transcription, that is, the synthesis of the primary RNA sequence, is realised by the RNA polymerases. Also in this case, as happens in the case of DNA duplication, there can be errors due to chemical agents or due to thermal fluctuations. However, the errors in case of transcription are less relevant compared with the errors in case of duplication. Here the errors may be lethal for the living being. In the former case the errors may be totally irrelevant since, even if a primary RNA molecule were transcripted in a wrong way, it would be only one of the many produced by the RNA polymerase. Let us recall that from the same active gene, thousands of primary RNA transcripts can be synthesised and from the same mRNA molecule thousands of copies of the amino acid sequences can be produced. For example, in each silk gland cell a single fibroin gene makes 10^4 copies of mRNA and each of them allows the synthesis of almost 10^5 proteins in 4 days (Alberts *et al.*, 1994, p. 105).
3) The splicing of the introns from the primary RNA transcript, so as to produce the mRNA, can happen by a sort of self splicing or can be catalysed by means of the spliceosomes. We know that there could be, in some cells, the alternative splicing and therefore from the same nucleotide sequence of primary RNA we can have the synthesis of more than one amino acid sequence. Nevertheless there could also be pathological splicing due to metabolic errors, or to wrong regulations.
4) The translation of the mRNA and the synthesis of the correspondent primary structure of the proteins occur in the ribosomes thanks to tRNA molecules which transport the right amino acids. However, there can be errors both in the base-pairing between mRNA codon and tRNA anticodon, and in the base-pairing between tRNA and the

amino acid. In both cases, the cell has evolved correction mechanisms. However, as always happens, errors "may escape".

In conclusion, by observing the standard amino acid sequences and the wrong amino acid sequences, we can affirm that the error probability in the amino acid synthesis is equal to 10^{-4}: low but not null (cf. Thompson, 1988; Soll 1990).

By taking all of this into account, Crick's claim ("By information I mean the specification of the amino acid sequence of the protein [from the nucleotide sequence of the DNA]") may be still kept, provided that all the steps of the protein synthesis work correctly; or provided that all the control mechanisms work correctly.

Unfortunately nobody can guarantee that all of this happens in any case. Therefore we are not allowed to claim that we can *exactly predict* the amino acid sequence from a given nucleotide sequence. We are allowed to claim that we can predict it with a probability approximately equal to $(1 - 10^{-4})$.[5] And, of course, this result does not consider other complexities in the gene expression such as the *trans*RNA splicing and the RNA editing.

I know, as Maynard Smith (2000a, p. 184) recalls, that biologists predict the amino acid sequences from the nucleotide sequence "all the time", but it should be quite clear that this apparent determinism actually is a phenomenological result due to a large number of instantiations and a low error probability.

3.2. *Considering Molecular Dynamics*

The cell, besides being considered as a biological system, can be considered as a physical system in which elements collide in various ways and because of many forces.[6] Surely, the most important forces in the cellular molecular dynamics are the electromagnetic ones. We have electrostatic forces (many molecules have a charge) and electrodynamic forces (remember that molecules in motion imply charges in motion and therefore electric fields in motion, and these had to do with the so-called Van der Waals forces). But we also have some non-fundamental (from the physical point of view) forces. I am thinking about the positional forces, such as those due to the hydrophobic groups or the steric repulsion connected with the three-dimensional configuration of the molecules and therefore with their degrees of freedom.

However, let us pause on the electromagnetic forces. They are the cause of the covalent bonds and of the non covalent bonds (ionic bonds, hydrogen bonds, bonds due to Van der Walls forces and indirectly to hydrophobic groups, which are held together by electromagnetic forces). Yet, so that two molecules electromagnetically bind, they have to interact, rather they have to interact in a correct way. I would like to stress that here there are two issues: the first one concerns the interaction; the second one concerns the correct interaction.

