The differential method and the causal incompleteness of Programming Theory in Molecular Biology

Giuseppe Longo, Pierre-Emmanuel Tendero

The differential method and the causal incompleteness of Programming Theory in Molecular Biology[1]

Giuseppe Longo

Laboratoire d'Informatique
CNRS et École Normale Supérieure, Paris
et CREA, École Polytechnique
longo@di.ens.fr ; +33144323328
http://www.di.ens.fr/users/longo

Pierre-Emmanuel Tendero

Laboratoire d'Informatique
CNRS et École Normale Supérieure, Paris

Summary
The " DNA is a program" metaphor is still widely used in Molecular Biology and its popularization. There are good historical reasons for the use of such a metaphor or theoretical model. Yet we argue that both the metaphor and the model are essentially inadequate *also from the point of view of Physics and Computer Science*.

Relevant work has already been done, in Biology, criticizing the programming paradigm. We will refer to empirical evidence and theoretical writings in Biology, although our arguments will be mostly based on a comparison with the use of differential methods (in Molecular Biology: a mutation or alike is observed or induced and its phenotypic consequences are observed) as applied in Computer Science and in Physics, where this fundamental tool for empirical investigation originated and acquired a well-justified status. In particular, as we will argue, the programming paradigm is not theoretically sound as a *causal* (as in Physics) or *deductive* (as in Programming) *framework* for relating the genome to the phenotype, in contrast to the physicalist and computational grounds that this paradigm claims to propose.

Key Words: Genome, programming theory, differential methods in Physics and Biology. System Biology.

Introduction

Mathematical modeling is the (implicit) aim of researchers, who refer to technical notions such as the notion of computer program; however, in our view, even the metaphor, even when used in a loose way, contains a very relevant scientific commitment[2]. Computer Science is a well-construed science, largely grounded on and directly originated from the formal approaches to Mathematical Logic and, as such, it has its own robust theoretical (and philosophical) commitment. In particular, we claim that, when its notions are

[1] *In* **Foundation of Science,** n. 12, pp. 337-366, 2007**.** A preliminary (and longer) French version of this paper is a chapter of **Evolution des concepts fondateurs de la biologie du XXIe siècle**, (Miquel et al., eds) DeBoeck, Paris, 2008.

[2] One can find uncountably examples in the litterature of the metaphoric or even technical use. A very high standard paradigmatic example is [Danchin, 2003] .

projected on the world of Nature, they impose to it a specific *causal structure*; in short, the metaphor, and more so the modeling, contain a non-neutral proposal for intelligibility (and an implicit Philosophy of Nature). In particular, as we will try to argue, the programming paradigm cannot capture the *causal relations* that should link the genome to the phenotype.

Our argument, forcibly informal as Molecular Biology is not a formalized discipline, focuses on an "incompleteness" and, thus, may be viewed as a "negative result" (or remark): the difficult alternatives are a matter of ongoing theoretical exploration by many (our own modest and preliminary attempts, printed elsewhere, are just hinted at the end). We claim though that getting rid or limiting the conceptual bias of a wrong theoretical (and philosophical) frame is a first step towards new ideas.

1. Modern Logic and Physical Space-Time

Let's try to hint briefly to a long history, which goes from the origin of modern Mathematical Logic to today's digital (*arithmetic*) computers. Surprisingly enough, this story is strictly related to the major crisis in our knowledge relation to physical space; a crisis that lead to radical revolutions in Physics and, in particular, made us understand the world, in terms of *causes*, typically, in a novel way.

It is the immense crisis caused by the birth of non-Euclidean geometries that forced many, Frege among others, to look for an alternative, arithmetic, foundation of Mathematics. The "delirium" ([Frege, 1884]) of the intuitive understanding of space in Riemannian geometries, made him make the courageous step of founding Mathematics out of space and time, in the "absolute concept" of *integer number* and on logical law of *arithmetic induction*: Arithmetic is Logic, for Frege. His novel and deep insight joined the dual approach by Boole (1854), who had arithmetized logic.

By Hilbert unifying approach, the Foundation of Geometry (1899) was then reduced to the formal consistency of Arithmetic, the bottom line of human certainty as for Mathematics. And here we are at the fantastic, yet purely mathematical, arithmetic functions and machines of the 1930s: the computable functions by Herbrand and Gödel, Church's lambda-calculus and the paradigmatic 0-1 Machine by Turing (1936). In that machine lies the logic core of formal computations and, then, of the notion of program: sequence-checking and sequence replacement. All what these systems can do is: check whether two sequences of numbers (or of 0s and 1s) are identical, move or change one or more digit (it really looks like a - parody of – genome). But the digital environment must be *exact* (and *absolute*, at least in the sequential machine, see below): it is a matter of a *Logical* Computing Machine, as Turing calls it, a man in the least act of thought (write 0 or 1, replace it by a 1 or a 0, move along the finite sequence), as exact as Frege's absolute logic.

And it is a *Cartesian machine*, as Turing introduces a crucial distinction: the program (the software) is totally independent from the hardware, a scientific realization of the soul/body dualism. Moreover, perfect iteration is at the core of computing: primitive recursion, the mathematical description of its core, is *iteration plus updating a register* (nothing else is needed). So, by the distinction software/hardware and identical iterability, one has the *portability* of software: without it, Computer Science as a science (and

Microsoft as a business) would not exist. In particular, you may take the soul of a computer and transfer it identically on another.

Finally, and most importantly, Turing's Discrete State Machine, his subsequent alternative name for its invention, containing now a clear reference to Physics, is *Laplacian* (similarly, we insist, the equivalent systems by Gödel, Church and the others, when one wants to force into them a naturalistic frame – for which they where not meant, as they were pure Logic). Turing observes this twice in [Turing, 1950; pp 47 and ff.] (see [Longo, 2002] for a discussion), and contraposes its predictable determinism to the unpredictability of what he calls deterministic "continuous systems" subject to the "exponential drift" as the morphogenetic systems he models in [Turing, 1952]. We call now these deterministic systems "sensitive dynamics to initial or border conditions", possibly described by non-linear equations. Of course, also a computer program may be *practically* unpredictable, as it may be *very* long and complicated, Turing observes; but unpredictability in non-linear systems is a key *theoretical* property. The theory even allows to evaluate the level of unpredictability: the value of the so-called Lyapounov exponents, say, or other criteria of divergence of initially close trajectory, like the exponential drift in Turing's model of Morphogenesis[3].

2. Networks of concurrent and distributed processes

Today's Computer Science is witnessing a major change in the hardware of computers, which is forcing, beyond expectation, a change in programming paradigms. The Church Thesis (the claim that all logical-computational systems compute the same class of functions) is getting inadequate or false, if one considers distributed and concurring computers. That is, different formal descriptions of synchronization mechanisms may yield different computational powers (see [Aceto et al., 2003]). Thus the myth of the absolute notion of computation is fading away, while the enrichment of this very notion may broaden its applicability (to biological phenomena, for example).

One thing should be clear though: if one enlarges the notion of "program" up to departing radically from Turing-Church frame, by identifying it, say, exactly to what DNA does, then of course, one may claim that the "DNA is a program". However, this wouldn't increase much our understanding of genes and it is not what it is meant by the programming paradigm in genomics.

The crucial issue with concurrent computing resides in the synchronization of processes that are distributed in space and may "concur" to a computation, i.e. they may share data bases, use partially the ongoing process one of the other…. The fact is that, by networks possibly distributed on the Earth surface, physical space-time stepped in computations, against the expectation of the founding fathers who thought of them as a purely logical, abstract activity: an isolated man in the least act of thought, said Turing, in an abstract, stepwise – sequential time. In contrast to this, in theories of concurrency, time is a matter of synchronization of possibly asynchronous processes and it may be "stretched", instead of being step-wise (processes may be in long transitions and cancellation states), composition of processes may be no longer associative. Relativistic

[3] Also Shroedinger, who uses the word "program" for the chromosomes, is aware of its Laplacian implication: given a complete knowledge of the code, "the all-penetrating mind, once conceived by Laplace" would access complete prediction [Shroedinger, 1951 ; pp. 22-23].

effects may modify the causal structure… (see the various articles in [Aceto et al., 2003]). Once distributed in a physical space-time that we better describe by continuous mathematics, Turing's Discrete State Machines and their programming are changing in nature. Is this enough to consider these novel paradigms for computing broad enough as to include genome's and proteome's dynamics?

Before discussing these various questions that go well beyond the usual programming metaphor or even the modeling for genome, let's move to a closer analysis of this core discrete chemical component of life.

3. DNA

From the few certitudes that we may have concerning the theory of computability, so well tied to that which is elementary and simple (all the while reaching the very complex, by composition of that which is simple), we shall move on to a critical analysis of the notion of genetic program. This is proposed as a framework of interpretation for the relationships between the very complex element which is the living cell and this discreet component of biological systems of utmost importance: DNA. The cell is an *elementary* component given that it dies if we cut it into pieces, all the while being *complex* (perhaps, even infinitely complex, in relation to any reasonable physical measurement). Further on in this paragraph, we will examine the question of the physiological role of genes and we will question ourselves, first of all, about the successive definitions that have been proposed, as well as about the compatibility of these definitions with one another (§3.1). In order to highlight that which, in our view, is problematic here, we will attempt (§3.2) to bring forth the fact that, despite the reiterated historical assertions of genetics, the existence of a direct causal relationship between genes and characters cannot be solidly established on the sole basis of experimental data, inasmuch as they present themselves *a priori* as results of *differential* experiments, in a sense that we will attempt to make explicit by means of references borrowed from Physics and Computer Science. The main question of interest to us will therefore be that of knowing if there exists an adequate theoretical framework enabling the conciliation of the assertions of genetics (classical, but also molecular) with experimental facts. We will particularly emphasize that, from this point of view, programming theory is unable *in principle* of providing such a framework (§3.2.2). Bypassing this first restriction (*which nevertheless remains a diriment for the theory of the genetic program*) and admitting, despite everything, that there exists a direct causal relationship between genotype and phenotype, *at least in the case of molecular biology*, we will then consider (§§3.2.3 to 3.5) whether the recourse to the metaphor of Turing machines or more recent programming methods and concepts as well as to certain notions related to information theory enables us construct a robust conceptual framework capable of unifying under the concept of gene the structural and functional properties of DNA.

3.1 The concept of gene as seen by the history of modern genetics

From the standpoint of the history of modern genetics, it seems that we may identify three great successive characterizations of the notion of gene having overlapped to finally produce an operational definition. This definition, while very precise, has become problematic in the case of multicelular eukaryotes. Indeed, the gene was first defined as:

1 - a *hypothetical functional unit of recombination* (of assortment and segregation), objectivizable within the framework of hybridization experiments inasmuch as it enabled the formulation of predictions, *in probability*, based on the frequency of traits in the lineage of a parental generation of which the genotype was assumed to be known [Mendel, 1907]. Later on, based on the works of T. H. Morgan and of his collaborator Hermann J. Muller [Morgan, 1926], this definition was complemented by the additional hypothesis according to which the gene would be the locus of a *structural modification at the chromosomal level*, modification which was supposed to intervene during *the differential transmission of characters*. In other words, the gene became:
2 - a *structural unit of mutation*, unit which, moreover, was still only hypothetical since the localization of genes at the chromosomal level rested solely upon a *simple formal analogy* between the empirical observation of the recombination of traits in the lineage of individuals and some of the remarkable behaviours of chromosomes at the moment of meiosis (*crossing-overs*, mechanisms by which paired chromosomes, separating each other in order to form haploid cells, "randomly" exchange some homologous portions of chromatin). Finally, with the discovery of the three-dimensional structure of DNA [Crick and Watson , 1953a,b; Crick, 1957], the concept of gene received, for the first time:
3 - a *characterization in terms of molecular biology*, both structural and functional, through the formulation of the hypothesis – or rather of the "central dogma" – of the existence of a *direct-causal* relationship between genes and proteins that the discovery of the genetic code would be soon to confirm.

