ORIGINAL ARTICLE

Information and Causality:
Mathematical Reflections on Cancer Biology

Giuseppe Longo [a, b]

[a] *Centre Cavaillès, République des Savoirs, CNRS, Collège de France et Ecole Normale Supérieure, Paris, France*

[b] *School of Medicine, Tufts University, Boston, USA.*

Corresponding author: Giuseppe Longo giuseppe.longo@ens.fr; http://www.di.ens.fr/users/longo

Abstract

Theorizing and measuring radically change in physics when using discrete vs. continuous mathematical spaces. In the 20th century, Quantum and Information Theories brought into the limelight the use of discrete observables and parameters (energy states, spin-up and down, 0 and 1 …). In these theories, the reference to discrete measurable values fundamentally contributed to knowledge construction; it similarly leads to dramatic consequences in life sciences, in particular when biological dynamics are identified with information processing. Following an early debate in physics, we briefly analyze the origin and the nature of the bias thus induced in biology, in particular in relation to the understanding of causality. We show how strong consequences have been derived from vague, common sense notions and then stress their role in cancer biology. Finally, we summarize new theoretical frames that propose different directions as for the organizing principles for biological thinking and experimenting, including in cancer research. Cancer is then viewed as an organismal, tissue-based issue, according to the perspective proposed in (Sonnenschein & Soto, 1999; Baker, 2015).

Keywords: continuous vs discrete modeling; causality in science; transmission of information; randomness in biology; Somatic Mutation Theory; Tissue Organization Field Theory

1. Introduction[1]

Computational virtuality is heavily affecting common and scientific knowledge. The new symbolic forms of interaction on electronic digital networks provide extraordinary new tools for humankind, from everyday worldwide exchanges to scientific modeling. Yet, they also suggest an image of the world disposed to a peculiar bias. It is in biology that the reference to informational, alphanumerical data structures has had the greatest impact throughout the second half of the twentieth century, by making DNA an "information carrier" or even a "computer program" for ontogenesis. As a consequence, development has been interpreted as the deployment of a program and organisms as "avatars" of genetic information[2]. We begin with an analysis of

---

[1]   This paper has been made possible by many years of a very stimulating collaboration with C. Sonnenschein and A.M. Soto, biologists at Tufts University. The third part of G. Longo, "Le conseguenze della filosofia" in "A Plea for Balance in Philosophy", R. Lanfredini ed., ETS, Pisa, 2015, is a preliminary version of this text (the first two parts of that paper in Italian are translated in http://www.glass-bead.org/article/the-consequences-of-philosophy/?lang=enview)

[2]   On Avatars (Gouyon, 2002; pp. 154-5), a well-known textbook on neo-Darwinian evolution, states: "To denote that which transmits genetic information or its physical carrier, we use the term avatar borrowed from the Hindu religion; it alludes to the physical forms adopted by the god Vishnu on his visits to Earth […]. The avatar, as noticed by J. Damuth, interacts with the environment, which provides for its needs and exerts an influence upon it but, above all, the avatar is produced by genetic information to ensure that this information is passed on. Individual organisms easily meet this definition. They interact with the environment, are produced

the peculiar interplay of discrete vs continua in the understanding of causality, a crucial notion in natural sciences. We will then mention some of the strong consequences of the weak conceptual frame based on a vague, common sense reference to computational notions, with particular emphasis on cancer research[3].

Information Sciences contain the grounds for a reading of the world through the "digital" or "discrete"[4] grid of numerical databases and computations, as soon as those fantastic tools for digital computing are transformed into "models" or true images of physical or biological phenomena. In other words, we claim that the intelligibility of the world proposed by discrete mathematical structures is far from neutral; in particular, it yields a peculiar relation to causality. Our claim is that the often implicit, but pervasive, reference to discrete mathematics, such as in the main theories of the elaboration and transmission of information, has biased causal analyses in biology, in particular the etiology of cancer.

2. Causality in physics and in computing

A comparative analysis between different theoretical frames will guide our critique of the informational approach still framing research, particularly in molecular biology. This critique stresses the linguistic/computational nature of that theoretical frame. As a consequence, biological investigation tends to be reduced to a molecular analysis, since DNA and molecules seem to be the natural locus for implementing programs and information. However, in our view, that perspective is not just "reductionist", if reduction means "reduction [of Biology] to the existing laws of Physics" (Perutz). As a matter of fact, major physical properties of macro-molecular interactions have been ignored, as they cannot be described in terms of discrete computational dynamics. In order to prove this, we need a short presentation of some aspects of the debate on causality and continuous vs discrete mathematics.

2.1. The mathematics of natural phenomena: causality in continuous vs. discrete manifolds[5]

"Discrete" here refers to the only good mathematical sense that can be given to this notion: the set of elements of a discrete manifold can be "naturally" given a discrete topology, that is, they may be all isolated, meaning that for each element there exists a neighbourhood, which contains only that element. Thus, we can merely count them, as they are all separated by a metrics intrinsic to the manifold, each in its own neighbourhood[6]. B. Riemann (1854) beautifully expressed this in his thesis that opened the way to differential geometry and then to Relativity Theory: "In the case of discrete manifolds, the comparison with regard to quantity is accomplished by counting, [but] in the case of continuous manifolds by measuring." (p. 3, Clifford's translation, 1873). "In a discrete manifold, the ground of its metric relations is given in the notion of it, while in a continuous manifold, this ground must come from outside. Therefore, either the reality which underlies space must form a discrete manifold, or we must seek the ground of its metric relations outside it, in binding forces which act upon it." (p. 12). In other words, in a discrete, complete manifold, the metric relations are intrinsic, as each point is "naturally" isolated, and one can merely count them[7]. By contrast, in a continuous one, a metrics, a scale for measurement has to be set - and one may then count also the number of measurements, of course. Moreover, Riemann dares to conjecture that the metrics must be grounded on the "forces acting upon it". Einstein will understand gravity as a cause of falling bodies, by identifying it with inertia in curving Riemannian spaces, where the metrics (and curvature) depend on the energy-mass and on momentum distribution. This work greatly contributed to a new understanding of causality in physics, by framing it in symmetries.

---

by genetic information, and copy the information [...]. Selection targets only genetic information, avatars are mere vehicles.". .

[3] A. Danchin (2003; 2009) is one of the few biologists who tried to search rigorously for computers' operating systems and compilers in molecular interactions, while even exploring a possible genetic meaning of Gödel's theorem, a non-obvious task (for more references, see section 3.2 below).

[4] For the non-mathematician reader, the finite strings of 0 and 1 in a computer's coding may be soundly seen as a paradigmatic example of a discrete mathematical structure. Below, we will try to define this notion in more general terms.

[5] By 'manifold' we mean a topological space in one or more dimensions.

[6] For example, the "natural" (integer) numbers are naturally isolated. Instead, on the continuum of the real number line, the discrete topology is surely not "natural": all maps are continuous on it and no relevant mathematics can be done on this basis. The so-called "natural topology" on the real line is usually considered the "interval topology" (or metrics); it is "natural" because it is derived from classical measurement in physics, which is always an interval (classical, and relativistic, measurement is approximated, and it is given as a continuous interval, in principle – no jumps, no holes). "Naturality" can also be defined in theoretical terms, see (Asperti & Longo, 1991).

[7] Typically, there are no accumulation points - that is, no actual limits points for a converging series - or operations increasingly converging to a limit make no sense.