So that two molecules can reach such a distance that the electromagnetic forces are effective, they have to be pushed in some way one against the other. We know that inside the cell there is, beyond the osmotic pressure from the outside, a random thermal motion. It is exactly the thermal energy which "keeps alive" this thermal motion that has to be considered the first cause of the probabilistic causal chains that happen inside the cell and which has, as one of its products, the amino acid sequences of the protein.

As stated previously, it is not sufficient that two molecules interact, they have to interact correctly. What does this mean? We know that the molecular chains are kept together by covalent bonds and that two parts of the same molecule or two different molecules can be bound by non covalent bonds. However, we know that a non covalent bond, since it is from 10 to 10^2 weaker than a covalent bond – especially in aqueous solution as there is inside the cell –, is not sufficient to keep bound the two sites of the two molecules (in the case of non covalent bonds, the binding energy is 1÷7 Kcal/mole; in the case of covalent bonds, the binding energy is 50÷100 Kcal/mole). Not only does the thermal motion allow the interactions among the molecules, but it also causes difficulties in the bonds due to both the thermal agitation of the molecules considered in themselves and the possible pushes caused by other molecules (at room temperature, the thermal energy of a molecule

is about 0.6 Kcal/mole). It follows that between two molecules there is a sufficient force to keep them bound if between the two respective binding sites a sufficient number of non covalent bonds are realised. This happens if the two binding sites have a particular suitable three-dimensional atomic configuration. In other words, the non covalent bonds between two molecules are sufficient, if the stereospecificity of the two binding sites is complementary, that is, if they match well. But why does a binding site of a gene regulatory protein exactly match with the binding sites of a particular nucleotides sequences of the promoter? For example, why, in yeast, does the protein MATα2 stick to the DNA sequence CATGTAATT? Why does a binding site of a particular enzyme exactly match with a binding site of a particular substrate? Why, for example, does $tRNA^{Trp}$ exactly match with the aminoacyl $tRNA^{Trp}$ synthetase? And why, immediately after, does this complex bind with tryptofan? Why does the $tRNA^{Trp}$ anticodon (ACC) exactly match with the UGG codon? There is no answer, except that given us by an evolutionary approach. This is the reason why Monod (1970) spoke about *gratuité* (on this term, cf. Appendix). The specificity of the biological recognition between two molecules inside the cellular nucleus or inside the cellular cytoplasm is a totally contingent evolutionary result, of course within the chemical constrains (cf. Knight *et al.*, 1999).

It is this play between the causality due to the forces involved and the two casualities (one concerned with the randomness of the thermal agitation and with the non predictable casual interactions so produced, and the other one with the evolutionary contingency of the specificity of the involved binding sites) that is at the basis of the synthesis of the amino acid sequences from the nucleotide sequences (with reference to the intracellular molecular dynamics, cf. Berg, 1983; Kirkwood *et al.*, 1986; Karplus *et al.*, 1990).

Let us try to be more formal and let us consider a generic intracellular molecular interaction. Let M be a molecule of a given type and let us suppose that there are m molecules of such a type. Inside the cell we have m^1, …, m^n molecules of, respectively, the types M^1, …, M^n. In such a case, p_{M/M^i} is the probability that a molecule of the type M interacts with an arbitrary molecule of the type M^i . Of course the probability strongly depends on the concentration of the two molecules and on the strength of their thermal motion. Just to have an idea of the enormous number of intracellular collisions, think that if there is the standard concentration of ATP (about 1mM), each site of a given protein will interact more or less 10^6 times per second with ATP molecules.

Yet we said that it is not sufficient that an interaction occurs; it has to be a good one, that is, it has to produce a bond. For this to happen a first event has to take place: the two interacting molecules have to have a sufficient kinetic energy, in particular a kinetic energy capable of overcoming the two repulsive potential barriers of their interacting sites. Of course this kinetic energy depends on the strength of the thermal motion, that is, on the temperature T. Let p_T be this probability.