In our view, it appears to be increasingly difficult to make sense of this history. It even seems fathomable that these three successive determinations of the notion of gene do not refer to identical entities and/or processes. The issue then is the precise contribution, in terms of theoretical explanation, of the metaphor of the computer's program in genetics.

3.2 Structure and function in the definition of genes.

Quite generally, in these successive characterizations of the notion of gene, one can distinguish two remarkable aspects, unfailingly related to one another despite that their conciliation proves to be clearly problematic. The first of these aspects, even if it appeared only secondarily from a historical perspective, relates to the *structural characterization of genes*, while the second concerns their *functional definition.*

3.2.1 A few points concerning the orientation of the following remarks.

At the beginning of the XXth century, Morganian cartographic works have quite distinctly suggested the existence of a structural characterization of genes. Nevertheless, it appears fundamentally problematic to peremptorily assert that this structural characterization is absolutely superimposable to the one which molecular biology proposes today.

First, the *loci* identified by means of Morganian hybridization protocols were indeed related to characters only on the basis of a *causal* interpretation, *scarcely argued on the physiological level*[4]. The "causes" were given as a set of *empirical correlations* established between, on the one hand, the observable distribution of recombinations of

[4] And it could not have been otherwise given the means of molecular investigation of the time.

pairs of characters amongst the lineage of a given parental generation and, on the other hand, the frequency of *crossing-overs* at the level of the individual's germinal cells. This type of statistical methodology leaves fundamentally open the three following questions, which are crucial today from the moment is posed the question of the reinterpretation of the assertions of classical genetics using the language of molecular biology.

1) On the basis of which physicochemical principle that is *a priori* or of indisputable empirical evidence can we assert that the locations in question *concerned only the DNA molecule?*

2) If the observed correlations must be interpreted using a vocabulary of causality, which is not as obvious as one would think (*cf*, for example, §. 3.2.2), is the causal relationship thus revealed necessarily of the *bottom/up* type such as exists for the mechanisms of transcription and translation of genetic information[5] ?

3) What causal role can we attribute, on the basis of these analyses, to the three-dimensional structure of chromatin, which clearly has a non-negligible effect on the kinematics of the chemical reaction of which the cell's nucleus is the locus?

Second, we can also observe that, since the 60s, with the structural definition of genes in relationship to the synthesis of protein, the question of the relationship between genotype and phenotype is no longer posed at the level of observable characters as such, but at the level of the cellular metabolism, thus indicating for molecular biology a very clear regression, in terms of domain of explicative validity in comparison to classical genetics. We will further see, moreover, that, strictly speaking, we do not even attain the level of cellular metabolism as such but only that of the *potential enzymatic role of proteins in the biochemical cycles specific to living organisms,* albeit problematically[6].

However, there still exist today very important works that reveal the role of certain *modifications* of the structure of the genome of species in the modification of a certain number of associated phenotypical elements. For instance, it is well known that modifications in the disposition of homeotic genes induce *teratogenic* effects on the development of Drosophila embryos. But there lies, precisely, something of a problem to our logician and computer scientist eyes: this teratological, or simply *differential* aspect, specific to the methodology of genetics, does not seem to lend itself as easily as it would appear to an analysis in *causal terms*, or *deductive*, as in programming. Let's discuss this issue, which, in our view, is fundamental.

3.2.2 Differential methods and causal structure.

The dominant empirical methodology in molecular genetics consists in the introduction of genotypical modifications and the observation of eventual phenotypical variations. In

[5] Which, moreover, are in this respect the only ones for which molecular biology possesses a theoretical explanation which is systematic in the strictest sense of the term.

[6] These cycles, moreover, are not limited to simple chemical reactions between proteins but involve, sometimes crucially, various chemical species, molecular or not, of which it is rather difficult to assert that their physicochemical properties are dependent of genes (think of the case of ions…).

what way is it possible to interpret empirical correlations of this type by means of causal vocabulary? How does such theorization relate to the usual practices of Physics, for which differential methods constitute a common experimental protocol? Or to those of Computer Science, where the notion of "program" has its own scientific origin and practical applications?

In Physics, we call "differential method" the experimental practice which consists in deriving certain causal relationships by setting all the parameters of a system, except one which is modified within a certain restricted domain of values. That can lead to propose a relationship (in fact, an equation) between parameters and the observable and, thus, *reveal causal relationships*, which, moreover, may be soundly immersed in the general frame of *formal symmetry breaking*.

Let's consider, for instance, the case of the equation of the perfect gas in a state of equilibrium, pV = KT. Once this equation has been asserted, a formal symmetry breaking corresponds to a breaking in the equilibrium since it will be interpreted as a *possible* causal relationship between the various parameters (for instance, a variation of p *to the left* will cause a variation of T *to the right*…). By applying the differential method in Physics, such an equation can thus be reconstructed *empirically* by inducing only a variation in pressure p, for example, within a certain window of values. In short, this equation continuously interpolates a (possibly large, but finite) set of empirical correlations between p and the other values. Yet, *an equation of the pV = KT type* (a remarkable step towards understanding) *is only one of the many possible formal determinations of finitely many empirical correlations,* and does not lend itself, as such, to a specifically physical analysis in terms of objective causality [Bailly & Longo, 2006]. Specifically, there are always an infinity of laws – of equations – of this type (polynomial, that is), enabling one to perform an interpolation between the same empirical data, within the window of admissible parametric values. However, as it was precisely the case for the equation of perfect gases 50 years after its historical formulation, a relationship of this type can also (and we would be tempted to say "must", in Physics) be *mathematically deduced* in an *a priori* fashion within a more general theoretical framework, that of Statistical Physics. What appears to be important to us is that, while the physical causality revealed in both cases is not the same, in the second (Statistical Physics) it is given by a differential equation proposing a framework of systematic *theoretical determination*. As a matter of fact, this equation is based on very general principles, the *geodesic principles* (or of least action), which govern the analysis of the trajectories of a gas, then transferred to the thermodynamic limit (an infinite limit, an integral). This fully justifies the empirical differential analysis (in an *a priori* way from a deductive point of view, even if it is generally subsequent from an historical point of view) as it allows to *derive* the pV = KT equation from those general principles (geodetics and related symmetries). Last but not least, the formal derivation within the framework of a well demonstrated theory produces *only one* solution out of many possible ones; thus, the recourse to experiments limits itself to the determination of the constant K – or to the eventual falsification of *all* of the theory, without permitting the trick of only changing the equation. Within well constructed theoretical frames, counterexamples destroy theories.

Let's return to the specific problem of concern to us. As for the existence of an eventual causal relationship between genotype and phenotype, it is thus necessary to

emphasize the fact that, although the differential method in genetics claims to attain the demonstrative rigor of physical theories, the lack of a more general theoretical framework does not enable to confer the status of *scientific proof* to the sole currently available experiments of teratogenics. What makes the discourse of genetics "causally incomplete" from this point of view is simply the fact that it is not in a position to establish the formal existence of a relationship of *direct causation* between genotype and phenotype. One may, in fact, suspect here the implicit persistence of a certain pre-theoretical (even pre-scientific) belief, such as the evocation of the metaphor of the computer program, could produce a systematic *theoretical determination*. And this without necessarily feeling the need to demonstrate it within a specific theoretical framework (physical, computational, or better: *specifically biological*), comparable, even vaguely, as for conceptual depth, to the frame proposed by the "geodetic principle" in Physics. What we are trying to say, and we will express it in other words later on (§§. 4.3), is that a simply differential study of the relationship between genotype and phenotype, though very rich in empirical data, is, *in principle*, insufficient to immediately derive a causal theory, or a deductive theory as in programming (because a program is a "deduction"), within which a given genotype would produce a specific phenotype, or even a range of phenotypes according to a given set of development parameters (that is, what molecular biologists call the "reaction norm")[7].

The physical singularity represented by living matter is so great from this point of view that a physicalistic theory of its determination would require as fine an analysis as in Quantum Physics or relativity, where probability correlations or Minkowski spaces, respectively, provided us with original insights into causal relationships. An analysis which we expect to be, in principle, completely different, however, from those specific to the theories of modern Physics (which are far from being unified and which fundamentally differ from classical frameworks), in view of the peculiarity of life phenomena, from afar, the Laplacian predictability that Turing himself attributes to his discrete-state machines (§ 1 and [Turing, 1950]). One of the stronger reasons in favour of this singularity of life follows from the intertwining and causal loops between the levels of organization specific to living organisms. Integration and regulation, concepts which are rather difficult to express within the framework of current physical theories, - and which are, in particular, quite different from cybernetics' concept of feed-back – constitute one of the components of this causal intertwining and recurrence[8]. To this, it is

[7] The problematic character of such causal relationships may be illustrated by a further example from Physics. Empirical evidence, at least since Aspect's work in 1980's, confirms that, in the case of an intricate state, a measure upon a particle of the system – measure which *is* a modification of the state of that particle – instantaneously induces an identical modification in the state of the associated particle. Now, that does not demonstrate that there is a relationship of causality between the two events, by an exchange of energy or information. Moreover, it could not be a question of such a relationship in this case, given the finitude of the speed of light. By a suitable interpretation of Bell's inequalities, the theoretical frame radically departed to the classical and relativistic understanding, with no need to metaphorical reference to other existing theories of information, programming or whatever: just an autonomous field theory to be later and possibly unified with other existing theories. Life phenomena departs from current physical theories, including the one implicit to the notion of program, at least as much as physical theories of different phenomenal levels - Astrophysics and Microphysics say - differ among them (all the while looking for a unification, more than a reduction).

[8] An example which is very simple but sufficient to highlight the difficulties of a direct causal analysis is

necessary to add that, these singular aspects having been integrated, it will still be necessary, from a strictly physicalistic standpoint, to account for this remarkable characteristic of biological systems which is the unity of an organism, that we attempt to analyze, further on, as a system of critical type having its *correlation lengths*, between parameters, of the size of the object itself (see Conclusion and [Bailly & Longo, 2006]).

It is clear that these different elements, particularly the complexity of the correlations between parameters and observables, will in all likelihood make all the more difficult the rigorous application of the differential method, which consist of the setting of all parameters (and observables) except one, that of causing variations in the latter and of observing only the direct consequences. But maybe it will be possible to do so thanks to a variational theory of phenomena *far from equilibrium* (doubtless of a novel type in Physics) and close to criticality (if not within it(!)), where the physical correlation lengths diverge (we will shortly return to this).

A further critique of the differential method, not as an empirical practice, beautifully mastered in Molecular biology, but as a direct ground for theoretizing, is suggested by the case of phenocopy, known since [Goldschmidt, 1938]. Developmental biologists are able to create "phenotypical clones" by simply modifying, in various manners, the experimental conditions of the development (see [Stewart, 2004] for recent references). So, from a differential standpoint, the pressure, the chemical composition of the environment or the local intensity of the electromagnetic field, as parameters of the ecosystem and, more generally, all the extragenic context, *are* as much causes of the development of the embryo as is the genome, if we remain at a *purely empirical* level. We can, for instance, induce teratogenic effects similar to those induced by the displacement of a homeotic gene (see § 3.2.1), simply by modifying the pressure at a certain moment during embryogenesis. And, from then on, the attribution of causality, or rather, of primal causality to the genome in the production of a given phenotype *does not appear to bear an indisputable empirical foundation and lacks a rigorous theoretical framework.* To return to the example that we have just mentioned, we do not see any reason in principle why to exclude at once the possibility that certain modifications in the disposition of homeotic genes would in fact *cause* a change in the embryo's reactivity to pressure…

To summarize, we have just attempted to demonstrate that the existence of empirical correlations between the modifications of two physical or simply logical structures does not demonstrate the existence of a direct causal relationship between them and that, reciprocally, the inexistence of such correlations does not demonstrate the absence of such a causal relationship. In the case of developmental genetics, we remain, in our opinion, and despite a remarkable experimental richness, at the level of taking notice of the existence of empirical correlations between the modifications of genotype and of phenotype. And this from the sole physicalist view point; we will show that the situation is even more unsatisfactory from the perspective of the paradigms of Theory of Programming.