Interlude: On Continuous Symmetries and Causality

Here symmetry means both the familiar symmetries in space (mirror, translation symmetries, etc.) and the "theoretical" symmetries extensively used in physics. For example, inertia is a conservation property (of momentum) and, by Noether's Theorems (Kosmann-Schwarzbach, 2010), it may be viewed as a continuous symmetry in the equations: momentum is an invariant by space translations, a symmetry group. Similarly, energy is invariant by time translation symmetries – it is conserved in time. In short, in either case, the "laws" do not change for space or time translations. So, by the relativistic unification of inertia and gravity, in the curving space-time of Relativity Theory, one may say that a body falls, a planet moves, or a light ray curves, for "symmetry reasons" – which is just beautiful. Thus, following Einstein's approach, enriched by Noether's and H. Weyl's work, 20th century physics largely understood causality as due to "symmetry properties", as extensively discussed in (Bailly & Longo, 2011). That is, leaving common sense behind, physicists can drop references to "causes" by treating causal phenomena within the broad theoretical frame of conservation properties. Now, these properties are described as continuous symmetries and their groups[8]. It should be clear that there is no ontological or absolute commitment here: we are just discussing how we understand (or, better, organize) natural phenomena by using different mathematical tools, either discrete or continua. However, their naive use contributed, in particular, to misleading research directions in cancer research, as we claim below.

In biology, it may be wiser to retain a "causal" terminology even when it is embedded in a broader theoretical context. A tentative step in this direction has been made (Longo, 2012) by focusing on the notion of "enablement". Typically, it may be still fair and useful to claim that "staphylococcus aureus caused pneumonia", but a good doctor should also analyze the organism's conditions that enabled the bacterium to reproduce, as they may also be relevant. More generally, a theoretical frame for organismal biology has been proposed in (Longo, 2015; Montévil & Mossio, 2015; Soto, 2016v; Mossio, 2016), one that is grounded on the singularity of organisms' default state – to which we will return,

since, we claim, the "causes" of cancer need to be analyzed in a broader biological context.

As for the analysis of the discrete vs. continuous mathematics in understanding causality, one has to distinguish between the epistemological issue (the analysis of causality) and the modeling problem. The use of continuous mathematics vs. discrete computational tools in modeling is also a delicate issue, see for example (Lesne, 2007) for a comparative insight into the mathematical modeling of physical dynamics by continuous vs. discrete mathematics. Of course, the two issues are inter-related as soon as one addresses an epistemological analysis of modeling techniques. This analysis must include a critique of the ontological assumptions often implicit in philosophically naive mathematical modeling, i.e. the view that the model is "objective" or that it is intrinsic to something, or that it coincides with "reality". As observed in (Lesne, 2007), "discrete objects are not really more 'objective' than an arbitrarily chosen partition of the space in cells," which is the intended "coarse graining" of the model. At most, a relative objectivity may be given by an analysis of the pertinent continuous vs. discrete or scale symmetries (Longo & Montévil, 2016); their "naturality" may be suggested by the chosen scale of measurement, as in physics - from quantum mechanics to hydrodynamics or astrophysics, where each theory fixes a "natural" scale of observation and measurement. As a further and dual link to causality, we (Longo & Montévil, 2016) show that in all existing physical theories, each random event corresponds to a continuous or discrete symmetry-breaking and to time irreversibility.

In summary, in contemporary physics one may understand causality within, or even replace it by, the broad frame of conservation laws. These are mathematically given as continuous symmetries in the intended descriptions, possibly given as equations (i.e. groups are the mathematical tools for their analysis). This story began with Galileo's notion of inertia[9] and its later associated transformations. These transformations form a continuous symmetry group preserving the laws of physics under any change of reference-system. In modeling and simulation on discrete-state (digital) machines – i.e. when discrete phase spaces and computations are used - symmetries are differently given and broken. In the discrete case, beginning with conservation laws in equations, everything changes and therefore their understanding, approximation, and convergence pose

---

[8]   More conservation laws and symmetries that govern physics could be mentioned: from TCP, Time-Charge-Parity symmetry, to the "supersymmetries", see (Brading, Castellani, 2003) and below for the quantum, "discrete" case.

[9]   The "default state" of inert bodies is linear, uniform and continuous movement, relative to an intended reference frame.

major mathematical and practical challenges, see (Goria, 2013). In particular, the physical analysis of causality, as a result of conservation properties in continua, is in principle lost and it is very hard to reconstruct it[10].

James Jeans, a major (quantum) physicist of the early 20th century, put it pithily: "when discontinuity gets in, causality gets out". A discrete manifold is totally discontinuous or totally disconnected: its scattered points have no topological connection with each other. From this perspective, the technical discussion in the Interlude above may be informally summarized as follows. We seem to understand causal relations either by direct contiguity (Aristotle's notion of continuity, like colliding balls) or else in a field with a pertinent conservation law. Again, this is what may be understood, in particular, by Einstein and Weyl's work.

Note that Quantum Physics and its indeterminism are presented in space and time continua: "discrete" structures appear in the dimension of energy or in the dimension of Planck's h, an action, i.e. energy × time. Typically, the energy spectrum of the bound electron is discrete, a true surprise in 1900, while the free electron has a continuous spectrum. As a special case of quantum indeterminism, 0-1 discrete alternatives may also result in measurement, such as the spin-up or spin-down of an electron; then the "standard" interpretation consistently and audaciously claims that this event "has no causes", it is pure contingency – that is, "causality gets out". From Einstein to Bohm and De Broglie, some physicists have rejected this interpretation and many still search for "hidden variables" or hidden causes varying in an underlying continuum. These scientists hoped that hidden causes (hidden variables in continua) could also justify quantum entanglement, that is, probability correlations in measurements of remote events (Jaeger, 2009). Note that entanglement is yet another phenomenon that prevents attributing a discrete structure to quantum observables or to spacetime. By entanglement, quantum observables cannot be "separated" by measurement, as there are instantaneous probability correlations, even at a distance. Thus, in quantum physics, we are particularly far from

a discrete space-time topology, made of isolated, totally disconnected elements sitting in well-separated neighbourhoods, like the pixels or the 0's and 1's in information processing.

In other words, discrete structures or discretized events provide an a-causal image of the world, far from physics or at most pertinent to Quantum Physics. Moreover, measurement is the only form of access we have to phenomena. In a discrete manifold, this is set aside, as one can just "count," as Riemann already observed. Classical and approximated measurement (an interval in continua) and the challenges of quantum measurement are forgotten (non-commutativity, indetermination, entanglement). Digital (thus discrete) databases are accessed exactly, pixel by pixel, and the causal relations are replaced by a discrete elaboration of "information" encoded by digits; and elaboration follows formal rules or instructions on how changes of digits have to take place, that is, elaboration obeys a "program". This view played a major role in biological theorizing. In physical computers, the replacement or re-writing rules (replace a 0 by a 1, or vice versa) function physically according to hidden flows that act on discrete structures, that is, by varying on underlying material continua, the computer's electric flows, and hardware. But, then, how does a digital computer actually work?

2.2. Computational dynamics

Modern computers are based on a fundamental idea by Turing (1936) regarding the elaboration of information; namely, the split between software and hardware. The autonomous science of software (or programming) was then born from logic, thanks to Turing, Gödel, Church and a few others, jointly in some fantastic areas of great mathematical rigor and achievement (computability theory, proof theory, type theory - by which the author of these lines used to earn his living!). The core idea is that programming (and its science) is independent of the hardware[11]. Similar conclusions can be drawn from Shannon's theory of information communication (1948): its analysis is independent of the material structure for the transmission (cables, waves, drums …).

---

[10] The use of the differential method (observe or induce a difference, such as a mutation in DNA, and then deduce a causal relation with a difference in phenotype) must also be critically analyzed. It may lead to wrong conclusions in biology when it is not framed in a sound theoretical context, as it is by geodetic principles or conservation properties in physics, see (Longo & Tendero, 2007) for a comparison. The case of cancer discussed below is a specific example.

[11] Note that a major difficulty in realizing Quantum Computing concretely is due to the uses and constraints that the quantum theory (the hardware) allows and imposes on programming and, thus, to the unavoidable blend of hardware and software: e.g. measurement, which co-constructs the quantum state, and entanglement have key programming (software) consequences.