There is a last step. This concerns the fact that the two molecules, which have interacted and whose kinetic energy allows the overcoming of the potential barriers, actually have to have the two binding sites 3-dimensionally conformed in such a way to permit a correct and stable sticking. Let p_C be the property that the two interacting molecules have the correct 3-dimensional configuration.

It follows that the probability p_B that a molecule M generates a bond with a certain other molecule M^i is

$$P_B = p_{M/M^i} \cdot p_T \cdot p_C$$

It may happen, and it actually happens, that the interaction between two molecules is mediated by other molecules: the enzymes. In this case the situation changes and we have to take into account both the probability of the interaction between the enzyme and the first substrate ($p_{e/s1}$) and the probability that this is a good (in the sense just considered) interaction between them (p_{e+s1}). Moreover we have to consider the probability of the interaction between the complex (enzyme + first substrate) and the second substrate, $p_{(e+s1)/s2}$, and, finally the probability that this is a good interaction. In conclusion, the probability of the bond between the two molecules is

$$p_B = p_{e/s1} \cdot p_{e+s1} \cdot p_{(e+s1)/s2} \cdot p_{(e+s1)+s2}$$

This extremely simple model points out that everything that happens inside a cell has to be considered both probabilistically and causally, that is, in a probabilistic causal way. It is remarkable that the causal events are all reducible to physical events, while the non causal events are either reducible to physical events (for example, the randomness of the thermal motion) or irreducible to physical events (for example, the evolutionary contingency of the complementarity between two binding sites).[7]

However, let us dwell for a while on the synthesis of the primary RNA transcript. We know that one of the two DNA strands works like a template and that the synthesis is realised by a particular enzyme, the RNA polymerase. This wanders randomly and, always randomly, collides with the chromosomes. Only if it finds the right DNA sequence, that is, if the gene regulatory proteins and the transcription factors permit it, does it bind with the DNA. It then opens the double helix and exposes the template strand to the random collisions of ribonucleoside triphosphate molecules. If one of these is the right one, that is, if it base-pairs well with the exposed DNA nucleotides, there is the beginning, or the continuation in the 5'-to-3' direction, of the primary RNA chain. At a certain point, the RNA polymerase stops synthesising the primary RNA chain.

We know that there is a particular DNA sequence where the RNA polymerase binds and that there is a particular DNA sequence where the RNA polymerase stops. For example, in bacteria, as starting sequences there are the so-called "–10 region" composed by TATAAT and the so-called "–35 region" composed of TTGACA. With reference to the stop sequence, we know that, for example in *Escherichia coli*, the synthesis stops when the RNA polymerase has synthesised a UUU sequence after that it has synthesised a self-complementary RNA sequence forming a sort of hairpin.

We might say that both the starting sequence and the stop sequence are signals of, respectively, the beginning and the end of the synthesis of the primary RNA transcript. But this language is not at all necessary and it may be carefully avoided. Everything, beyond the evolutionary contingent matching of the binding sites, can be explained simply by resorting to pure molecular genetics, or, if we want a finer explanation, to bio-physics.

I wish to pause upon this point for a while. *I am not preaching or pleading a sort of physical reductionism*, since, as I said, there is at least one physically irreducible aspect, that is, the evolutionary contingent matching of the binding sites. I am simply claiming that *in this case* (I repeat: *"in this case"*) reduction works well. *In this case* we have a good bio-physical explanation of the process from the nucleotide sequences to the correlated amino acid sequences; and that's all.

Let us consider, to stress this aspect even more, what happens when two binding sites match. We know that in this case there is a many-atom system in play, that is, a many-nuclei plus many-electrons system. Whenever we meet systems like this, Quantum Mechanics enters the scene. This means that we have to deal with the time-independent Schrödinger equation:

$$Hy = Ey$$

where H is the Hamiltonian of the system, E is the energy and y is the wave function.