---

the following: a modification in the blood of the concentration into certain ionic categories, for instance, may cause a drop in blood pressure, contrarily to the anticipated physical effect, following regulation mechanisms which over-react in the opposite way (ago-antagonistic effects, [Bernard-Weil, 2002]).

Concerning the issue of the role of context with regard to the differential method and its interpretation within the framework of the metaphor of genetic program, there remain two remarks to be made. First, if it is true that the *real* functioning of a computer program also presupposes the existence of a context (operating system, a certain type of hardware), programming theory was nevertheless created *in abstracto*, outside of the world, as a theory of computability, from Turing's mathematical machines (which do not necessitate a *physical* context, as we mentioned earlier – it is "man in the act of minimal thought"). This legitimate independence in relation to the physical world remains essential to the practical developments of Computer Science – even of concurrency and networks – as a science of the portable software, that is to say, of software which, within adequate operational and physical contexts, may be identically iterated (see § 1). The identical iteration of processes is so central to computing, that, usually, the reference to the physical structure of computers is made only in order to explain dysfunctions and not behaviour deemed "normal". In a certain sense, even the problems posed by concurrent processes within distributed networks are exactly the "problems" to which we attempt to remedy by means of a good theory: we do want that our web page is opened *identically* millions of times independently from the access path – and it usually works.

Following this, it seems dubious that the eventual information coded within DNA may be interpreted according to terms borrowed from programming theory as such: the context of the expression of genomes is, indeed, certainly not a passive locus of identical iteration, but, rather, can be described as an active space of ontogenetic constitution (cf. §3.5), where *the variability is as important as the stability*, which, strictly speaking, *does not have an equivalent in Computer Science*. Second, the differential method, as it is applied in Physics, supposes, as we have seen, weak, - even lacking – correlations (finely analyzed) between the contextual parameters and the variables of which one is studying the behaviour in function of certain parameters judged *a priori* to be more relevant. This hypothesis would not seem legitimate for the analysis of extragenomic or epigenetic contexts of ontogenesis, because the aspect of biological systems, which constitutes the greatest challenge for current physical theories, consists in the existence of this particular "causal field", specific to the living cell, where (almost) everything is correlated to (almost) everything else and where the effects of "resonance" seem infinitely (and we use this word in a mathematical sense) more complex than in any physical dynamics, even non-linear.

More modestly, we claim that an eventual primacy of the genome, even if making itself known by means other than the differential methods, would have absolutely nothing comparable to the *structure of logical determination specific to the formal notion of program*[9].

[9] This restriction does not exclude, of course, that there exists very local cases (such as those, for instance, of certain rare genetic diseases) for which the relationship between genotype and phenotype follows precisely from this type of causality. However, what we want to indicate here is the fact that it is not a given that these particular cases constitute, in fact, the general model on the basis of which all possible cases can be analyzed. And even if these specific cases where not rare, as they are, it could actually be the exact opposite of a paradigm: in Physics (...Physics, once more... but don't molecular biologists want to be physicalist ?), we can recall that the Aristotelian principle, grounded on large empirical evidence, according to which all mobiles immobilize from the moment one ceases to apply motion to them, is precisely opposite to the Galilean inertia principle, which is alone to be theoretically relevant.

In any case, for the moment and with regard to the answer to these questions, it appears to us, to return to the problem of the historical succession of the three paradigms of the genetic theory from where we started, that the discourse of the theory of heredity in molecular biology (laws of Mendel included), has now shifted to no longer be restricted to the explication of molecular mechanisms of replication and expression of the genetic information within cellular masses [Pentris *and al.*, 1983]. This assertion and those preceding it have not, in this sense, any other pretension than to indicate the existence of an eventual interpretational bias which may introduce itself in an *a posteriori* rereading of the propositions of classical genetics before the advent of the molecular paradigm and, particularly, of the metaphor of the genetic program. The issue which we propose to examine following this is, in fact, that of the relevance of the latter metaphor and of its related notions, from the viewpoint of the understanding of the way by which the characteristic structural and functional aspects of genes can be synthesized, by means of the sole conceptual tools of molecular biology, into a coherent whole capable of providing a plausible explanation of the way by which genes *participate to the forming of living organisms*.

3.2.3 One gene – *many* proteins *vs. one* protein - *no* genes.

If one considers a few decades of a selective set of works in contemporary molecular biology – selection which we readily imagine to be too thin with regard to the fantastic density of empirical data –, it appears to us that structural genes are today assimilated to portions of DNA. These are potentially associated, *via* the genetic code, to proteins whose sequences of amino-acids are determined by their constitutive series of nucleotide triplets. Nevertheless, we can already see that this definition is not, strictly speaking, a purely structural characterization of genes since it is fundamentally indissociable from their functional definition, which is that of serving as matrix to the synthesis of "one" protein, at least in the case of the simpler organisms.

Indeed, in the case of superior eukaryotes, it appears to be rather more complex, inasmuch as it would appear that, for the latter, genes can no longer be simply conceived as *uninterrupted* segments of DNA in linear correspondence to specific proteins, modulo the inclusion of some "detritus". In this respect, we notice here that the very term of detritus, generally employed to designate non-coding portions of fragmented genes in the genomes of eukaryotic species, appears to be most inappropriate. It is clear, as many publications tend to demonstrate, that the presence of these portions not associated to proteins *contribute to the adaptation capacities of cells to their environment*: after the transcription of genes; the phenomenon of alternative splicing[10] would provide the latter with an assured plasticity, [Brett et al., 2001]. In particular, proteins have been found that, while in principle associated to the "same" gene, differ from the standpoint of the regulation of their metabolic activity, in relation to their organic function (one gene, many proteins)[11].

[10] That is, the set of post-transcriptional regulation "mechanisms" destined to bring the primary transcript to maturation by excising the messenger RNA segments associated to the gene's intronic sequences and by rearranging the coding portions in various manners.

[11] The number of proteinaceous variants which may be synthesized from a same gene can thus reach the several hundreds, as is the case, for example, for the cSlo gene of the chicken's inner ear's hair cell which

Now it is precisely here that there arises a fundamental difficulty with regard to the definition of gene in structural terms, if one wants to interpret it using functional vocabulary. What is missing, in a great number of cases, is a direct relationship between genes and the expression of the information contained therein. Moreover (and conversely), this degenerate relationship (see also § 4.2.1) is sometimes brought to its acme through the existence of mechanisms of extragenomic modification of the sequence of primary transcripts. These then become susceptible of conducting to the synthesis of certain proteins *having no antecedent within the genome of the cell itself* (one protein, no gene). We could also add to this that the distinction between intron and exon is no longer even relevant for the definition of the physiological role of genes inasmuch as it appears that the synthesis of certain proteins effectively involved portions of intronic DNA.

There is, nonetheless, a solution to this problem of the "empirical degeneracy" of the notion of gene, a solution consisting in relating the definition of genes no longer to certain portions of DNA, but to their equivalents in terms of RNA. Yet this latter alternative would not prove satisfactory inasmuch as it seriously throws into question the traditional interpretation of the role of genes in the transmission of characters. As a matter of fact, it would then be necessary to integrate the problematic idea that the latter may have, such as segments of RNA leading to the synthesis of proteins, a *discontinuous* existence *in time and space*, thus invalidating, by the same token, the absolute primacy conferred to DNA within the process of development.

3.3 The notion of information in the definition of the concept of gene.

It appears to us that the recourse to the vocabulary of information theory by molecular biology generally remains rather informal. Thus, the very notion of genetic "information", employed in reference to the existence of a code – which is, moreover, redundant or rather, degenerate (see §.4.2.1) – linking certain series of nucleotides to amino-acids, seems today incapable of accounting fully for the relationship between genes and proteins. The reason being, among others, that it perpetuates, by means of an Aristotelian wordplay[12], molecular biology's central dogma according to which the function of a gene could be deduced from the sole sequence of its nucleotides and, given the existence of alternative splicings, this would henceforth be *largely subject to caution from a strictly physiological point of view*.

From this standpoint, one of the best examples that one could give of the very informal aspect of the recourse to this vocabulary is certainly that of the interpretation of experiments involving the inactivation of some *knockout* genes. Indeed, it has been often observed that this type of experiment does not necessarily conduce to the pure and simple

---

reaches 576 variants (see [Black, 1998]). The concerned proteins, moreover, have this remarkable property that they intervene within the cell in order to modify its various resonance frequencies, which seems to indicate an increase in plasticity and also, therefore, an increase in the level of adaptability of the chicken's inner ear.

[12] The concerned wordplay *is* Aristotelian, in the sense that it refers to the idea that genes would "inform" the proteins, meaning that genes would "put them into form", in the same way that, for the Philosopher, the coming into being of creatures of nature is translated by the imposition of a sensible form <*morjh*> - of which the ideal model refers to an intelligible form <*eidoV*> - through the implementation of a certain organization <*logoV*> which determine them as the actualization of a certain function <*enteleceia*> in the broadest sense of the term.

suppression of the functions to which the studied genes had been traditionally attached; yet, sometimes biologists obtain the unexpected consequence, to say the least, of the latter seems even to be improved. And this fact is precisely interpreted as highlighting the great redundancies of metabolic pathways in living organisms [Tautz, 1992; Thomas, 1993]. Now, if we take seriously the idea of introducing the scientific rigor of information theories into genetics, then we must acknowledge that the very term of redundancy is itself inadequate here. This notion is absolutely unable to account for the *originality of this biological phenomenon* with regard to cybernetic systems. The concept of redundancy has, indeed, imposed itself in the field of cybernetics, from a strictly pragmatic standpoint, in view of regulating the introduction of errors in the automatic processing of messages. The redundancy thus defined is therefore *a structural property of coded information* which in no way affects its function/meaning. Now, in the case of genomes, the term "redundant" is not only applied to the genes of which the structure is repeated, with a few variants devoid of any functional role in the view of faithfully rendering at least one of them, but also to those of which the associated protein(s) *have a function sufficiently similar to substitute one another*. It would therefore be more in accordance with the consecrated use of the term "redundant" to speak in their case of functional degeneracy, a term which appears to have great resonance in the analyses of Gerald Edelman [Edelman and Gally, 2001][13]). But then, all the while taking notice of the enormous importance of an analysis of this apparent informational "surplus" within the genome, we must also emphasize that the cybernetic analogy is here of the most lax and *does not enable to introduce all the rigor of the concept of redundancy in information theory*[14].

3.4 The status of the concept of gene in the metaphor of genetic program.

It also appears to us that the notion of genetic program poses serious theoretical problems of interpretation inasmuch as, in practice, it seems to be employed only very locally, so

[13] We believe that there is no probing argument in favour of the existence of a *generic* original relationship between structure and function, lost over the course of evolution, as some molecular biologists seem to imply (opening the way by this to an original "intelligent design"), and to the explicit difference of Edelman.