Thus, programming may be identified with a general form of "term (re-)writing": programs are an alpha-numeric writing of instructions on how to transform or re-write alphanumeric strings into new alphanumeric strings (Bezem, 2003), a space-and-time discrete dynamics. In computer networks, distributed in space-time continua, this presents some peculiar difficulties that are adequately dealt with by the difficult mathematics of concurrent and network programming. These analyses are also based on continuous dynamics in complex network structures (Baccelli, 2016a; 2016b), whose nodes are digital computers (Aceto, 2003). Moreover, if one looks closely into today's computer's hardware, the instructions that modify a discrete (possibly digital) data-type actually work by variations in electric tension's levels in continuous fields and/or by driving electric currents in continua into two stable states, through discrete thresholds. That is, in silico, continuous dynamics undergo "critical transitions" and pass through various sorts of switches that stabilize current or no-current states in a material component of the hardware (the 0 and 1 at the base of digital computing). So, physical causes and flows still refer to continua, yet the physical structure of computers displays for the user only a discrete interface, as bits or pixels, where causality is hidden and only the writing and re-writing system appears (the changing 0 and 1's). This is an amazing technological achievement: by fine engineering, one may forget the underlying physical hardware and its continuous flows and just consider (and work on) the discrete software processes by writing alpha-numeric programs.

I would dare to say that this invention of ours – the discrete, visible, programmed dynamics on a computer screen - is as important and as far from the natural world as the invention of the alphabet some 5,000 years ago in Mesopotamia (Herrenschmidt, 2007). At that time, paleo-anthropologists claim, humans first discretized the continuous flow of language, originally a song, by strings of meaningless signs. Indeed, modern digital computers are the latest advancement of that atomistic creation of ours, which cuts the flow of language into a discrete notation: this is their "linguistic" origin, well beyond nature. Moreover, today, the once-static alphabetic writing moves on a screen: it is not only written, but it is automatically re-written according to written (alphabetic) instructions. Causality gets out from the image of the world that is proposed on the screen of digital re-writing machines, as it is hidden behind a cascade of major technological inventions that separate software from hardware. We see only pixels, re-written

from other pixels, 0's transformed in 1's and vice versa, following exactly-defined instructions in a discrete structure, with the biologist having no idea - and no relevance for the computer scientist - of how this is physically obtained: causes are replaced by instructions, like in the instructional theory of the DNA in ontogenesis. In computing, this is a fantastic accomplishment for programming theory, the science of software that has been broadly developed, independently of the hardware support and its causal dynamics. However, the analysis of a causal structure may be relevant in the natural sciences (e.g. one searches for the "causes" of cancer), possibly to the extreme of excluding causes, as standard quantum physics dares to do (the a-causal nature of the spin-up or -down of a quanton mentioned above).

And here is another fundamental feature of discrete computations and information technologies: any set of isolated points can be encoded in just one dimension, with no loss of information - a sequence of 0's and 1's suffices, meaning that discrete data and computations are insensitive to dimensional coding (up to a modest, linear, coding cost). This is an essential property in order to write Turing's Universal Machines, and therefore today's operating systems and compilers: they are encoded like programs and data, all in the same, unique dimension, the "Type" (or dimension) of integer numbers. The expressiveness of computing, i.e. the class of functions that can be computed, is based on the self-referential power of recursion[12] and compiling. These are both encodable in the one-dimensional type of integer numbers, an invention by Gödel, Church and Turing, in the 1930s, and a key tool also in generalizing recursion theory to all countable types – computable functions acting on computable functions, and so on, also a useful theoretical basis for computing all sorts of data types, see (Kreisel, 1982) for references.

Once again, these are very effective tools, but may yield a totally distorted image of the physical and biological world[13]. Typically, everything changes in physics and, a fortiori, in biology when dimensions are changed.

---

[12] Recursion allows the writing of self-referential equations on integer numbers, i.e. to define an x satisfying $x = f(x)$. It is easy to solve this sort of equations over many continuous domains, as fixed points, while it is a non-obvious feat of computability theory to be consistently grounded on them.

[13] Wolfram and his followers claim that the Universe may be seen as a (big) Turing Machine (Wolfram, 2013). From this perspective, an apple falls because it is programmed to fall, like a falling apple on a computer screen. Today's physics of symmetries (see the interlude above) is not much affected by this sort of claims. But, lacking a theory of organisms, the myth that an embryo develops because it is programmed to do so has been unfortunately more successful than the computational explanation of falling bodies.

The dimensional differences are crucial in physics, e.g. when distinguishing energy vs. force, or when describing, say, wave propagation - heat for example, where dimensional differences deeply modify the diffusion. In biology, if one forgets dimensionality then one misses the bodily material structure of organisms, which necessarily have three spatial dimensions[14]. For example, in vitro experiments were immensely enriched by the recent practice of three-dimensional cultures (Mroue, Bissel, 2013).

In conclusion, the informational/computational approach has for too long diverted attention from the rich networks of causal and enablement relations, within an organism and an ecosystem, in favor of an instructional a-causal perspective. A change in a phenotype must derive from a change in the instructions that are encoded in discrete data types, and this may totally bypass the physico-chemical causal structure - or even force a wrong causal structure (see below). Moreover, by the loss of dimensional analysis and due to the split between software/hardware, this approach misses the proper dimensionality as well as the radical materiality of biological organisms. Organisms are composed of their specific matter: the bases of DNA and the molecular components of membranes have no physico-chemical alternatives in a space that we strictly understand in three-dimensions. In summary, the crude, naive dualism and immateriality of the vague references to discrete unidimensional coding and software is sufficient to hide the proper causal structure as well as the spatiality, materiality, singularity and historicity of the living being, which can be always surmised as this living thing here, in this three-dimensional space, with this material body and this history. There is no way to transfer the biological "information" from DNA to Lego, like in the toy Turing Machine constructed in homage to Turing in Manchester in 2012[15], and have it work for ontogenesis. Synthetic biology extracts and re-combines fragments of DNA, or their exact chemical replica, and

places them in cellular membranes with their proper physico-chemical and dimensional structure. Dualistic perspectives - software vs. hardware, or soul vs. body, are a fantastic invention for the purposes of computing with machines or expressing a religious dimension, but they constitute a major distortion of knowledge when imported into the natural sciences. And they place the analysis of the causes of cancer on shaky grounds.

3. Strong Consequences of Weak Hypotheses

Once we focus on term re-writing as the programming structure of selfish genes, (physical) causality is removed and the search for coded "instructions" or "recipes" (Maynard-Smith, 1999) guides the analyses of biological phylogenetic and ontogenetic dynamics. So, François Jacob explicitly identified genes with alphabetic writing[16], while W. Gilbert (1992) claimed that, once the human DNA was fully decoded, we would be able to encode it on a CD-ROM and say: "Here is a human being, this is me". In the same dualistic/mystical vein, Francis Collins, director of the National Human Genome Institute, publicly asserted in 2000: "We have grasped the traces of our own instruction manual, previously known to God alone."

3.1. Exact Codings

The informational approach in biology conflates the concept of programming on discrete data with the common-sense understanding of "information" and "computer program", which are vaguely familiar to everybody, at least as a precise form of what is meant by "concert program", "instructions", "recipe". Yet scientific knowledge usually emerges when common sense notions are rejected (Bachelard, 1940), like "sunrise" and "sunset" which corresponded to an immobile earth. In fact, the use of "information" and "programming" in biology is not scientific because it neither applies the mathematical invariants proper to information and programming, nor the theorems proper to the corresponding scientific disciplines. Instead, it transfers a vague, everyday notion and refers to "weak" meanings[17].