In our case, that is, in the case of many-atom system, the Hamiltonian, in a semiclassical approximation, is made up of the kinetic energy of the electrons and of interaction energy of the electrons with the nuclei and among themselves, that is,

$$H = \sum_i \frac{-\nabla_i^2}{2} + \sum_{i,I} \frac{-Z_I}{|r_i - R_I|} + \frac{1}{2} \sum_{i \neq j} \frac{1}{|r_i - r_j|},$$

(to simplify we use atomic units: $\hbar = e = m = 1$; the index I runs over the nuclei and the indexes i and j run over the electrons). If E is known for a given nuclear configuration, therefore we can find the corresponding potential energy for the nuclei. However, the difficulty is to find out the expression of the electronic wavefunction $y = y(r_1, ..., r_n)$, since it depends on the spatial coordinates of the n electrons in question.

This is the main problem for any attempt to give a physical account of the matching between two binding sites of two molecules (DNA sequence, RNA sequence, gene regulatory proteins, enzymes, …, that is, any biological molecules). From the 1960's on many efforts have been made to

try to solve this many-body problem. A many-body problem can be tackled neither *via* analytic and deterministic tools (we cannot deal with more than two bodies), nor *via* statistical tools (the bodies are not enough); therefore some other solutions has to be found. Actually these have been found and they are numeric solutions. Thanks to the progress in the algorithms theories and to the increase of computer power, we are able to simulate what happens in the binding sites of the biological molecules and so to give a good model, based on Quantum Mechanics, of the stereospecificity.

To be honest we are only at the beginning of this field of research, but it promises to be extremely fruitful for a more profound inquiry and understanding of the bio-chemical bonds.[8] However, what is interesting for us is that we can use again "the old notion of specificity described in terms of new physics" and so Sarkar's 1966 desire seems to be satisfied.

4. *Conclusions*

If what has been said holds, it is possible to provide a probabilistic causal account, based on bio-physics, of the process which leads to the synthesis of the amino acid sequences . Of course *this does not imply a strong global physical reductionism, but simply a physical reduction of what can be physically reduced plus the awareness that not everything can be physically reduced*. For example, the reasons why there is a good matching between certain two binding sites cannot be physically reduced, but it must be studied and explained *via* evolutionary biology. Nevertheless, stereospecificity can be physically explained and modellised. I believe that a *global form of physical reductionism* is wrong, but I am sure, as shown, that a *local form of physical reductionism* works.

All of this leads us to conclude that, on the one hand, we have a precise bio-physical way of grasping the gene expression and, on the other hand, the information talk is not necessary. Therefore, at least in this case, *biology without information* can be done. Naturally this means neither that one must necessary abandon the latter, nor that the information talk could not have some value at different biological levels.

5. *Appendix on the term* gratuité

We know that philosophy is not made by playing with the meaning of the terms, but we know also that philosophy has to pay great attention to the meaning of the terms. This implies that philosophers should be well-aware of the etymon of the terms, of course without falling into the rather comic fallacy of the false etymon.

Maynard Smith (2000a, p. 185) suggests that the term *gratuité* introduced by Monod (1970) has been "translated, not happily, as gratuity" and proposes, it seems followed also by Godfrey-Smith (2000, pp. 203-204), that "it is the arbitrary nature of molecular biology that Monod calls gratuity" and therefore that "it is a pity that he [Monod] used the term 'gratuité' rather than 'symbolic' to describe the essential arbitrary nature of such a signal" (Maynard Smith, 2000b, p. 216).

Is he right? Let us consider the terms *gratuitous* and *arbitrary*. In French there are two terms: *gratuit* and *arbitraire*. The same terms are present, for example, also in Spanish (*gratuito* and *arbitrario*), and in Italian (*gratuito* and *arbitrario*). This is not surprisingly since French, Spanish and Italian are neo-Latin languages. What is interesting is that also in English we have the persistence of the same words, i.e., as said, *gratuitous* and *arbitrary*.

Unfortunately, but it could not be otherwise, the two terms have different meanings which are correlated with the two different Latin roots: 1) 'gratuitous', 'gratuit', 'gratuito' derive from the Latin term *gratus* and therefore from *gratia*; 2) 'arbitrary', 'arbitraire', 'arbitrario' derive from the Latin term *arbitrarius* and therefore from *arbiter*. This difference in the Latin implies an obvious difference in their contemporary meaning.