[14] We would like to emphasize right now the specificity of the concept of degeneracy for biology: it means much more than a simple informational "surplus" or "deficit". As we have seen, the "knock-out" experiments demonstrate that *sometimes* the suppression of entire sections of DNA can have absolutely no observable "consequence" for the development of individuals, at least, within some reasonable limits. In contrast, there also exist cases where the modification of a sole nucleotide "conduces" to important phenotypical changes. And, doubtlessly, there must exist cases where these two phenomena are observable *for a same portion of DNA*. The possibility of a co-occurrence of these two major phenomena (*a possibility which nothing enables to discard in principle*) suggests that one cannot favour *a priori* the explanation of a single of these crucial aspects of the physiological role of genes at the expense of the other one, by postponing the explication of what remains unknown, by composition of what is known. It appears to us, in this respect, that a rigorous conceptual distinction, not only between redundancy and degeneracy, but also within the latter notion itself, between structural degeneracy and functional degeneracy, could enable us to have a clearer grasp of that which distinguishes these two remarkable aspects for eukaryotic organisms within which a quantifiable *functional deficit* (cf., for example, the successive and different assessments provided over the course of the sequencing of the human genome) is conjugated with an apparent *structural surplus* (the genes are fragmented, repeated, can substitute one another…).

that *it is unable of making intelligible the physiological role of genes as a whole*. There are obviously many reasons to this, among them, the following:

Originally, the notion of genetic program was developed within the perspective of the works of Monod, Lwoff and Jacob concerning the regulation of the expression of the lactose operon in *Escherichia Coli* [Jacob and Monod, 1959]. Nevertheless, it is clear that this latter notion was more or less implicitly present as early as the 30s in some reflections of [Morgan, 1934] and [Goldschmidt, 1938] for instance, concerning the role of genes in the process of embryonic development, with the only difference, however, that these authors precisely believed that this role was not *preponderant*, but were closely dependent on cytoplasmic regulations. What we mean is that the works of Monod, Lwoff and Jacob cannot truly be held as indissociable from the formulation of the notion of genetic program in molecular biology, in the sense, for instance, that they would have provided the experimental basis which was lacking before. During the same period, Ernst Mayr had introduced this notion totally independently, during a theoretical discourse concerning the issue of causality in biology [Mayr, 1959].

Nevertheless, regardless of the theoretical and/or experimental foundations of the concept as such, it appears that these regulatory mechanisms, isolated in the case of the bacterial genome, constitute in fact a legion in that of eukaryotes, which leads to think that the metabolic processes within cells are subject to controls of the same type over the whole course of the individual's lifetime. The central role of these mechanisms within the cellular metabolism is, in fact, absolutely undeniable, but one can ask oneself whether the recourse to the very metaphor of genetic program is necessary to make intelligible these phenomena specific to living organisms. In fact, we observe that, by certain of its aspects, the notion of program introduces an important theoretical bias into the reading of the experimental data. Here are a few examples.

3.4.1 Structural genes *vs.* regulatory genes.

First, the initial model of the lactose operon introduces a very ambivalent distinction between the notion of structural genes, the latter being supposed to have as sole function the enabling of the synthesis of the molecules necessary to the continuation of the cellular metabolism, and that of regulatory genes, which can be of different types according to their mode of intervention upon the regulation of the expression of structural genes. Now it is precisely from this point of view that they rather hardly lend themselves to an interpretation in terms of genetic information inasmuch as the portions of DNA to which they correspond are sometimes not associated to the synthesis of any particular protein. Fundamentally, they serve only to favour the bonds of molecular complexes inducing epigenetic modifications of the genome's structure, which enable the opening or closing of certain adjacent frameworks of interpretation.

The case of regulation genes therefore involves a new difficulty in the general definition of genes with regard to an enzymatic activity of proteins since there visibly exist *relationships between these two entities which in no way involve mediation by means of a code*. This problem is also not exclusively a theoretical issue relating to the sole definition of genes in molecular terms since it also leads to formulate very specific questions on the practical level. As a matter of fact, one needs to number the portions of DNA susceptible of behaving like genes in the genomes of "decrypted" species [Bernot, 2001; Boffelli, Nobrega and Rubin, 2004]. In addition, part of the regulation mechanisms

of the expression of genomes does not rest, strictly speaking, upon relationships which are analyzable in terms of code in the formal sense of information theory, but rather upon dynamic physicochemical interactions, such as those studied by statistical chemistry. This leaves open the possibility of alternate interpretations of the physiological role of the gene *stemming from the same corpus of experimental data*.

There is nothing, indeed, which allows to exclude *a priori* that DNA can only be, with regard to cytoplasm, a simple "analogical databank" submitted to sorting by a program located downwards of the process of translation, that is, in the organelles participating in the cellular metabolism, if we do insist on preserving the computer program metaphor [Atlan and Koppel, 1990]. One can also oppose the deterministic aspect, which we have said to be of a Laplacian type, that the notion of program *imposes a priori* to the relationship between genes and proteins (be it a relationship of translation or regulation), and this because *this type* of determinism is *fundamentally incompatible with the probabilistic character of the predictions of statistical chemistry*[15] [Creager and Gaudillière, 1996 ; Kupiec, 1996], at least as much as it is incompatible with the (possibly) deterministic, but dynamic processes that prevail in a living cell. And we see no physically admissible reason according to which one could suppose that the molecules that intervene in the reactions specific to vital phenomena would be excluded from this type of restriction for chemical analysis.

3.4.2 The notion of stereospecificity.

For a long time, an issue was the question of the existence of *specific* molecules capable of regulating the synthesis of the proteins necessary to the cellular metabolism. It seems however that the great majority of molecular factors isolated to this day are never specific, as such, but rather, indeed, *ubiquitous and totipotent* so that it is frequent that many still non-isolated co-factors are summoned to explain the *noted absence of specificity with regard to the studied proteins*. In fact, the concept of stereospecificity is highly problematic for the interpretation of physicochemical phenomena specific to living organisms. It signifies that the molecules which intervene in the regulation of genetic expression are endowed with *specific bonding properties* on particular substrates, *excluding all others*. Now, the plasticity of macromolecules is a phenomenon well known to chemical kinetics since it is not uncommon for them to have several enthalpic isomers between which continuous oscillations take place, under the effect of thermal agitation *as long as no constraint is applied to them*. Furthermore, when the stated constraints consist in the establishment of *non-covalent* bonds with certain substrates, the stability of the resulting three-dimensional complex *is only temporary* in an enzymatic reaction. There lies a fact well-known to chemical kinetics given that most reactions of this type *are partially reversible*, so that the differential equations which characterize them *concern only the evolution of concentrations and not the behaviour of individual molecules themselves* (see previous footnote).

There is therefore not, *a priori*, any fundamental reason justifying the fact that the apparent non-specificity of proteins in living organisms would be an object of surprise

[15] These predictions are, indeed, probabilistic inasmuch as they concern the *global* behaviour of populations of molecules and not the individual behaviour of each of these molecules, which remains *submitted to the perturbing influence of thermal agitation*.

and give way to interminable empirical and/or theoretical inquiries given that the hypothesis of stereospecificity is not a *sine qua non* condition of the validity of the analyses of statistical chemistry. Stochastic behaviours within structural stability may also explain genetic processes, even if they are largely incompatible with the notion of genetic program (unless one is talking of a very different notion of "program").

3.4.3 The issue of the developmental role of genes and cloning.

Finally, it would be difficult to provide a better illustration of the highly problematic character of the notion of genetic program than with the example of cloning, which is however often used to vouch for the *technical* mastery over the genome of species by molecular biologists.

The problem of embryonic development is indubitably the issue that has for the longest period of time held all the successive theories of heredity in check. In fact, the aspect of development which poses the clearest difficulties to the theory of genetic program is the fundamental fact that splitting cells are capable of differentiating themselves selectively before replicating themselves *more or less* identically, according to the developing tissue. If we insist on favouring the physiological role of genes with regard to the metabolic activities of the oocyte over the course of development, this process can also be presented as that of the *differential activation of genes*, in the formation of tissues and organs. The problem posed by this is nevertheless not insurmountable *a priori*, be it simply because there is no reason in principle why to exclude the existence of certain genes which promote the process of differentiation itself[16]. Today, however, we must acknowledge that *such genes have still not revealed themselves, except very locally*, in the case of the morphogenesis of some organs [Goodwin, 1985]. Even if such genes had a more general role, the question of *their primacy in the developmental process* as such would still remain no less problematic since it would preclude explaining that the process of differentiation is *irreversible*, indicating, therefore, the possibility of the existence of epigenetic modifications in the genetic material itself, which would reveal themselves to be "somatically hereditary", if such a notion can make sense. But, fundamentally, it is the fact that cloning requires the transfer of the nucleus of a differentiated cell in an *adequate* ooplast that most clearly suggests that the unfolding of the genetic program is not independent of the extragenic conditions and, particularly, of certain cytoplasmic constraints.

The first thing which appears to be necessary, practically speaking, would be to "reprogram" the nucleus of a differentiated cell in order to proceed to its implantation in an ooplast; in other words, to induce a complete restructuring of its chromatin. Thus, the success of Dolly's cloning by Ian Wilmut truly appears as the consequence of a notable enrichment of knowledge in terms of the developmental role of genes within suitable extragenomic contexts. And it is so because ultimately, we are, today, only discovering the very great complexity of the relationships between the receptivity of genetic material

[16] In fact, the difficulty here is that *from this sole point of view*, that is, if we only postulate, without empirical evidence or any theoretical framework, that each stage of the process of differentiation is dependent of certain combinations of genes, the program theory remains unfalsifiable, in the sense that the explanations of cellular differentiation will be at best circular, and, at worst, indefinitely regressive (regulating genes which regulate the expression of regulating genes…).

to certain cytoplasmic constraints and the epigenetic modifications of the expression of genes – and we could almost say that cloning is justified only to specifically elucidate these aspects. The birth of Dolly was, moreover, a surprise for everybody, including Wilmut and his team who, at the time, emitted the hypothesis that their success rested on a particular method of donor cell culture in the absence of albumin, this supposing to have placed them in a state of latency favourable to such modifications [Wilmut *and al*., 1997, Lewontin, 1997, Ahouse and Keller, 1997]. But, as we know, it later appeared that lack of albumin had absolutely nothing to do with their success since similar results have been obtained without the recourse to albumin privation.

4. The machine and DNA.

But then, is it still possible to believe that DNA is a program in the Computer Science sense of the term? As correctly stated by Danchin, it is question here *of a metaphor*, more than a "mathematical model", to be used as a suggestion for intelligibility, since DNA is not as human construction, practical or conceptual, as our digital machines. But when we project, upon natural phenomena, a human construction (the alphabet, clocks – for long, digital computers – more recently…), it is necessary to think, as we tried to do in §. 1 and 2, about the constitutive history of these constructions, very rich in human practices and internal logic which mark their meaning and their possible role for intelligibility. Metaphors are far from neutral: they project an understanding and even a Philosophy of Nature.

4.1 The rule, calculus and their context.

In the first two sections, we have highlighted the Laplacian and Cartesian roots of modern computing and we mentioned the ferocious dualism that characterizes it: Turing's great idea, as we mentioned, consists in the very clear mathematical separation between software and hardware. Here lies that which initiates the modern notion of program and the logico-mathematical theory of programming. This separation, purely conceptual in 1935, produced modern Computer Science, where software portability is at the center of any possible application, as is its perfect theoretical (but also practical) *iterability*. This property is a *local* one, in a given context (starting on a specific, local digit, exactly, allows perfect iteration of a computation), in contrast to the *global* nature of the *structural stability* of living beings and processes. Locally, life is always different: it is in the local description of the phenotype that one best appreciate variability, as core component of phylogenesis and ontogenesis. And this begins with the degeneracy phenomena at the bio-chemical level (see §. 4.2.1).