---

[14] Through "mean field theory" in physics, we know that more than three space dimensions force a mean field and forbid singularities, such as borders, membranes ... an impossible world for organisms. In two dimensions, it is hard to have ducts. We humans seem to be suited just for three space dimensions, no more and no less. One dimensional discrete encodings miss or bypass this fundamental aspect of the topological/geometric structures. As a matter of fact, one may claim that "everything "geometric" or spatial is sensitive to coding and to dimensions.

[15] On the centenary of Turing's birth, a computationally complete computer was constructed in Lego. It could indeed compute all computable functions ... rather slowly though!

[16] "The surprise is that genetic specificity is written not with ideograms like in Chinese, but with an alphabet." F. Jacob, Leçon inaugurale, Coll. France, May, 1965. See also the explicitly "linguistic model" for biology in (Jacob, 1974).

[17] In an attempt to have a precise notion about information theory in biology, an extensively quoted paper, (Maynard-Smith,

The specificity and historicity of organisms also seem unsuitable for a description by the conceptual invariants proper to general mathematics, by the a-historic and generic nature of mathematical objects (Longo, 2017); this is also shown by the little or no invention of new mathematical concepts and structures inspired by biology, when compared with the marvellous role of physics in producing new mathematical ideas. Even less can an organism be scientifically reduced to the uni-dimensional and immaterial invariance of computer software and its arithmetic coding. Their beautiful mathematics has demonstrated its incompleteness even for proving relevant properties of programs and of arithmetic (Longo, 2011). Perhaps, new mathematical ideas are being found, both by an original modeling of morphogenesis, based on properly biological principles (Montévil, 2016b), and by a calculus of "heterogenesis" as changing spaces of observables and parameters, a true novelty with respect to the existing mathematics for physics (Sarti, 2018), partly inspired by our work.

In summary, can information be used in the Shannon-Brillouin sense? Does it partake of Turing-Kolmogorov Algorithmic information theory? Or, should it be viewed as software on the discrete data type? In spite of the lack of scientific specification of what biological information means exactly, the informational approach was justified by - and implied - several important consequences. First, molecular structures became the obvious discrete data types and codes for programs and the ultimate information storage for organisms and for all biological dynamics. Then the functional specificity of nucleic acids was supposed to be entirely due to the sequence of its bases, as complete codes for the se-

quences of the amino acids of proteins. Moreover, exact macromolecular specificity, e.g. the key-lock paradigm, a very strong property, if ever there was one, of macromolecules, was derived from (or revitalized by) the analysis of how to elaborate and transmit information: "Necessarily stereospecific molecular interactions explain the structure of the code ... a Boolean algebra, like in computers" (Monod, 1970). Similarly, chemical and stereo-specificity allow the "oriented transmission of information", as assumed by Crick's 1958 central dogma of molecular biology, (Monod, 1970)[18].

In synthesis, strong and still now accepted consequences/implications follow from the information theoretic framework, as summarized in Stanford's "Biological Information" chapter:[19]

1. The description of whole-organism phenotypic traits (including complex behavioral traits) as specified or coded by information contained in the genes.
2. The treatment of many causal processes within cells, and perhaps of the whole-organism developmental sequence, in terms of the execution of a program stored in the genes.
3. Treating the transmission of genes (and sometimes, other inherited structures) as a flow of information from the parental generation to the offspring generation.

As synthesized in (Griffiths, 2001, pp. 395–96) "Genes are instructions—they provide information—whilst other causal factors are merely material…. A gay gene is an instruction to be gay even when [because of other factors] the person is straight." Under this computational perspective, the informational cascade from DNA to phenotypes is centered on molecular exact ("stereospecific") interactions, that turns out to be the only way ("it is necessary") to transmit and elaborate information, as in a re-writing system. The Boolean, key-lock model refers to a formal chemistry that may then be analyzed in terms of computational re-writing

---

1999), explicitly mentions Turing-Kolmogorof (Elaboration or Algorithmic Information Theory) and Shannon-Brillouin (Transmission or Communication of Information). This paper thus acknowledges their literal use, but confuses these approaches in their dual relation to complexity and entropy, see (Longo, 2012; Perret & Longo, 2016) for a critique. Yet, most authors claim that their function is "just metaphoric". But "When metaphors have been used too often, they die: people cease to be aware that the metaphoric use of the words is not a literal one. Then they become illegitimate forms of predication and of discourse" (Ricoeur, 1975). Indeed, rarely do metaphors entail strong consequences: it is the literal discourse (the reference to alpha-numeric instructions), combined with the reference to common-sense knowledge of what information and the program mean, that imposes a vision of the world. For example, commonsensical observation of the immobility and centrality of Earth, interpreted literally, allowed for centuries the deduction of a philosophically strong and mathematically detailed geometry of epicycles in order to describe the planets' orbits. The geocentric hypothesis was not metaphorical, but literal and commonsensical. And so is the programming-genocentric hypothesis of the encoded Aristotelian homunculus. It entailed the strong, precise consequences we discuss in this section (3.1).

[18] "Biological specificity ... is entirely ... in complementary combining regions on the interacting molecules" (Pauling, 1987). "The orderly patterns of metabolic and developmental reactions giving rise to the unique characteristics of the individual and of its species ... the shapes of individual molecules allow them to selectively recognize and bind to one another. The main principle which guides this recognition is termed complementarity. Just as a hand fits perfectly into a glove, molecules which are complementary have mirror-image shapes that allow them to selectively bind to each other" (McGraw-Hill Dictionary of Scientific & Technical Terms, 6E, 2003).

[19] Philosophy of Biology: http://plato.stanford.edu/archives/fall2008/entries/information-biological/ (October 2016).

processes, which transform sequences of letters into sequences of letters, following the instructions, in a deterministic and predictable way, plus some unavoidable noise, (Monod, 1970). Evolution would be the result of noise in an "exact, Cartesian, machine".

3.2. Stochasticity and the Creation of Novelty

It should be clear that the founding fathers of molecular biology discovered fundamental physico-chemical structures and mechanisms at the core of cellular activity. However, their amazing experimental insights and observations, such as the double helix structure or the allosteric and lac operon mechanisms found by J. Monod, F. Jacob and J.-P. Changeux (1961 – 1962), were later embedded in the theoretical frame we criticize here.

The computational approach typically excludes physical stochasticity from being an essential component of gene expression, and, more generally, excludes stochastic and low-affinity macromolecular interactions. This exclusion is contrary to evidence of these chemical phenomena, which dates back to the late 1950s – as for the role of Brownian motion in a cell and stochastic gene expression, see (Kupiec, 1983; Elowitz, 2002; Paldi, 2003; Raj, 2006 and 2008; Fromion, 2013; Marinov, 2014) for references and contributions. Moreover, chemistry has long dealt with macromolecular interactions in stochastic terms (Gillepsie, 1977). Macromolecules have large quasi-chaotic oscillations (shaped by their enthalpies), they are "very sticky" and low affinities are relevant: they are thus treated in probabilistic terms, whose values depend also on the context - see the references above and (Creager & Gaudillière, 1996; Kupiec, 1996) for more on the origin of this debate. In short, the physico-chemical analyses are based on the global stochastic behavior of ensembles of macromolecules and not the individual behavior of each of these molecules, which remains submitted to the perturbing influence of thermal agitation and other "random" dynamics such as affinities with low probabilities. However, this stochasticity has a different nature from the one dealt with in statistical physics: many molecular types, in a cell, contain "small" numbers of molecules and their behavior is highly constrained by chemical affinities, membranes, compartments … and by the "Physics of Epigenetics", see below. On these grounds, a recent research track, derived from chemistry, radically departs from the "information-programming" approach, where each gene would act like a Laplacian demon "instructing" molecules individually, but this track differs from a purely statistical approach based on physics. Statistical physics

refers to "huge" numbers in the passage from the micro to the macro-level, on the order of Avogadro numbers, 1023, in contrast to molecular analyses, which may refer to small numbers for a molecular type in each cell. The aim is to find a good "mesoscopic" level of description for understanding the regulated stochasticity of macro-molecular interactions in a cell, including gene expression, see (Giuliani, 2010) for more.