Consider now a Latin dictionary (for example, Lewis, 1879), we read

Arbitrarius means "done by way of arbitration", "uncertain, not sure" exactly because is under the *arbiter*'s judgement, and therefore somehow subjective to arbiter's will.

[…]
Gratuitus means “done without pay, reward or profit” or “for no particular reason”; that is, *gratis* or “out of favour or kindness”.

These meanings are, of course, considered in the *Oxford English Dictionary*.

Arbitrary [ad. L. *arbitrarius*, f. *abiter*] … Dependent upon will or pleasure
[…]
Gratuitous [f. L *gratuitus* (cogn. w. *gratia*, *gratus*) + ous] … Freely bestowed or obtained; granted without claim or merit; costing nothing to the recipient; free

Already from this short etymological analysis it follows that *gratuit* and *arbitraire* are two terms whose meaning is totally different, and the same happens for the two correspondent English words 'gratuitous' and 'arbitrary'. As a consequence, to demand to translate *gratuit* with ‘arbitrary’ instead of with ‘gratuitous’ is totally wrong.

Moreover, we may ask ourselves: “Was Monod aware of the difference between *gratuit* and *arbitraire*?” I think he was, even if, to be honest I do not know at which degree. What is sure is that he wanted to use the term *gratuité* for cultural reasons. His *Le hasard et la nécessité* was not simply a popularisation book, it was something more. It was also a product of the cultural climate in which it was written, that is, it was strongly embedded in French existentialism. As proof, we may recall the epigraph that opens the book: it is a quotation from *Le mythe de Sisyphe. Essai sur l’absurde* written in 1942 by A. Camus. Moreover, the closing words of the book are extremely similar to Meursault’s last thoughts in 1942 Camus’s *L’étranger*. And it should be noted that Camus was a good friend of Monod.

These are not only curiosities around the book but clues that the content is to be read in a particular light, that is, in the light of existentialism. And according to French existentialism some acts or phenomena are exactly “Freely bestowed or obtained; granted without claim or merit; costing nothing to the recipient; free”, that is, as the entry of the dictionary, they are gratuitous. By the way, exactly this word compares a lot of times in existentialist writings, for example in J.-P. Sartre, *L’être et le néant* (1943).

Therefore we can conclude that Monod with the term *gratuité* did not want only to mean that there are no physical-chemical reasons for the matching of the binding sites, but something more, something concerning existentialist meditations on the meaning of human existence and human structure.

Do we prefer to abandon Monod’s existential interpretation of that bio-physical fact? We can, but unfortunately not by translating *gratuit* with ‘arbitrary’, since there is no arbiter’s will (and no arbiter at all) entering the play (unless we do not want to argue for a divine arbiter and for a divine arbiter’s design). If we wish to eliminate the existential taste and if we prefer a philosophically more neutral term, we can. It is sufficient to adopt the modal term ‘contingent’ (i.e., the contradictory term of ‘necessary’ and the sub-contrary term of ‘possible’, according to the modal square). That is, whenever we want to speak of the bio-physical independence of the function of a molecule from its structure and we want to avoid any existentialist taste, we may speak of contingency (even if, to be honest, we should be well aware that this term, of course with another philosophical taste, was used also by Sartre).

6. ***Acknowledgement***

I would like to thank L. Floridi and P. Vezzoni for fruitful comments on earlier versions of this paper and the unknown referees for their suggestions and critics.

References

Ageno, M. (1989), *Le radici della biologia*. Milano: Feltrinelli.