The central issue with a comparison genome-phenotype vs. program-semantics, or with intelligibility by metaphors, consists in the analysis of the various contexts of expression of the discrete-state machine or of the genome. Because, prior to any expression, there is something very important common to these two structures of the intellect and of the world: the essentially discrete nature of the notational, conceptual, or chemical support. Discrete in the following mathematical sense: discrete topology *naturally* organizes these structures (in the informal but rigorous geometrical sense). But *it is the operational context which generates the meaning, or the calculus, or even still the structures of living matter* (the phenotype).

We have already briefly summarized the thesis at the center of comparative analyses, between Physics and Computability in [Longo, 2002; Bailly, Longo, 2006]: computability is a theory of iteration, from primitive recursion (Herbrand, Gödel), up to software portability (a most remarkable form of iterability). Even in concurrent networks, of which the (difficult) mathematics is in the process of development, all is in place to command the certitude of iteration. Physical space and time, even of a relativistic type (Matherat, and Jaekel in [Aceto et al., 2003]), pose difficult problems to the synchronization of asynchronous concurrent processes. Yet, we all know that if we open a distant Web page a thousand times, we will obtain, except in rare cases, always the same page, identically, independently of the spatiotemporal access (and we become furious if things are not this way: the exactitude and independence of Physics, material and spatiotemporal, remain the objective of this extraordinary mathematical construction which is Computer Science). Programming, be it sequential or concurrent is intended for that: the rule *must* dominate, without exception. And if exception does occur, if interaction in the time and space of a network, not predicted by Turing, can, in rare cases, be of any hindrance, a science of concurrent processes is established to avoid it, to contain it. And this is effectively achieved, with the mathematical strength of the rules (the instructions of a program) which pile up upon rules.

Exactitude and iteration: the rule directs the calculus. The computer performs no calculus without rules, the program itself is a set of rules (Regulae ad directionem … calculi, Descartes would say). *Randomness does not exist in computability* (unless as a disguised, but relevant analysis of uncompressibility, a la Chaitin). Pseudo-random generators are small programs which generate sequences of 0s and 1s, *perfectly iterable under the same initial conditions*, but which provide good distributions of probability in the sequence spaces (they *seem* random, as they just "imitate" randomness and do not model it). Just as is for Laplace, also a great mathematician of probabilities, chance is perfectly foreign to determination by rule[17]. Now, since Poincaré, it has been understood that a deterministic (classical) process is random when, *iterated in the same border conditions, in the sense of physical measure, it does not undergo the same evolution*. This does not exist in sequential programming theory (of computability); in concurrent systems, it is rare. It is due to the physical (or human) context and it is avoided at all costs (and successfully so).

In fact, the practice of concurrent programming over distributed networks of computers (distributed in space and time, this is the novelty) essentially produces the same type of evolution as sequential processes from the standpoint of determination and predictability (and not the same set of processes!), as the data bases are *exact* (digital in a digital context). All is done (semaphores, interleaving, see [Aceto et al., 2003]) so that such be the case, despite the lack of a convincing (and unique) theoretical framework. The objective, largely met in practice, is to implement the imagined process in a manner essentially insensitive to initial or contextual variations below the level of observability, including those which present a theoretical problem in the conception of the piece of software; that is, those which are due to the concurrency-synchronization of processes distributed over physical time and space.

[17] This Laplacian split between normative law and randomness may be found identically in Monod and Jacob's approach to necessity and randomness.

Yet, some insist that the novel programming structures invented for concurrency fit the genomic patterns: high parallelism, concurrent processes, data flow, message-passing, term-rewriting, Petri Nets… the fascinating terminology of Computer Science seems to permeate theories of genes. However, one by one, it is possible to find counterexamples to the adequacy of these structures of information flows, as there is much more than the addition of space and time in moving from Turing Machines to genome. So data-flow is inadequate, as it doesn't support loops nor proteins' decay. If one looks at protein cascades as message-passing, their re-combination and decay seems far away from the computing analysis (no distinction seems possible between a message and a process, in genomics). Parallelism and communication is affected, in the cell, by omnipresent feedbacks, which are badly handled by synchronous and asynchronous automata; proteins processes are inhibited and this is not described by term-rewriting or Petri Nets; in operator algebras the gates' output is predetermined, it cannot be a function of the input, as it seems to be in molecular dynamics.

However, beyond these technical remarks, the issue concerns the general conceptual frame that computing, this extraordinary, alphabetic, cartesian, invention of ours which did not exist in Nature, forces upon natural phenomena. In spite of the growing trend of interaction and interactive processes in concurrency and alike, the underlying Philosophy of Nature still sees rules are *previous* to any digital process: there lies the classical, Newtonian (and Fregean) conception of *physical* (and logical) *law,* an absolute of nature (even of thought), embodied at last within the discrete-state machine because the law *governs* its course. And the exactitude of the digital universe guarantees predictability, at least in terms of iteration: the Laplacian prediction is possible in discrete-state machines, even if we have to work at it when they are distributed in space. And this, since, in a structure where there is only the discrete, the dynamics sensitive to variations below what is observable (the intended discrete topology) *are in principle excluded*: the digital data base fixes for ever the level of observability and the *theory* does not admit influences from anything below (which is extremely rare and to be avoided by all means). Thus, run twice the program which generates the strangest of Lorentz attractors and one will obtain exactly the same digital image, *this being absolutely impossible for a physical turbulence* – the fact is that *the causal structure has changed during its implementation upon the machine* [Bailly, Longo, 2006] (and if one time in a million times are not as such, measure the probability of these exceptions in relation to the non-iterability of natural turbulences! The addition of a radomness generator in the context of this discussion, where the aim is understading, is closer to a cheat than to scientific analysis from the perspective of causal intelligibility: a turbolence is causaly sensitive to border conditions, that is the theoretical issue).

What about this *programmed insensitivity to context,* specific to the digital machine, in the case of the genome? We doubtlessly find elements of a discrete structure in the sequences of nucleotides, these letters reminiscent of a four letter alphabet. But the role held by the context of expression in the analysis of data and the evolution of processes changes radically. First, being of natural phenomenality, it would be necessary to abandon, as we have said above, the classical conception of the laws of Nature: a normative conception, derived, since Aristotle up to Newton, from social and religious practices (civil and penal law, divine law). It is precisely this normativity which makes the strength of programming, but *it is not this way that Physics proceeds today when*

trying to make intelligible the phenomena surrounding us. The co-constitution of meaning and of scientific intelligibility are the result of a practice of knowledge which attempts to propose conceptual spaces (phase-spaces, adequate metrics), which integrate the knowing subject in an active subject-object polarity. Relativity and quantum mechanics proposed this type of way of seeing things, as mentioned above, and their revolutionary perspectives *changed the scientific approach*. The trajectory of a planet *is not* a course obeying Netwon's laws/equations, *but a geodesic within an adequate Riemanian manifold*. The evolution of a quantum system can be understood in Hamiltonian terms and in terms of symmetries. That the programming of a digital machine follows old paradigms is, on one hand, *intrinsic to its essentially artificial nature*. We are God to the machine, we dictate the rules. On the other hand, this is due to the *historical exclusion of new scientific paradigms* (the "delirium" of non-euclidean geometries and their physical consequences) *in post-Fregean analyses of the foundations of mathematics*, as if Mathematics were out of the world, away from physical time and space, grounded on the formal concept of Integer Number, whose theory should have proved its own consistency (Hilbert); these are the logico-formal foundations from which the invention of the machine is largely derived since the '30s.

Can we, despite all the above, transfer this Laplacian scientific paradigm (equational-formal determination implies predictability of evolution) to the framework of the living, particularly to that of genomics? Once again, the allure of the discrete is highly justified and we will return to this, as the notion of computer program (sequential, concurrent) is *much more than alphabetical writing*: it derives, we hope to have emphasized this enough, and contains/imposes a strong organization of the world and a firm logic, which find expression within the framework of formal laws of which the *iterative certainty*, even *the predictability, independently of the context*, is the main priority. And the discrete at the genetic level, which is surely there, within the chemical structure of DNA, *in which context does it find expression?*

4.2 DNA and its context.

At the nanoscale level, locus of its primary expression, DNA is submitted to a violent and rapid vortex (with activity variations, within the limits of the thermodynamic interval of viability): thousands of particles bombard its nucleotides. From time to time, it occurs that a nucleotide collides with a compatible base or polymerase, after thousands of clashes having resulted in nothing. But this compatibility, or even complementarity, can not be characterized by a clear yes or no: there are *degrees* of stereocomplementarity and gross geometrical correspondences can at times enable, by means of induced adaptation, a coupling between biomeres of which the compatibility can depend on the slightest variation of context, for example a thermal oscillation. At this level already, the "rule" is fuzzy, adaptive, *highly contextual*, based on a process of which the totality of its evolution matters as much as the local encounter. There exists no type of computer programming, even concurrent, that starts off in this turbulent fashion (thankfully so, given the reasons for which we have invented computers: programming is *teleologically oriented* and, for example, the same *physical* notion of "program attractor" would make no sense).

Let's continue. Sometimes, the duplication system associates, for instance, a G to a C; however, it is also possible that we obtain the GT coupling. This less frequent duplex

may remain, but it can also be mutated. It is not convincing to enframe these phenomena by stating that there are universal laws with exceptions, even frequent, because a little mutation of this sort, or another one yet unseen in the laboratory, can be at the origin of a phylogenetic tree… like the one to which we belong. It is better to say that to certain situations correspond a wide range of possibilities, a bit in the manner of geneticists who use the concept of "reaction norm" (a genotype can engender a range of phenotypes according to context) but all while stressing, from our point of view, that the list of possible phenotypes associated to a reaction norm is not given (programmed) *a priori* (at the onset of the evolution of species, insist the claimants of the intelligent design/ programming approach). They would rather be *possible cases* in a non-stationary framework, one would say in Physics, almost all of them, but not all, equivalent: the selection, local (within the cell), or global (ontogenetic, even phylogenetic), will sort those which are viable at a given moment of ontogenesis or phylogenesis. What a catastrophe it would be if we programmed in such a way: and if one day we invented, in Computer Science, a method of "Darwinian selection" of programs within a concurrent framework, it would be necessary to simulate the variability of the living, which precedes selection, with all its framework of individuation and degeneracy, which we shall address. Because variability, in phenomena relating to living matter, *is not a defect*; to the contrary, that which is very important, for living matter, *its physical singularity*, is that *never* is the cell identical to its parent cell. And this variability is the very *condition of possibility* of somatic as much as of phylogenetic selection and, therefore, of the development of individuals as much as of the evolution of a species.

In short, even in classical Physics, the rule, the law, is an instrument of intelligibility, it is not inscribed within the world nor in a pre-established program (except if we are still Galileo-Newtonian or if we are programmers): *it is of an epistemic nature*, the result of symmetries and geodetics in a co-constitute interface with phenomena, as we learned from Relativistic and Quantum Physics. In any case, in Physics, *a law having exceptions is false and must be discarded.* It would be absurd to give ourselves, as instrument of intelligibility, in biology, a notion of rule which would not enable us to understand a great part of that which matters. In fact, in biology, there are almost never any perfectly rigid laws, à la Newton-Frege-Turing, with exceptions, but there are rather *possibilities within a framework which is globally (relatively) stable, but not too much so.* The mobile margins of an attractor, of which the components would follow viable trajectories within the attractor, may provide a better image of the ontogenetic and phylogenetic processes, just like the margins of an extended criticality which we will address. But it is also necessary to add to these physical metaphors *the non-stationarity of the phase-space*, which form over the course of the evolutionary process itself – and this, in contrast to *all* current physical theories (see [Bailly, Longo, 2006]) [18].