Informational language, instead, constructs an autonomous conceptual universe independent from the underlying physical processes and their causal structure: that is, causes are replaced by information flows, signals, control, programs, whose necessary physical support is the assumed exact complementarity of keys and locks, hands and gloves. Of course, some randomness cannot be excluded. Yet, in view of the predictable determinisms of Boolean re-writing systems, randomness is described as "noise" affecting evolution in particular: "Evolution originates in noise, imperfections …" (Monod, 1970). Also in this respect, the information theoretic terminology is not neutral. In particular, it biases the understanding of biological variability, adaptivity, diversity: they are all (or are derived from) unavoidable noise. A typically "informational" identification of randomness and noise, two distinct concepts in physics. The latter is usually eliminated from information processing or, at best, averaged out, as randomness at the tail of a Gaussian. It is thus treated by "central limit" theorems like in the recent area of Noise Biology, based on statistical physics (see Bravi & Longo, 2015, for a critique).

In the programming approach to biology, some researchers have been looking for a proper form of randomness to be added to noise. This approach, based on the literal and rigorous understanding of "genetic program", justifies biological novelty and diversity creation as a form of unpredictability (as randomness) derived from Gödel's theorem. The evolutionary "creativity" of the DNA should be understood as the "creativity" (its name in Logic) of the set of provable theorems in Arithmetic. This happens to be "incomplete" and allows to "create/construct" the unprovable sentence. This is not an idea conceived by extremists, but by a few competent and coherent molecular biologists, largely quoted, such as A. Danchin (Danchin 2003; 2009). It is a tentative, more rigorous "DNA-as-a-program" approach that, at least, goes beyond the usual vague, common-sense use of "information" and "programming" in biology – in principle, one should closely look for arithmetic recursion and logical negation (!) coded in DNA

(they are both needed to encode Gödel's theorem). It is thus possible to criticize more precisely this severe form of Gödelitis (Longo, 2018).

Thus, with regard to biological randomness, as unpredictability of phenotypes and as a component of evolutionary novelty, information-oriented frameworks do not require the complex blend of physical, classical and quantum randomness in a cell or in a more complex organism: either it is noise in information-elaboration channels or it is … "Gödelian" as above. Even Brownian motion is seen as disturbing noise in exact, Turing-machine-like, genetic expression. Instead Brownian motion, along with macromolecules' enthalpies-based oscillations, dominates the physical dynamics and the energetic landscape and it has a constructive role in both prokaryote and eukaryote cells (see the references above as for stochastic gene expression, an approach harshly marginalized for decades by the informational mainstream; (Richard, 2016) presents further experimental evidence). Thus, the stochastic approach highlights a fundamental causal component of bio-chemical interactions in physical dynamics that we better understand *in continua*, while "stochastic" transmission and elaboration of digital information would make little sense. Moreover, different forms of randomness, at all levels of organization, may causally contribute to phenotypic changes and to biological stability by adaptivity and diversity, see (Buiatti & Longo, 2013) for the further notion of "bio-resonance" and (Calude & Longo, 2016a) for a survey.

The analysis of the physical structure and environment of the cell (the "Physics of Epigenetics" mentioned above) provides an understanding of some of the key physical constraints that canalize molecular dynamics. In particular, movements, torsions and compressions of the chromatin fiber structure enhance and control "DNA transactions by an epigenetic tuning of its mechanical and topological constraints", as stressed in the seminal work by (Lesne & Victor, 2006). More precisely, "steric hindrance, conformational changes at various scales, topological constraints (on DNA and the fiber), elastic properties (of DNA and the fiber), electrostatics" … crucially contribute to chemical interactions, as those authors observe. Cortini (2016) offers a broad survey of "the physics that governs the three-dimensional organization of the genome in cell nuclei". The authors stress that even the terminology of "histone code" is inadequate; this is so, we think, because of the independence from the material and dimensional structure of the notions of information and code that contradicts these

analyses of the physical dynamics in cells. Note also that torsions and elastic deformations are not used to elaborate information in computers (please, do not try with yours) nor, more generally, in the implementation of alphanumeric re-writing systems.

The informational perspectives bypass or are incompatible with additional physical phenomena in cells, such as the possibly very relevant role of the "super-coherence" of water. This is due to a Quantum Electrodynamic effect in highly partitioned structures containing water, such as an organism made of 10^{13} cells, further divided by internal membranes; water molecules are then organized by co-oriented spins (Del Giudice, 1983, 1986; Arani, 1995). This accelerates the Brownian motion of non-water molecules at constant temperature and, thus, enhances the rate of (stochastic) biochemical interactions.

In conclusion, the necessary physical analysis of macromolecular dynamics requires a complex blend of discrete and continua. Quantum Physics, in a completely different realm, faces a similar challenge. In biology, the role of macromolecules, beginning with a fundamental physico-chemical trace of evolution, DNA, and their discrete structure, is evident. In particular, as a discrete template for proteins, DNA constitutes a relatively stable constraint on molecular random movement and formation; it resists thermal fluctuations (Sarpeshkar, 1998), while it uses molecular Brownian motion and is "opened" to interactions also by pressures and torsions, all notions that are better understood in continua.

In section 6, we propose a properly biological investigation of constraints and of unpredictable changes, which uses but goes beyond purely physical notions of constraints and randomness for the analysis of both biological stability and novelty creation.

3.3. The Software of Life and reductionism, both away from Physics

In summary, in the Theory of Programming, a robust science on its own, but also in the information/programming approach to biology, the underlying hardware has no interest for the program analyst, provided that it works correctly, in spite of some noise. In Computer Science, the needs of Programming set the standard of "correct" working for the physical, material structure, which followed Turing's mathematical distinction between software and hardware by more than 10 years. It is the engineers' job to have the hardware work according to the programmer's needs and, thus, realize an interface appearing as a (Turing-vonNeuman) discrete state architecture, with whatever material structure they

have. And they can have it work correctly, which is just fantastic. In modern computers, we thus implemented the strongest form of Cartesian soul/body split, by radically subordinating matter (hardware) to an independent soul (software). Similarly, the material cell must follow the genetic instructions; it is just an Avatar (see the footnote above on Avatars). Yet, the genome may escape from them and generate novelty internally, independently from physics, from the organism, and from the ecosystem. This may be due to noise in information processing or it is justified by the literal, non-metaphorical form of Gödelitis, mentioned above.

When developmental biology follows this extreme Cartesian dualism, it is reduced to purely formal laws of a derived symbolic chemistry, a virtual interface handled in terms of information and programming theory, with a reference to physics in occasional reductionist claims[20]. But, if the causal structure of this presumed formal biochemistry of information in macromolecules is absent, doubtful or incomplete, the laws of which physical theory are actually invoked in these reductionists perspectives in biology?