Alberts, B., Bray, D., Lewis, J. Raff, M., Roberts, K. and Watson, J.D., (1994[3]), *Molecular Biology of the Cell*. New York: Garland Publishing.
Atlan, H. (1972), *L'organisation biologique et la theorie de l'information*. Paris: Hermann.
Bar-Hillel, Y. and Carnap. R. (1953), "An Outline of a Theory of Semantic Information", in Y. Bar-Hillel, *Language and Information*. Reading (Mass.): Addison Wesley, 221-274.
Barwise, J. and Perry, J. (1983), *Situation and Attitudes*. Cambridge (Mass.): MIT Press.
Berg, H.C. (1983), *Random Walk in Biology*. Princeton: Princeton University Press.
Boniolo, G. (2003), "Theories and models: really old hat?", *Yearbook of the Artificial*. Bern: Peter Lang Academic Publishing Company, Issue II, forthcoming.
Boniolo, G. and Vidali, P. (1999), *Filosofia della scienza*. Milano: Bruno Mondadori.
Brillouin, L. (1956), *Science and Information Theory*. New York: Academic Press Inc.
Car, R. (2002), "Introduction to Density-Functional Theory and ab-Initio Molecular Dynamics", *Quant. Struct.-Act. Relat.* 21: 97-104.
Car, R., Parrinello, M. (1985), "Unified Approach for Density Functional Theory and Molecular Dynamics", *Phys. Rev. Lett.* 55: 2471-2474.
Carloni, P., Rothlisberger, U. and Parrinello, M. (2001), "The Role and Perspective of Ab Inizio Molecular Dynamics in the Study of Biological Systems", *Acc. Chem. Res.* 35: 455-464.
Chargaff, E. (1963), *Essays on Nucleic Acids*. Amsterdam: Elsevier.
Crick, F.H.C. (1958), "On Protein Synthesis", in *Symposia of the Society for Experimental Biology*. New York: Academic Press Inc., 138-163.
Devlin, K. (1990), *Logic and Information*. Cambridge: Cambridge University Press.
Dretske, F. (1981), *Knowledge and the Flow of Information*. Cambridge (Mass.): The MIT Press.
Fantini, B. (1988), "Utilisation par la génétique moléculaire du vocabulaire de la théorie de l'information", in M.Groult, J. Roger (eds.), *Transfert de vocabulaire dans la sciences*. Paris: Ed. du CNRS, 157-170.
Floridi, L. (2002), "What is the Philosophy of Information?", *Metaphilosophy* 33: 123-145.
Floridi, L. (2003a), "Information", in L. Floridi (ed.), *The Blackwell Guide to the Philosophy of Computing and Information*. Oxford: Blackwell.
Floridi, L. (2003b), "Open Problems in the Philosophy of Information", *Metaphilosophy*, forthcoming.
Gatlin, L.L. (1972), *Information Theory and the Living System*. New York: Columbia University Press.
Godfrey-Smith, P. (2000a), "On the Theoretical Role of 'Genetic Coding'", *Philosophy of Science* 67: 26-44.
Godfrey-Smith, P. (2000b), "Information, Arbitrariness, and Selection: Comments on Maynard-Smith", *Philosophy of Science* 67: 202-207.
Griffiths, P.E. (2001), "Genetic Information: A Metaphor in Search of a Theory", *Philosophy of Science* 68: 394-412.
Hesse, M. (1966), *Models and Analogies in Science*. Notre Dame (Ind.): University of Notre Dame Press.
Hohenberg, P. and Kohn, W. (1964), "Inhomogeneous Electron Gas", *Phys. Rev. B* 136: 864-871.
Israel, D. and Perry, J. (1990), "What is Information?", in P. Hanson (ed.), *Information, Language and Cognition*. Vancouver: University of British Columbia Press, 1-19.
Jablonka, E. (2002), "Information: its Interpretation, its Inheritance, and its Sharing", *Philosophy of Science* 69: 578-605.
Judson, H.F. (ed.) (1979), *The Eight Days of Creation*. New York: Simon & Schuster.
Karplus, M. and Petsko, G.A. (1990), "Molecular Dynamics Simulations in Biology", *Nature* 347: 631-639.
Kay, L.E., (1997), "Cybernetics, Information, Life : The Emergency of scriptural Representations of Heredity", *Configurations* 5: 23-91.