Furthermore, in Computer Science, a different result from the predictable determination with regard to a programmed rule is simply an error. For this reason, for

[18] The elevators claim that 40% of fertilizations, in mammals, fail: what a flawed program the genetic one! But these failures, in general, are not "errors", "exceptions"... they are part of the phylogenetic/ontogenetic game of variability and selection, of which the theory should take care. With regard to "exceptions" in Computer Science, a notion which is present in programming and refers to deterministic bifurcations, devoid of physical criticality (with its associated fluctuations). The notion is therefore still of a Laplacian type – laws with foreseen exceptions.

instance, we do not trust computers to evolve by themselves and so we prefer to pay for new software[19].

How to make intelligible this singular mixture of global stability (the formed or forming living individual) and local instability, sometimes coupled to their dual: an extreme local rigidity (*some* steric complementarities are very rigid) and a global plasticity (that of an organ such as the brain, even over the course of adult life), as well as a global dynamic stability and cascading local variations? Even the physics of the most dynamic of systems seems insufficient: mathematical invariance is always the center and physical *variation*, so well acknowledged in its importance (the Lyapounov coefficients of some systems very well describe the role of minimal perturbation/variation, in time), is quite different to biological *variability*. The latter bases itself on *individuation*, a notion foreign to Physics, just as is the notion of pathology.

To conclude: individuation, variability, changing framework of stability, instability maintained within plastic limits (the pathological, the deviant may become a new evolutionary direction), the non-stationarity of the space of possible evolutions, with this stemming from the action of the genome, discrete sequence *immerged within a cell*, a quasi-fluid environment. All this therefore remains to be grasped, possibly by means of a complex game to be invented for the mathematically discrete vs. the continuum, with its own invariants and stabilities where functioning uses dysfunction. This is highly remote from a theory centred on the arithmetically discrete such as computability.

4.2.1 Degeneracy and the program.

Even this discreteness, so important to the phenomenon which is living matter, has a rather singular property, for computability and even for Physics, that we already mentioned in several places: *degeneracy*. This notion was introduced by Edelman and Tononi, and revisited by other authors [Edelman, Tononi, 2000 ; Edelman, Gally, 2001] with regard to the functioning of the brain; it has its origins in the works of Edelman concerning the immune system. Shortly, a structure is degenerative if non-isomorphic sub-structures can give place to a same functionality and if a given structure can find expression in several functionalities. This concept differs from *redundancy*, which is well known to information theory (a theory of communication [Shannon, Waever, 1975]), as well as in Computer Science. From this perspective, we can make a distinction between *"functional" degeneracy* (of non-isomorphic systems participating to a single same function) and *"systemic" degeneracy* (a same system participating to distinct functions), see [Bailly, Longo, 2006]. In [Edelman, Gally, 2001], it can be observed that degeneracy is *ubiquitous in systems of living matter*, starting with the genome. Even for a discrete structure, it is at the origin of a very specific phenomenality. It is necessary to consider, also, that in biology, when one says "same" (function, for instance), it absolutely does not mean "perfectly identical", as can be said of a mathematical object, of a function calculated by a program, a program, a sequence of numbers… For biological degeneracy, non-isomorphic sub-structures (although "analogous", in a sense to be specified

[19] The fantasies of classical artificial intelligence have constituted in this respect one of the most momentous wastes of money in the history of science: we could invoke numerous projects, form 1956 in the USA up until the Japanese "5th generation"; would molecular biology find itself upon the same path, once again rooted in the myth of the rule, of the program, of the Laplacian calculus which computes/ engenders everything?

according to context) or a given structure generate *almost* the same function, within a similar context. A small fluctuation of the context, even of physical origin, can then generate in time an expression, a function for instance, which is quite different. This causal analysis of the variability of living matter is compatible, but considerably enriches that based on the notion of mutation (both have no meaning for programming theory and are seemingly very seldom used in practice… thankfully).

To conclude, far from providing clear insight into living phenomena, the notion of genetic program, basing itself on the hypothesis according to which there would exist specific (rather than degenerate) molecules capable of regulating the synthesis of proteins necessary to the cellular metabolism, has encountered, in practice, numerous difficulties, some of which having been recalled here and in § 3.

Specifically, from a standpoint that one would readily qualify as nominalistic, it is not very convincing, in the context of a scientific discipline, to always recourse to explanations consisting of a simple description of the observed phenomena and which are interpreted (even theorized) using borrowed terms, problematically, from other disciplines without, at the same time, having imported their methodological rigor. This because these explanations then present the major flaw of multiplying the number of hidden co-factors (hidden variables?) from the moment that an experiment appears to no longer readily accord itself with the model's prescriptions.

We have, in this sense, already emphasized that, account taken of these elements, the notion of genetic program, far from having reinforced the role of genes in biological processes, has very seriously eroded the very possibility of providing it with a univocal definition in terms of genetic information. And this because genes can no longer be without ambiguity associated to a succession of nucleotides encoding a protein, and because *even if such was the case*, the succession of amino-acids within the given protein would not enable to determine the function associated to it, *unless referring to the constraints which will be applied at the cytoplasmic level*. And, in fact, it does appear today that it may be impossible to do without a reference to the epigenetic and extragenic contexts within which the genome finds expression, with all their immense biological complexity, in order to give a minimally coherent interpretation of them [Allis and Junuwein, 2001, Boguski and Hieter, 1997, Turner, 2002].

4.2.2 The differential method from Physics to Programming.

In our critique of the differential method in molecular biology (§ 3.2.2), it had been question of a theoretical incompleteness with regard to the inference of "causal laws". Specifically, we have demonstrated the fact, important to us, from a purely logical point of view, that simple empirical correlations between certain modifications of DNA and the differences observable in the phenotypes of individuals do not immediately and/or evidently provide a "law" which causally correlates the various structures of DNA to their "consequences" deemed to be normal. It is necessary to emphasize here, once again, the specificity of the differential method in Physics, from the moment that it is a question of formally establishing any causal dependency between variables. Indeed, beyond the analyses of the correlation/decorrelation of variables which we have already addressed, physicists generally need to endow themselves with a theoretical framework of interpretation, essential to the construction of scientific objectivity.

Firstly, the finite number of experiments and, further on, the discrete character of the data does not enable physicists to propose a *unique* equation otherwise than by means of a certain "conceptual optimality": in other words, we are looking for the equation (the polynomial, the curve…) having the "simplest" mathematical form. It is obvious that the validity of this criterion *is not an absolute* but *its effectiveness* is at the center of the methodological strength of Physics, since Copernic, Kepler and Galileo (and more explicitly, Lagrange and Hamilton). Secondly, the differential method in Physics finds, in a large measure, its mathematical justification in the variational methods of differential calculus: small (infinitesimal) perturbations enable the reconstruction of geodesics. In both cases, we can see, the mathematical analysis presupposes a continuous framework – smooth spatiotemporal (even conceptual) "surfaces" – where these criteria of optimality have a strictly physico-mathematical meaning and coherence.

Now, this framework is precisely lacking in biology's differential analyses, probably because of the huge difficulty of sorting out, due to their physical singularity, the "interlinked" causal relationships characterizing living matter, which appear to us to be actually stemming from Physics' most dynamic theoretical frameworks. And in no way does the "model" (the metaphor) of program provide such a framework by way of its causally Laplacian computational roots. We have already recalled, in fact, that the latter find their origin in a theory of computability (sequential or, for the little that exists, of concurrency) of which the consequences for causality regimes have been analyzed elsewhere [Bailly, Longo, 2006] and are very removed from the smoothness of the physico-mathematical theories which are implicitly referred to in biology, by proposing a differential analysis largely unfit to Computer Science, as argued further below. In short, the "flaw" in terms of explication which we have noted in genetics does not stem from differential analysis itself, a technique which has given us, throughout the XXth century, information and results of great interest, but it must, in our opinion, be attributed to the sole attempts to frame it within more or less naïve notions of "genetic programming".

As already mentioned, in Physics strong and explicit principles (symmetries, geodetics) justify the theoretical proposals of which the differential experiments are the counterpart. In spite of the presence of physical constraints in life phenomena, living processes do not seem to rely *only* on similar principles (symmetries and geodetics), which of course participate to it (think of phyllotaxis); so the differential methods can hardly rely *only* on similar theoretical background. The point now is that the Computer Science theory and practice cannot help in this regard: there is no way to analyse the behaviour of a program by forcing small changes and... see what happens. Let's be more precise.

First, a modification in the order of the instructions of a program does not necessarily modify the result of a calculus despite that there are strong reasons, namely within the paradigmatic framework of Turing machines, to speak of a relationship of causality with regard to the relationship between programs and calculi (see §. 2). And, following that, it is clear that in logic and Computer Science, the analysis of the *semantics* of deduction or of programs (the "meaning" of the theorem or the "function" that is computed) would not make do with a simple differential, or perturbation method, to establish a relationship of causality between deductions or programs and deduced meaning or calculated functions.

A programmer may be willing to experiment "local mutations", by changing here and there one or more instructions or part of them, while looking for a variant of a known

program. But in no way one would use such a practice to understand the "meaning" of (the computed function by) a given well-formed program: the rare, but possible practice above leads to no theory. The nowadays rich semantic theories of programs are based on completely different mathematics. The major streams are called Denotational Semantics (see [Amadio, Curien, 1998]) and it is largely based on categorical meaning of logical theories, [Asperti, Longo, 1991], or the Abstract Interpretation and a few more.

There nevertheless exists a type of analysis, in logic and programming theory, which may resemble a differential method: the "Böhm-out" technique [Barendregt, 1984]. The theorem at the basis of this technique enables to demonstrate that a *difference* between two programs, as terms of Church's lambda-calculus (in their so-called Böhm-trees, at a finite level) induces a difference in their semantics, *given* very specific operational semantics or certain mathematical domains of interpretation (or, at least and more technically, it demonstrates that two terms in normal form, which are syntactically different – even for a very small "mutation" - do not calculate the same function, *in any already given semantic environment*, i.e. in any mathematical meaning of programs). But this result, which may be considered of the "differential" type, *does not* provide the semantics of a program (or of a lambda-term), its "phenotype", if we may say so. It may instead help to characterize syntactically the terms having *the same semantics*, within an already given mathematical structure [Barendregt, Longo, 1980].

More so than in Physics, therefore, a coherent and sound, conceptual interpretative framework must first be given (with its metrics or topology, its perfectly clear mathematical interpretation… a sort of framework for signification-determination comparable to that of Statistical Physics for thermodynamics, to recall the analogy in §. 3.2.2), and *then* the analysis of the equality between programs is refined by means of a differential method. Once more, the analysis of variations contributes to better make explicit the generated meanings or even processes/functions/"phenotypes", once a direct and sound interpretation *has been given*; in short, that which signifies/calculates a program very generally. Once more, then, *the notion of genetic program does not enable to account for the relationship between genes and characters by means of a sole differential analysis*, because it lacks a determination of this relationship in the *direct* sense (for "normal" genes or programs, before "mutations" or Böhm-out) as it is given in semantics theory of programs.

4.3. More on causal structures and finalism.

We can ask ourselves what pushes so many biologists, of the highest scientific level, towards this myth of genetic program. It is clear that the discrete structure of DNA, so well described by sequences of letters that do not fail to be reminiscent of the encoding of a formal language, suggests such an analogy by means of a metaphor, as a convenient representation. However, that does not suffice to justify in itself the reference to the notion of program as such. The need, so strong in biology, to find a justification (or an allegedly reductionist explanation) to the *finalism* of processes relating to living matter does transpire here all the more clearly. By means of language play around the formal notion of program, we indeed surreptitiously slide towards a conception of living matter which is more or less permeated with finalism (that thing which we try to hide, but which we cannot do without…). And here, salutary, comes the notion of program, a secular

notion which, without needing God, contains *an end* within its lines of code so nicely embodied in our modern machines. The computer program is *made for*…, it has an objective, but it is, or can be, materially encoded. However, we do not know of a working computer program which has not previously been entered via a keyboard: is it God then who types on the molecular keyboard? Surely not: it's Evolution, one may say. Like a Deus ex-machina, evolution is introduced here to fill the explicative shortcomings of the program metaphor itself, by apparently making intelligible the encoding within matter of a project for the future… At least, if we exclude the question, highly problematic from a strictly physical standpoint, of the origin of genetic material, inasmuch as it constitutes the condition for the possibility of this very inscription.