Physics, from Galileo to Quanta, has never ceased to construct and modify its laws by confronting unprecedented phenomena or by novel insights into known phenomena, or just by changing the scale of observation. There is no reduction within physics, as it proceeds by "unification", from Newton to Boltzmann to current attempts to unify Quantum and Classical/Relativistic physics or chemistry or hydrodynamics, see (Chibbaro, 2015) and (Longo, 2016) for a review. For example, hydrodynamics, as a science of incompressible fluids in continua, is not understood in terms of quanta; physicists try instead to invent a new theoretical frame that could unify these theories (note that there is a lot of water in an organism … thus, which "physics" are reductionists in biology referring to?). Moreover, classical and quantum random phenomena, which are far from unified, are both present and interact in cells and may have phenotypic consequences (Buiatti & Longo, 2013). In physics, all existing unifications were based on very strong theoretical hypotheses, grounded on revolutionary ideas. For instance, Newton's equations and infinitesimal calculus, which unified Galileo's falling stones and celestial bodies; Maxwell's equations for electro-magnetism and optics; Boltzmann's asymptotic construction of Statistical Physics, which unified

particles' dynamics and Thermodynamics on the grounds of the ergodic hypothesis, an incredibly strong and precise statement. String Theory or Non-Commutative Geometry, as for today's attempts to unify quantum and relativistic fields by incredibly strong, revolutionary assumptions and concepts, surely not derived from common sense. And none of these is a "reduction" to a "lower" level. Moreover, as it is in the cases above, unification, in science, should always be provisional and "local," not dogmatic and a priori reductionist, but constructed as a new theoretical frame. Developing unifying frames of biological analyses and of relevant physical theories may be a challenging and long-term task, since even the analysis of molecular dynamics requires original physical treatments, as the "mesoscopic stochasticity" hinted in the previous sub-section.

The deduction of strong consequences from weak, fuzzy, a-scientific, vaguely metaphorical or literal and "common sense" hypotheses, such as the "information" or "programming" assumptions in biology, is unacceptable as a scientific praxis. Note finally that, this pre-scientific reference is only made to a Theory of Information on discrete data types; elaborated, transmitted, and encoded by programs, written as alpha-numeric instructions. No reference is ever made, that I know, to the well-established discipline of Geometry of Information, where symmetry changes in continuous symmetry groups propose a radically different conceptual frame (Barbaresco & Djafari, 2015). Perhaps, such an approach would broaden the reductionist focus on DNA and molecules by a possible analysis of morphogenesis as a dynamic of "geometric information in three dimensions".

In addition to the few listed above, we will see some specific strong consequences of the weak hypotheses transferred from common sense notions of "information" to biology and how these still affect developmental biology and, thus, cancer research. These domains have been for too long dominated by the myth of the computer program and information, centralized in the DNA (the Central Dogma of Molecular Biology). The increasing attention to "epigenetic information" did not modify the focus on information-program-signal, as drivers of development, nor the idea that ontogenesis as well as its pathological developments should be always (or first) studied as a DNA centered issue. In this context, cancer has been consistently analyzed as the result of DNA de-programming either inherited or provoked by a carcinogen disrupting the DNA

---

[20] «Life can be explained on the basis of the existing laws of Physics» (Perutz, 1987). Fortunately, Galileo and Einstein did not have a similar conservative stand, within physics.

encoded instructions (the mutagenic effect of carcinogens). Let's briefly summarize some steps of this still prevailing view of the etiology of this life-threatening disease, a view that recently received further support from the software industry, in spite of massive negative evidence.

4. An announced debacle

We will follow the story of a wrong path as courageously acknowledged by one of the founding fathers and a major actor of the dominant theory in the biology of cancer, R. A. Weinberg, in his 2014 paper (see references). The so called Somatic Mutation Theory (SMT) postulates that cancer originates as a one-cell disease, it is then clonal, and is due to one or more driver mutations, see (Nowel, 1976, Cairns, 1981; Strauss, 1981) for classical surveys of this century-old theory originally proposed, in a different language, by (Boveri, 1914).

Since 1971, generously funded projects have heralded the final victory against cancer thanks to genetic therapies able to "reprogram" the "deprogrammed DNA", within a few years. In particular, this approach was at the core of President Nixon's War on Cancer (see below for more quotations on this). The common sense notion of "program" was indeed understandable also by Nixon; a major advantage of using an everyday, a-scientific language, as it facilitates the understanding of the message by everybody. Moreover, programs can be debugged, thus the promise of genetic therapies as DNA debugging is still ongoing (see below): an easy to understand, short path to therapies. In spite of providing neither therapeutic solutions nor plausible explanations of the carcinogenic process since 1971, the complete decoding of human genome, a major technological achievement, by the year 2000, was seen as a further tool to solve the cancer puzzle and generate, once again, genetic therapies. These were anticipated within 10 or 15 years at the latest, while sound diagnosis and prognosis were promised much sooner. Genetic analysis of cancer cells were expected to provide diagnosis of malignant vs. benign forms of this disease, primary vs. metastatic cancers etc. These optimistic papers are too many to be listed; it may be enough to quote (Collins, 1999), written by the head of the Human Genome Project, (Hanahan & Weinberg, 2000) (over 20,000 quotations in a few years), (van Eschenbach, 2003), all major personalities in the field. In (van Eschenbach, 2003), cancer is viewed as both "a genetic disease and a cell signaling failure. Genes that control orderly replication

become damaged"; on the grounds of this causal analysis, the paper promises, by 2015, genetic therapies for "eliminating suffering and death due to cancer". Incidentally, this claim was supported by the American Association for Cancer Research in 2005.

Thanks to knowledge of DNA sequences in normal and cancer cells, these proposed upcoming therapies were supposed to be based "on scientific laws as robust as those of chemistry and physics" (Hanahan & Weinberg, 2000). The proximity of the notion of "program" to common sense, as always, promoted these promises among funding agencies and the public. The enormous financial efforts and the ruthless exclusion of alternative hypotheses have both been motivated for decades by the idea that any phenotype, including "pathological" ones, is determined by the genes, or their mutations. However, a half-century of genetic research has produced no plausible gene-based cancer therapy, see (Baker, 2014; Huang, 2014) - two elegant syntheses and highly recommendable reading to the non-biologist (but so worrying!). As Weinberg (2014) himself acknowledged, "We were, after all, reductionists, who would parse cancer cells down to their smallest molecular details and develop useful, universally applicable lessons about the mechanisms of cancer development […]. Half a century of cancer research had generated an enormous body of observations about the behavior of the disease, but there were essentially no insights into how the disease begins and progresses to its life-threatening conclusions". So, Weinberg (2014) observes, against the extensively quoted (Hanahan & Weinberg, 2000), that "a particularly jaundiced cancer researcher" commented to him "one should never, ever confuse cancer research with science!''.

How could DNA be de-programmed according to the early research projects? At the beginning of the 1971 War on Cancer, retroviruses were considered as DNA de-programming agents. "Few seemed deterred by the well-established observation that most types of human cancer did not represent communicable diseases" (Weinberg, 2014). Ramazzini, anatomist and physician in Bologna had already made this observation in early eighteenth century.

Weinberg continues his auto-critique (pp. 267-9) by summarizing further spurious key steps in the SMT approach to cancer. Since 1973, the search focused on "chemical species correlated directly with mutagenic activity". He then recalls the progressive move, between 1982 and 1999: from "just one mutation" to "a specific sequence of mutations". "Only later was it clear

that most human carcinogens are actually not mutagenic ... but fortunately I and others were not derailed by discrepant facts" (sic).

This is a crucial remark. As a matter of fact, there is increasing evidence that many (most?) carcinogens interfere on tissue organization, not by sending (chemical) signals that de-program DNA. For example, Maltoni (1980) observed the disruptive role of asbestos microfilaments on the tissue matrix, on cell connections and membranes, but could not point to any direct mutagenic effect. This observation was in contrast with the claims of the dominant theory (SMT) and, hence, received little consideration. As a matter of fact, when asbestos is delivered as micro particulate (powder), it ceases to generate cancer: "fiber dimension is one of the important determinant factors of asbestos carcinogenicity" (Huang 2011). Also, by subcutaneously inserting diverse inert objects (plastics, metals, etc.) it has been shown that their carcinogenic effects depended not on their chemical nature but on their peculiar physical structure (e.g. the carcinogenic effect may depend on the presence and size of micropores in plastic membranes, a fact known since the seventies (Karp, 1973)). Mutations may follow as consequences (called, in SMT, "passenger mutations") not causes ("driver mutations", in the SMT terminology) of cancer, as we will recall soon.