Kay, L.E., (2000), *Who Wrote the Book of Life ? A History of the Genetic Code*. Stanford: Stanford University Press.
Keller, E.F. (1995), *Refiguring Life: Metaphors of Twentieth Century Biology*. New York: Columbia University Press.
Kirkwood, T.B., Rosenberg, R.F. and Galas, D.J. (eds.), *Accuracy in Molecular Processes: Its Control and Relevance to Living Systems.* London: Chapman and Hall.
Kitcher, P. (2001), "Battling the Undead: How (and How Not) to Resist Genetic Determinism", in R.S. Singh, C.B. Krimbas, D.B. Paul and J.Beattie (eds.), *Thinking about Evolution. Historical Philosophical and Political Perspectives*. Cambridge: Cambridge University Press, 396-414.
Knight, R.D., Freeland, S.J., Landweber, L.F. (1999), "Selection, History and Chemistry: The Three Faces of the Genetic Code", *Trends in Biochemical Sciences* 24: 241-247.
Kohn, W. and Sham, L.J. (1965), "Self-consistent Equations Including Exchange and Correlation Effects", *Phys. Rev. A* 140: 1133-1138.
Lewis, C.T, (ed.) (1879), *A Latin Dictionary*. Oxford: Oxford University Press.
Manher, M. and Bunge, M. (1997), *Foundations of Biophilosophy*. Berlin: Springer Verlag.
Maynard Smith, J. (2000a), "The Concept of Information in Biology", *Philosophy of Science* 67: 177-194.
Maynard Smith, J. (2000b), "Reply to Commentaries", *Philosophy of Science*, 67, 214-218.
Monod, J. (1970), *Le hasard et la nécessité.* Engl. transl. *Chance and Necessity.* New York: Knopf 1971.
Morange, M. (1983), "Schrödinger et la biologie moléculaire", *Fundamenta Scientiae*, 4, 219-234.
Olby, R.C., (1971), "Schrödinger's Problem: What is Life?", *Journal of History of Biology*, 4, 119.148.
Oyama, S. (1985), *The Ontogeny of Information: Developmental Systems and Evolution*. Cambridge: Cambridge University Press.
Quastler, H. (1958), "The Status of Information Theory in Biology", in H.P.Yockey (ed.), *Symposium on Information Theory in Biology*. New York: Pergamon Press, 399-402.
Quastler, H. (ed.) (1953), *Essays on the Use of Information Theory in Biology*. Urbana (Ill.): University of Illinois Press.
Sarkar, S. (1996), "Biological Information: A Skeptical Look at some Central Dogmas of Molecular Biology", in S. Sarkar (ed.), *The Philosophy and History of Molecular Biology: New Perspectives*. Dordrecht: Kluwer Academic Publishers, 187-231.
Sarkar, S. (2000), "Information in Genetics and Developmental Biology: Comments on Maynard Smith", *Philosophy of Science* 67: 208-213.
Shannon, C.E. (1948), "A Mathematical Theory of Communication", *Bell System Tech. J.* 27: 379-423 and 623-656.
Soll, D. (1990), "The Accuracy of Aminoacylation Ensuring the Fidelity of the Genetic Code", *Experientia* 46: 1089-1096.
Stryer, L. (1995), *Biochemistry*. New York: W.H. Freeman and Company.
Thompson, R.C. (1988), "EFTu Provides an Internal Kinetic Standard for Translational Accuracy", *Trends Biochem. Sc.* 13: 91-93.
Yoxen, E.J., (1979), "Where does Schrödinger's "What is life?" belong in the history of molecular biology?", *History of Science* 17: 17-52.