One may say that maybe we are using trickery, because it has only been here a question of language. The problem, nevertheless, is that this language tends to impose itself as *the* dominant paradigm within the scientific community, including for that which concerns the formulation of research projects which, it appears, should still remain reasonably receptive to the possibility of a paradigmatic failure (we ask ourselves, indeed, how could science progress if it was otherwise…). We seek, for instance, mostly and since a long time, the cause of cancers in the mutations of suddenly badly programmed DNA and it has only been a few years that the attention of some audacious individuals has shifted *also* towards the context of expression of genomes, [Sonnenschein, Soto, 1999].

But let's return to the issue of finality, by trying to approach it from a possible physicalistic standpoint. Let's consider a boulder forming a slope endowed with a set of given irregularities and upon which rain gradually forms a cavity, by the repeated action over time of friction and small turbulences. It may happen that this first cavity will, in turn, generate a new turbulence which will also cause a second cavity to appear, a "child", of sorts, of the first one. But then, would it really make sense to say that the first notch is a *program*, written by physical evolution, in order to generate a second one? Or further, that it contains *information* about the form of future notches with their apparent structural specificity? With this example, we clearly see that it is *always* possible, from a simply descriptive point of view, to "manufacture" an apparent finality by playing the rhetoric game of programming or by employing the notion of information, so laden with meaning within Aristotelian culture; but does it make the phenomena thus described *physically* intelligible? Modern Physics has chosen a different approach to explain this kind of phenomena.

Up until the first half of the XIXth century, many authors indeed saw in the criteria of optimality of physical trajectories a form of finality in actuality within the sciences of Nature: the low-pitched or the light was thus reputed to "choose" the optimal trajectory to "go towards...". It is only with the advent of variational methods that this type of reasoning was finally discarded to the benefit of explications within which the geodesics where induced (caused) by the structure of space itself (physical space or, more generally, phase-space). Over, then, with finalism for Physics… Yet, still now many claim to answer the question of finalism by integrating it to the description of its objects (the DNA *is* a *program*, that is it is "programmed for"), all the while claiming to be physicalistic, in a way which is, in our view, rather paradoxical. But the issue of the eventual finality specific to living organisms extends, in our opinion, well beyond that of a possible relationship between genes and characters, even simply at the molecular level.

It should be dealt with general principles (comparable to the geodetic principle in Physics) or adequate conceptual structures, proper to Biology (see [Bailly, Longo, 2006] for a discussion).

It would then not seem extravagant, from this standpoint, to ask ourselves whether a study of the properties of biological systems employing the methodology of Physics should refrain from accepting a minimal form of finality, despite that it does exist. However, it would indubitably be wiser to reserve this inevitable and constantly renewed reference to finalism in biology to the probably least accessible road to knowledge, that of the original formation of living matter; and this, precisely, to use an expression often used by computer scientists, by *lack* of knowledge[20]. But for the rest, especially if one claims to maintain an authentically physicalistic perspective, it will be necessary to leave aside this obsessional reference to the finality which is implicitly related to the term 'program' and to that of 'gene', identified to these "notches" in DNA, its chemical structure! The vascular system is, from this standpoint, as finalized as the canalization which forms progressively by accumulation of notches on a boulder, with the difference, however, conversely to the geodesic canal bored by water in a pre-existing physical structure (a phase-space), the vascular system forms *at the same time* as the rest of the body… And it is here that is added the material memory which is DNA *with* its cellular context; a memory which appears to constitute, by means of frictions repeated all along the course of phytogenesis and ontogenesis, a characteristic specific to biological processes, totally removed from usual physical dynamics.

To conclude, let's replace the boulder's physical slope (or even the *cause* of the slope, perfectly unknown to biology), with the sole contingent finality of survival, and let's try to construct conceptual spaces, as much as possible of the physico-mathematical type, or better, purely biological, in order to produce an intelligibility which avoids the insertion of an end, of the Computer Science type, into any piece of DNA. Understanding this specificity is indubitably one of the greatest future issues for biology and will certainly bring forth the emergence of theories on the fringe of existing physical theories, as we suggest in the conclusion.

Conclusion

The complexity of the elementary components of natural phenomena remains a great scientific issue of our day. Firstly for Quantum Mechanics, with the non-locality and non-separability of elementary particles, with "strings" and ten-dimensional theories of which six are compactified in order to grasp that which is elementary, but also for Biology, where the complexity of the cell, necessary context of the expression of DNA, is

[20] Within the context of the huge debate concerning finalism in biology, and in this search for minimality, we dare refer here to that of which we are talking about in [Bailly, Longo, 2006]. We are thinking about a finality which is "contingent" because physical, non-programmed, and which could be otherwise than it is, or not be at all; a finality which constitutes, moreover, the implicit foundation of any discourse on living organisms, that is, *survival* (of the individual, of the species). From the constitution of the first living structure, indeed, an evolutionary passage among the least understood, what has counted in this novel and singular physical formation and what has definitely differentiated it as such from the remainder of inert matter – even though, moreover, it was not foreseen at the onset nor inscribed within a project - is nothing else than its ability to survive. Otherwise, this matter could precisely not be considered as living, or would no longer be there.

probably mathematically infinite with regard to any physical complexity measure. The historical recourses to a mechanistic theory which attempts to grasp that which is complex by reduction to that which is elementary and *simple* have had their day, in these disciplines. All our mechanical engines, all our artificial constructions, having been designed from "bottom-up", are produced by composition of the *simple* elementary (clocks, computers, programming languages… programs). It is in all likelihood not the case for structures of which the natural history is rich in back-tracks, interactions between various levels of organization, where the elementary is dynamically co-constituted, resulting from very complex global units, and thus, in principle, very complex itself. Embryogenesis, typically, is a "top-down" process: all begins with a cell, a complex organic unit, which differentiates itself (and besides, this complexity of the elementary for Biology is *the* condition for the possibility of the processes of differentiation and individuation themselves).

But even the most deterministic of physical systems, our planetary system, *is not a huge clock*, as has been believed for a long time (despite that Newton and Laplace doubted this), neither it is run by a predictably deterministic program (iteratable). Its *essentially chaotic* nature ([Laskar, 1992 ; 1994]) shows that the global equational determination, which corresponds to the geodesics of spacetime, does not suffice to grasp its evolution in causal terms; it is in fact *causally* sensible to variations/perturbations below any theoretically possible measurement. It is therefore the elementary of local phenomena – which can be very complex – that interferes with global evolution. And in this case specifically, a theory of perturbed geodesics enables us to understand something; while a vision in terms of rules or of programs, if it can lead to good computational *imitations* with their own contribution in terms of intelligibility, must nevertheless be fundamentally distinguished from an explicative and/or predictive mathematical modelization in the most strict sense of the term (see [Longo, 2002] for the distinction between computational imitation and mathematical modelization, also implicit in [Turing, 1950 ; 1952]).

Biologists should dare to clamor *the specificity of their theoretical needs*. Science proceeds firstly *by differentiation of phenomenal fields*. Darwin did not take ideas from the physical theories of the time, even less from the highly sophisticated mechanics of his day, but proposed (following others also: ideas always have a history) an absolutely novel theory – at most, with some sociological contaminations, some would say. What matters here is that with his work of theoretical reflection, *he distinguished the theory of living organisms from all existing physical theories*. We can see here an analogy with another great theoretical moment, within Physics itself, which broke the so-called unity of late XIXth century's microphysics and astrophysics, supported by the proposition of a planetary model of the atom: the invention of Quantum Mechanics. Some physicists had the audacity of saying: no, the structure of determination which we propose is *radically, even irreducibly, different*. And the central linchpins of physical intelligibility have thus been reversed: no locality, no separability, no trajectories (!), the field is linear.

Biologists should do as much, a little bit like Darwin (and we can see signs of it, here and there). A clear conceptual separation aids to grasp a very difficult phenomenality; *afterwards*, we can aim for unification, language bridges and for the logical derivation of theories, relating (unifying!), typically, the Chemistry of macromolecules to a theory of the living unity of the cell. But we will not be seduced, in the meantime, by misleading

metaphorical analogies, endowed with a strong history and an autonomous logical structure. In Physics, theoretical separation does not prevent us, almost 100 years later, from perceiving some elements of a possible unification between quantum and relativistic fields (actually, *two* possible unifications, so far, apparently incompatible between themselves…); but, in the meantime, Quantum Physics, with a very particular intelligibility and numerous interpretations, will not have refrained from changing the world by its efficiency nor from changing scientific culture by *its methods and its explicative autonomy*.

The comprehension of the discrete spatial structure of DNA is one of the great scientific breakthroughs of the XXth century, but its activity within a highly complex and dynamic context *remains to be understood*. Our first impression persists: in the absence of a strong and autonomous theory of the living cell as an organism, with all its levels of organization, a theoretical specificity which must be thought of as both *dynamic and physically singular*, unknown to current physical *theories*, it will be difficult to provide ourselves with a framework of intelligibility comparable to that of Physics. Once again, we are not putting into doubt that there is only "physical material" in the world, but we believe that current physical *theories*, with their structures of determination (particularly the causally Laplacian theory of programming), are not adequate to make intelligible the organized matter that interests Biology, just as classical and relativistic dynamics provide little understanding of microphysics. Without a radical change of point of view, there will always be insurmountable difficulties in grasping the place occupied, be it at the level of the cellular metabolism, of individual development or of the evolution of species, by this discreted trace of the history of living beings (and of no other of the systems addressed by physical theories) which is DNA *with* its context of expression.

Concerning determination and causality, one avenue among many other possible (and proposed) ones is fascinating to us and we are working at it. The issue of the correlation length between variables (and their interactions with observables), which, in a living organism seems to attain the dimensions of the organism itself, could possibly be analyzed at the edges of specifically physical phenomena, as we attempt to do in [Bailly, Longo, 2006]. It would be a question of analyzing the dynamics of living phenomena as *extended critical situations*. In fact, physical criticality enables to grasp the passing from the local to the global, for instance in phase transitions. It describes changes by means of passages which are isolated points of the physical control parameter (instantaneous if the time parameter is at stake), and where global correlations establish themselves between all components of the phenomenon, producing a new unity; this new "coherence structure", with regard to the previous scale, gives rise to mathematically infinite measures. In Physics, then, one says that some values diverge beyond the physically observable/measurable at the intended scale: they go to infinity, as in a mathematical singularity. These physically diverging values may acquire meaning for Biology, if they help us to grasp the situation of a critical but *extended* unity, the cell, the organism, extended in time and space, *far from equilibrium*, stable and unstable, or even preserved within the frame of stability by this unity of correlation (of auto-organization) which is *physically* implausible, because mathematically infinite. One of the technical difficulties would reside in the intervention of sound renormalization techniques, with several (an infinity of?) parameters, far beyond current physical theorizations. The immediate advantage, with regard to that which we have been discussing, would consist in a possible

intelligibility of the role of the slightest variation in this discrete component of heredity, DNA, in the induction of enormous changes at the phenotypical level: the effects of resonance and intrication specific to the correlation lengths of extended criticality would justify these changes of causal scale, without conferring a role of causality, in the sense of Laplacian determination, even "direct and whole" to DNA, as a program[21] . It may also be possible to better understand along these terms the genetic notions of mutations, of pleiotropy, of polygenesis or of epistasis: it is the activity of the organism (cell, metazoan) in its critically extended unity, its organized *action-reaction*, which would provide them with meaning.