Other commentators of note have expressed their views on carcinogenesis for the record. In a very interesting interview, Venter (2010), whose team first decoded the human genome in 2000, acknowledged that "'We Have Learned Nothing from the Genome''. Wrong expectations were due to "the ill-founded belief that those who know the DNA sequence also know every aspect of life … That is nonsense". However, cancer biologists did learn something from the Genome Decoding. The extensive decoding of the DNA of cells in cancerous tissues showed that, in the same tissue, cells might have very different mutations and chromosomal changes: "Genome sequencing also came of age and documented myriad mutations afflicting individual cancer cell genomes" (Weinberg, 2014). More precisely, "63 to 69% of all somatic mutations [are] not detectable across every tumor region … Gene-expression signatures of good and poor prognosis were detected in different regions of the same tumor" (Gerlinger, 2012), see also (Kato, 2016). Genomics in the analysis of metastasis did not provide much help either, as acknowledged also by proponents of SMT. "Despite intensive effort, however, consistent genetic alterations that distinguish cancers

that metastasize from cancers that have not yet metastasized remain to be identified. [.] The idea that growth at metastatic sites is not dependent on additional genetic alterations is also supported by recent results showing that even normal cells, when placed in suitable environments such as lymph nodes, can grow into organoids, complete with a functioning vasculature" (Vogelstein, 2013). In the interpretation hinted in the next section, normal cells in a context that cannot control and canalize their "normal" reproduction with variation may yield a "pathological" situation. Moreover, no driver mutations specific to metastasis have so far been documented (Zhang, 2013: Alshaya, 2014).

Finally, it is remarkable that cells in healthy tissues may have the genetic hallmarks of cancer: "aged sun-exposed skin is a patchwork of thousands of evolving clones with over a quarter of cells carrying cancer-causing mutations while maintaining the physiological functions of epidermis" (Martincorena, 2015). Equally noteworthy is that cell aneuploidy and polyploidy, which used to be considered as another chromosomal signature of cancer, are present in 50% or more of normal liver cells and are considered to be beneficial by assuring resilience to toxic shocks and for liver regeneration (Duncan, 2013).

Following the quotations referred above, a few relevant facts have become clear from the massive DNA decoding of cells in cancer tissues. They are, in summary:

1. Gene-expression signatures for benign and malignant cancer may coexist in the same tumor.
2. Genetic analyses do not allow to discriminate between a tumor that (has or) will metastasize(d) from another that (has or) will not.
3. DNA sequencing does not help in distinguishing a primary from a metastatic cancer.

Note that 90% of lethal cancers are metastatic (Sporn, 1999; Cook, 2011). This fact dramatically stresses the relevance of the last two points.

In conclusion, the etiology of cancer, that is the origin of primary cancers, remains an open problem. However, proponents of SMT acknowledge that 99.9% of mutations found in cells of all cancer tissues are passenger not driver mutations of cancer, see (Vogelstein, 2013) and the next section[21]. Therefore, increasingly many

---

[21] In reference to the percentages mentioned in the last few lines, it may be fair to claim that most publications in biology of cancer (90% ?), in the last few decades, focus on geno-centric approaches and that the vast majority of research funding (90% ?) has been

authors seem to acknowledge that the "primary and immobile motor" of ontogenesis - and thus of cancer as of any phenotype, the DNA as a program, is actually a passive recipient of orders, resulting in passenger mutations. The possibly messy situation of cells' chromosomes in a cancer - not just passenger mutations, but massive polyploidy, aneuploidy, etc., may negatively exert a retroactive effect on tissues' unhealthy dynamics: their deregulating effects may further disrupt the cells' dialogue, hormonal control of reproduction etc. see the next section and (Baker, 2014, 2015; Huang, 2014) for surveys. Yet, evidence has also been obtained that "cancer cells can display a seemingly paradoxical state in which their mutational burden is similar to and perhaps even lower than that of adjacent normal cells" as acknowledged by (Gatenby, 2017), where the driving role of tissue and organismal environment is stressed. Even more radically, (Versteg, 2015) mentions tumors without mutations. Moreover, as observed since (Sonnenschein & Soto, 1999), mutated cells from a cancer tissue may functionally normalize when transferred in a healthy tissue, see also (Soto & Sonnenschein, 2011).

4.1. The strength of vague theories

"… You cannot prove a vague theory wrong"
Richard Feynman (1964)

In view of the remarkable empirical knowledge that DNA decoding has provided, are we approaching the end of a (de-)programming DNA centered view of cancer and of ontogenesis in general? Hopefully, empirical negative evidence in the natural sciences should have the same role as "negative results" in mathematics or mathematical physics. In principle, they should modify scientific thinking and scientists may invent or become more open to new theories, and new scientific paradigms (Longo, 2010; Longo, 2018). To the contrary, the genocentric informational/ programming views cannot be falsified by experience nor "in theory", because they are not scientific: those views are based on common sense notions of information and program and on the "homunculus" ancient myth, modernized and made literal by encoding it in chromosomes[22].

Thus, the presence of mutations and chromosomal alterations in cancer tissues continues to be perceived as the cause of the disease, since according to that theory any phenotype must have an antecedent in the genotype and the genotype is supposed to completely control the organism-avatar – recall the footnote above on Avatars. We know from human history that when common sense and myths combine, they are unassailable and any change requires a true revolution. Evidence, observations, mechanisms etc. may be otherwise totally misinterpreted (see the previous footnote).

So, following the still dominating trend, Microsoft proposes to help in solving the cancer puzzle by its technical (or commercial?) skills in software production. Microsoft's "computing cancer project" (Microsoft, 2016) claims that one has to understand how the cell's programs work, then "If you can figure out how to build these programs, and then you can debug them, it's a solved problem". Their motto is "Our approach to solving cancer: debug the system". Is this just surplus money that goes to cancer research? Not necessarily, because joint ventures in this enterprise are meant to apply for funds to bio-medical research institutions: thus, the support is first used to prepare huge research project, an unfair advantage and a bias on research. And, more importantly, Microsoft's talent for commercials and publicity, which are the actual aim of these announcements in spite of the sufferings they refer to, may confirm common sense genomics by reaching the general public, politicians and managers who decide about funding; in short, it sets a reference[23]. IBM also offers DNA decoding services for cancer diagnosis and prognosis, in spite of the evidence mentioned above. And Big Data enter massively in the game.

In view of the very heterogeneous and unexpected genetic situations of cancer cells, of the "myriad mutations afflicting individual cancer cell genomes" thus

---

[22] On epicycles and Galileo's inertia. As we mentioned in a previous analogy, also the geocentric approach to the planetary system

was a literal and commonsensical approach. Yet, it implied the precise and mathematically possible Ptolemaic description. As a matter of fact, any finite number of points, thus of observations, on an ellipse around the Sun can be interpolated by enough epicycles centered on Earth – it is a matter of an approximation by finite series and, thus, the mechanics of epicycles may be empirically corroborated. Epicycles though happen to be incompatible with the Galilean default state for physics: inertia as the fundamental conservation property of momentum. The analysis of mechanisms can only make sense if framed in a sound theory, since it is the theoretical explicit, scientifically arguable proposal that guaranties a sound interpretation of observations.

[23] S. Knapton, "Microsoft will 'solve' cancer within 10 years by 'reprogramming' diseased cells", The Telegraph, 20/9/2016. As a former user, now a Linux fan, I think that Microsoft should better and first debug its own software, see (Di Cosmo, 1998).