[1] It suffices to recall that two alleles made up of the same number of nucleotides have the same Shannonian informational content. Unfortunately this is not what we want, if we accept Crick's definition, as many authors have clearly written since the years in which Shannon's theory was proposed. For example, the French information scientist Brillouin (1956) and the Italian biophysicist Ageno (1986) did it. Nevertheless the warnings of Brillouin and Ageno were not taken into due consideration and others have continued speaking in terms of mathematical theory of information (for example, Quastler, 1953; Gatlin, 1972; Atlan, 1972). Note that I am not claiming that the mathematical

theory of communication cannot be used in dealing with molecular genetics, but that it is totally useless if Crick's definition is concerned (cf. Sarkar, 1996, pp. 209-210).

[2] I do not even consider the possibility that someone really supports the thesis according to which genes actually carry information, in an ontological sense.

[3] This distinction is crucial for Sarkar (1996), also because his historiographical thesis, that I share, claims that if the notion of information had a heuristic role at the birth of molecular genetics, then it lost any useful explanatory role. For the heuristic role, cf. also Judson (1979); Fantini (1988); Keller (1995); also Maynard Smith (2000a, p. 179) mentions that "scientists need to get their idea from somewhere". As well known, this heuristic role seems to begin with Schrödinger (1944); with reference to this work, cf. Olby (1971); Yoxen (1979); Morange (1983). On the historical and sociological relevance of the notion of information, cf. Kay (1997) and (2000).

[4] By taking into account the well know difficulties of interpreting biological probability as a frequency or as a propensity, especially with reference to probabilistic causality, I would suggest to interpret it in a subjective way, that is, *à la* B. De Finetti. Nevertheless, I do not discuss here such an interpretation. Note that all the probability attributions mentioned in the text are taken from the quoted references and, of course, all of them are based on empirical statistics.

[5] Of course, this probability attribution derives from what said some lines above. See the reference indicated.

[6] We know that in nature there are four fundamental forces: gravitational, electromagnetic, weak and strong forces. Of course neither the weak nor the strong forces has a primary role in the intracellular molecular dynamics. However, if one wants to study accurately and at a physical level some mutational processes which occur in the DNA structure, one is obliged to consider the weak force to understand the radioactive decay which can produce some of these mutations. It seems that not even the gravitational force has a lot to do with the intracellular molecular dynamics, due to the extremely small masses of the molecules involved. This is only partially correct since, from the point of view of the development from the zygote on, the gravitational force between the masses of the cellular molecules and the mass of the Earth has a role in breaking certain symmetries and so in determining certain axial structures.

[7] By the way, exactly this model permits us to understand that the metabolic errors inside the cell are totally physiological, since all the interactions may be realised, also the more energetically disadvantageous ones. These metabolic errors would not happen only if there were (or if the cell had evolved) infinite potential barriers. Yet this does not occur, since it is physically impossible. Exactly this physical impossibility permits that engine of evolution which is the modification of the primary DNA structure that may lead to DNA mutations. In short, it would not be wrong to claim that *it is a physical impossibility which permits biodiversity*.

[8] Two attempts are worth recalling (for a review, cf. Car, 2002; Carloni *et al*., 2001). The first one is called "Density Functional Theory". In this approach, instead of directly trying to find out the expression of the many-electron wave function, attention is focused on the single particle density

$$r(r) = n\int dr_1,...,dr_{n-1}\left|y\ (r_1,...,r_{n-1})\right|^2 .$$

In this way to perform calculi is easier, since $r(r)$ depends on only one spatial coordinate (cf. Hohenberg *et al*., 1964; Kohn, 1965). Instead in the second approach, the so-called "*ab initio* Molecular Dynamics", only the valence electrons are considered: only these are relevant from the point of view of the bio-chemical bonds, while the other ones are considered negligible as to the effects on the binding properties and therefore irrelevant for the calculi (cf. Carr *et al*., 1985). This is not the mere showing off of useless formulas, but the suggestion that there are physical methods capable of grasping what happens inside the cell at molecular level, and therefore capable of giving us both a good reductionist explanation and a precise reductionist description of what occurs when we deal with gene expression.