Acknowledgements Very instructive discussions and gentle disagreements with Antoine Danchin spurred the writing of this article. F. Bailly, J.-J. Kupiec, M. Mossio, B. Saulnier and J. Stewart have been essential to the redaction process, thanks to their extended and... critical remarks.

References

(Longo's articles are "*downloadable*" *de* "*http://www.di.ens.fr/users/longo*" ou Google: search : Giuseppe Longo).

Aceto L., Longo G., Victor B. (eds.) **The difference between Sequential and Concurrent Computations**. *Special issue,* **Mathematical Structures in Computer Science**, Cambridge U. Press, n. 4-5, 2003.

Ahouse J. C. et Keller E. F., "Writing and Reading about 'Dolly'", **BioEssays**, 19, 741-742, 1997.

Allis C. et Jenuwein T., "Translating the Histone Code", **Science**, 293, 1074-1080, 2001.

Amadio R., Curien P.-L. **Domains and lambda-calculi**, Birkhuaser, 1998.

Apter M. J., **Cybernetics and Development**, Pergamon, 1966.

Aristote, **Physique, Paris, 1862**

Aristote, **Traité de l'Âme,** édition numérique www.clerus.org, 2007.

Asperti A., Longo G. **Categories**, **Types and Structures**, M.I.T. Press, 1991

Atlan H., **L'Organisation biologique et la théorie de l'information**, Hermann, 1972.

Atlan H. et Koppel M., "The Cellular Computer DNA ; program or data", **Bulletin of Mathematical Biology**, 52, 335-348, 1990.

[21] And that would be comforting with regard to the sole 1% of difference, difficultly interpretable within genetics' dominant paradigms, which seems to exist between the DNA of certain great simians and that of this note's reader. In fact, we rather rely on the sensitivity-to-context and structural stability specific to the extended criticality of our reader, as well as to his or her extragenic richness and epigenetic experience.

Bailly F., Longo G. **Mathématiques et sciences de la nature. La singularité physique du vivant**. Hermann, Paris, 2006.

Barendregt H. **The lambda-calculus: its syntax, its semantics**, North-Holland, rev. edit., 1984.

Barendregt H., Longo G. "Equality of lambda terms in the model T," *in* **To H.B. Curry: Essays in Combinatory Logic, Lambda-calculus and Formalism**, (Seldin J., Hindley R. eds.) Academic Press, London, pp. 303-337, 1980.

Bartel D., "MicroRNAs : Genomics, Biogenesis, Mechanism and Function", **Cell**, 116, 281-297, 2004.

Beadle G. W. et E. L. Tatum E. L., "Genetic Control of Developmental Reactions", **American Nauturalist**, 75, 107-116, 1941.

Bernard-Weil E. "Ago-Antagonistic systèmes" *In* **Quantum Mechanics, Mathematics, Cognition and Action.** (M. Mugur-Schachter ed.), Kluwer, pp. 433-463, 2002.

Bernot A., **Analyse de génomes, transcriptomes et protéines**, Dunod, Paris, 2001.

Beurton P.-J. et al, **The Concept of the Gene in Develoment and evolution : Historical and Epistemological Perspectives**, Cambridge University Press, 2000.

Bjedov I., Denamu E., Gerard B., Matic I., Tenaillon O., Rardman M., Souza V. et Taddei F., "Stress-induced Mutagenesis in bacteria", **Science**, 300, 1404-1409, 2003.

Black D. L., "Splicing in the Inner Ear : a familiar tune, but what are the instruments ?", **Neuron**, 20, 165-168, 1998.

Boffelli D., Nobrega M. et Rubin E., "Comparative genomics at the vertebrates extremes", **Nature Reviews Genetics**, 5, 456-465, 2004.

Boguski M. et Hieter P., "Functional Genomics : it's all about how you read it", **Science**, 278, 601-602, 1997.

Bonner J., **The Molecular biology of development**, Oxford University Press, 1965.

Brenner S., Dove W., Herskowitz I. et Thomas R., "Genes and Development : molecular and logical themes", **Genetics**, 126, 479-486, 1990.

Brett D., Pospisil H., Valcárcel J., Reich J.. Bork P. "Alternative splicing and genome complexity". *Nature Genetics* **30**, 2001.

Burnet F. M., **Le Programme et l'Erreur** (1978), Albin Michel, 1982.

Campbell K. H. S., Kind A. J., Schnieke A. E., McWhir J. et Wilmut I., "Viable Offspring Derived from Fetal and Adult Mammalian Cells", **Nature**, 385, 810-813, 1997.

Cappuccio M.. "Traces of computational mind : From wax tablets to Turing machine", **Géométrie et cognition**, (Longo ed.), Editions rue d'Ulm, 124, s.5, 43- 60, 2003.

Chapouthier G. et Matras J.-J., "La Néguentropie : un artefact ?", **Fundamenta Scientiæ**, 2, 141-151, 1984.

Collectif, **Biologie moléculaire de la cellule**, Flammarion, 1995.

Collectif, **Le Concept d'information dans la science contemporaine**, Cahiers de Royaumont, Minuit-Gauthier-Villars, 1965.

Creager A. N. et Gaudillière J.-P., "Meanings in Search of Experiments and *Vice-versa* : the invention of allosteric regulation in Paris and Berkeley, 1889-1968", **Historical Studies in the Phisical and Biological Sciences**, 27, 1-89, 1996.

Crick F. H. C., "On Protein Synthesis", **Symposia of the Society for Experimental Biology**, 12, 138-163, 1957.

Crick F. H. C. et Watson J. D., "Molecular Structure of Nucleic Acids", **Nature**, 171, 737-738, 1953a.

Crick F. H. C. et Watson J. D., "Genetical Implications of the Structure of Deoxyribonucleic Acid", **Nature**, 171, 964, 1953b.

Danchin A. **The Delphic Boat: What Genomes Tell Us**, Harvard Univ. Press, 2003.

Eddy S., "Non-coding RNA Genes and the Modern RNA World", **Nature Reviews Genetics**, 2, 919-929, 2001.

Edelman G. M. et Gally J. A., "Degeneracy and Complexity in biological systems", **Proceedings of the National Academy of Science**, 24, 13763-13768, 2001.

Edelman G., Tononi G. **A Universe of Consciousness. How Matter Becomes Immagination**, Basic Books, 2000.

Enard W. *et al*., "Intra- and Inter-specific Varaition in Primate Gene Expression Patterns", **Science**, 296, 340-343, 2002.

Feingold J. et Serre J.-L., **Génétique humaine, De la transmission des caractères à l'analyse de l'ADN**, INSERM-Nathan, 1993.

van Frassen B. **Lois et symetries,** Vrin, Paris, 1994.

Frege G. **The Foundations of Arithmetic**, 1884 (english transl. Evanston, 1980.)

Gilbert W., "Why Genes in Pieces ?", **Nature**, 271, 501, 1978.

Godell B., Gouyon P. H., Maynard-Smith J., Radman M., Taddei F. et Toupance B., "Role of mutator alleles in adaptative evolution", **Nature**, 387, 700-702, 1997.

Goldschmidt R., **Physiological Genetics**, McGraw-Hill, 1938.

Goodwin B., "What are the Causes of Morphogenesis ?", **BioEssays**, 3, 32-36, 1985.

Goubault E. (ed.) **Geometry in Concurrency**, *Special issue,* **Mathematical Structures in Computer Science**, Cambridge U.P., vol.10, n.4, 2000.

Gros F., **Les Secrets du gène**, O. Jacob/Points-Seuil, 1991.

Havelock, E. A. **Origins of Western Literacy**. Toronto: Ontario Institute for Studies in Education, 1976.

Herrenschmidt C. et al. (eds), **L'Orient et nous**, Albin-Michel, 1996.

Jablonka E. et Lamb M., **Epigenetic Inheritance and Evolution**, Oxford University Press, 1995.

Jacob F. et Monod J., "Gènes de structure et gènes de régulation dans la biosynthèse des protéines", **Comptes rendus de l'Académie des Sciences de Paris**, 349, 1282-1284, 1959.

Kupiec J.-J., "A Chance-Selection Model for Cellular Differentiation", **Cells, Death & Differentiation**, 3, 385-390, 1996.

Kupiec J.-J. et Sonigo P., **Ni Dieu, ni gène**, Seuil, 2000.

Laskar J., "La stabilité du système solaire" *in* **Chaos et déterminisme** (Dahan E. et al. eds), Seuil, 1992.

Laskar J., "Large scale chaos in the Solar System", **Astron. Astrophysics,** 287, L9 L12, 1994.

Lewontin R., "The Confusion over Cloning", **New York Review of Book**, octobre 1997.

Longo G. "Laplace, Turing et la géométrie impossible du "jeu de l'imitation" : aléas, déterminisme et programmes dans le test de Turing". Conférence invitée, Colloque **Cognition, meaning and complexity**, Univ. Roma II, June 2002; paru dans **Intellectica**, n. 35, 2002/2.

Mayr E., "Cause and Effect in Biology", **Science**, 24, 1-14, 1959.

Matherat P., Jaekel M.-T. "Concurrent computing machines and physical space-time", *special issue,* **Mathematical Structures in Computer Science**, Cambridge U.P., vol.13, n. 5, 2003.

Mendel G., "Recherches sur des hybrides végétaux", traduction de A. Chappellier, **Bulletin scientifique de la France et de la Belgique**, 41, 371-419, 1907.

Morange M., **Histoire de la biologie moléculaire**, La Découverte, 1994.

Morange M., **La Part des gènes**, Odile Jacon, 1998.

Morgan T. H., **The Theory of the Gene** (1926), Yale University Press, 1928.

Morgan T. H., **Embryologie et Génétique** (1934), traduction de J. Rostand, Gallimard, 1936.

Nijhout F., "The Nature of Robustness in Development", **BioEssays**, 24, 553-563, 2002.

Nouvel P. “Modèles et métaphores” *dans* **Enquête sur le concept de modèle**, Nouvel P. (ed.), Presses Univ. de France, 2002.

Ong W. J**. Orality and Literacy: The Technologizing of the Word**. New Accents. Ed. Terence Hawkes. New York: Methuen, 1988.

Pentris S., Tooze J., Hunt T. (eds.) **DNA makes RNAmakes Protein**, Elsevier, 1983.

Portin P., "The Concept of the Gene : short history and present status", **Quarterly Review of Biology**, 68, 173-223, 1993.

Shannon C. E. et Weaver W., **Théorie mathématique de la communication**, traduction de J. Cosnier, G. Dahan et S. Economidès, Retz-C. E. P. L., 1975.

Shroedinger E., **What is life?** , Cambridge U. P., 1944.

Simpson G. G., **The Major features of Evolution**, Harvard University Press, 1953.

Sonnenschein C., Soto A., **The Society of Cells**, Bios Scientific Publishers, Boston, 1999.

Stewart J. **La vie existe-t-elle ?** Vuibert, Paris, 2004.

Sturtevant A. H., **A History of Genetics**, McGrawn-Hill, 1965.

Tautz D., "Redundancies, Development and the Flow of Information", **BioEssays**, 14, 263-266, 1992.

Thomas J. H., "Thinking about Genetic Redundancy", **Trends in Genetics**, 9, 395-399, 1993.

Turing A. "Computing Machines and Intelligence", **Mind**, LIX, 1950 (page references to its reprinted version *in* M. Boden ed., Oxford Univ. Press, 1990).

Turing A. M. "The Chemical Basis of Morphogenesis" **Philo. Trans. Royal Soc**., B237, 37-72, 1952.

Turner B., "Cellular Memory and the Histone Code", **Cell**, 111, 285-291, 2002.

de Vries H., **Espèces et variétés, leur naissance par mutations**, traduction de L. Blaringhem, Alcan, 1909.