---

allocated to those analyses. These two aspects of research trends are also the result of the amplifying effects of bibliometrics and "impact factors", that reinforce main stream, fashionable areas (Longo, 2014), and thus enhance positive unstoppable retroactions between publications and funding.

of "cancer's infinite complexity" (Weinberg, 2014), and of the failure to turn cancer biology into a science, many researchers follow Big Data 'philosophy' in (Anderson, 2008). Namely, collect all "-omics" available data (genomics, proteomics, metabolomics…), then "...throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot … Correlation supersedes causation, and science can advance even without coherent models, unified theories … No semantic or causal analysis is required". Of course, the supposition is that the larger the database is, the better it is for prediction and action with no need for understanding.

We are coming full circle back to the more than 100 years old remarks by Riemann, Jeans and others quoted above: if you have only discrete manifolds, you give up causality. Thus, consistently claim the purest Data Miners, just look for correlations without causal explanations – scientific understanding is no longer needed. Note first that these provably wrong claims against theorizing allow for the neglect of measurement as well: classical, relativistic and quantum challenges as for physical measurement are forgotten – a digital database is exact, the metrics is intrinsic. The pre-given discrete structure of the databases may thus help to forget how these data have been collected in complex biological systems. The (often implicit) *a priori* in the choice of observables and of their metrics do not need to be discussed, as this would be "theorizing". In our view, instead, Data are "Compressed Theories" (and not vice versa), since to collect them supposes a theoretical perspective, the choice of observables, measurement theory and tools …, see (Longo, 2016).

Moreover, as formally shown in (Calude & Longo, 2016), sufficiently large sets of numbers, even when produced by a random process, necessarily contain correlations, which are then spurious. More precisely, a nice and non-obvious combinatorial theory of numbers, Ramsey Theory, proves the following: (Informal) Set the criteria for a correlation in a database: its n-arity (you want to correlate n variables), the length p of the correlation (you want it to be long enough, e.g. n data must correlate every minute, for a year, say), the number c of parts you divide your database (you give the same "color", say, out of c colors, to numbers that you consider correlated: they are close or happen simultaneously or whatever). Then, for any n, p and c, one can compute a number, d say, such that for any set A with d or more elements and for any partition of the collection of sets of n elements of A in c colors, there exists a subset B of A that contains p elements and is monochromatic, i.e. it is entirely contained in one partition – thus B realizes the correlation given by n, p and c.

The number d above is truly "huge", but isn't the bigger the better?[24] Then the data miner may happily exclaim, "we've got a correlation!", even when the data set A has been produced by a random process. That is, in any immense numeric databases, one has a deluge of spurious correlations, in a very strong sense, as the set A and the partition of A above are arbitrary. Thus, A may have been obtained by … throwing dices, flipping coins, quantum measurements … and by arbitrary choices of observables and measurements. It is hard to predict and act on these grounds. Moreover, when you are dealing with very large sets of numbers, most of them are "random", in a precise, algorithmic sense (Calude & Longo, 2016). It may be wiser, then, to try some scientific theorizing.

5. Towards TOFT

Following a different research path, an approach proposed by Sonnenschein and Soto (TOFT, Tissue Organization Field Theory, see the references by these authors) is based on Darwinian principles that we further extended to a tentative theory of organisms (see the next section). The TOFT approach to cancer refers to early intuitions by C. Waddington, J. Needham and a few others (1930s), later forgotten by the subsequent genocentric perspective, see (Soto & Sonnenschein, 2011) for references. The novelty and the corresponding "paradigm instability" brought in by TOFT vs. SMT is analyzed in (Baker, 2014; 2015), see also (Sonnenschein & Soto, 2017).

The key principle of TOFT is that all cells, be they unicellular organisms, or those of multicellular ones, proliferate constitutively as long as nutrients are available; in (Sonnenschein & Soto, 1999) terminology, cell proliferation and motility is the "default state" of all cells. We extended this default state to the idea that all organisms as well as the cells in multicellular organisms, tend to reproduce with variation and to move, as more closely spelled out in the next section. This is an extension to cells within an organism of Darwin's principle of heredity in evolution as descent with modification, which occupies three out of the first six chapters

---

24   Some theorems of Ramsey Theory may be used to produce huge numbers (sizes) by "incredibly fast" growing functions, beyond arithmetic provability, see Paris-Harrington or Kruskal-Friedman non-provability results in (Longo, 2011). But the one used in (Calude & Longo, 2016) yields a size accessible to today's databases (less than exponential in n or p).

of the Origin of Species, see (Longo, 2015; Montévil, 2016; Soto, 2016). This revolutionary principle is essential to Darwin's second principle, selection. It is a "limit-state" analogous to Galilean inertia, but specific to life forms. Note that inertial movement is a limit principle, as movement is always constrained and modified by gravitation and frictions. Analogously, somatic cells, and also organisms in an ecosystem, are constrained/controlled by the organism or the environment in their free reproduction and movement. As Darwin observes, an unconstrained organism would quickly cover the entire Earth, by reproduction. Galileo's inertia, Darwin's principles and the default state of reproduction with variation and motility are all derived from observation and posed as principles of intelligibility at the core of their theoretical approach. By positing inertia, asymptotically (no physical body moves like a point on a Euclidian straight line at constant speed), Galileo could analyze what affects it, gravitation and frictions. On the grounds of his first principle, Darwin could propose selection as acting on organisms. TOFT's central idea then is to analyze what controls and canalizes cell reproduction with variation and motility in an organism, see (Montévil, 2016; Soto, 2016) for more on Darwin and the conceptual analogy with Galileo's principle of inertia. Under this perspective, since also somatic cells, as all cells, reproduce with variation and move, if not constrained, cancer is a tissue-based, organismal problem, akin to the process of morphogenesis during development.

In general, morphogenesis is the result of complex interactions among different components of a tissue and cannot be reduced to cellular events. It is a tissue-based, organismal phenomenon, inextricably linked to the three dimensions of space (topology) and to a developmental history (time).

In this sense, "Carcinogenesis takes place at the tissue level of biological organization, as does normal morphogenesis .... Chronic abnormal interactions between the mesenchyme/stroma and the parenchyma of a given morphogenetic field would be responsible for the appearance of a tumor" … "cancer is development gone awry", (Soto & Sonnenschein, 2011). In other words, in carcinomas (about 85% of cancers), carcinogens would disrupt the reciprocal relationship between stroma and epithelium that generates and maintains the morphology and function of an organ. The less constrained, or abnormally constrained tissue reorganizes into a structure that deviates from the normal tissue but is still recognizable. In particular, cells in the interior of the reorganizing structure are less constrained and thus proliferate, and move more than required by the organ's function. The

tissue actually becomes more complex in a precise histological sense, while reducing tissue (organ) functionality, a hallmark of cancer as observed in (Longo, 2015).

Note that, in contrast to the claim by the SMT that "once a cancer cell, always a cancer cell", cells from a mammary carcinoma (an epithelial cancer), when placed into a normal mammary stroma (the normal microenvironment of the mammary epithelium) revert to normalcy (Maffini, 2005). Moreover, the carcinogen does not need to act on cells that will be recognized as "cancer": exposure of the stroma may suffice in causing cancer of the epithelium (Maffini, 2005).

The idea is that cancer does not depend on a "triggering signal" at the molecular level, which would de-program the DNA of an a priori quiescent cell by inducing a driver mutation and enhancing re-production. Instead, cancer can be considered as the failure of the regulatory relations of and be-tween cells in a tissue and of the tissue in an organism. Passenger mutations may follow (also for to-day's SMT supporters, they are 99.9% of mutations in cells in cancerous tissues, see above), as mutations are one of the main modes of variation at the cellular level. The fact that the cells can be normalized shows that those mutations are not "drivers", as well as their secondary role even in reinforcing the pathological behavior and interaction with other cells and the tissue.

These hypotheses, and their therapeutic consequences redirect the attention of researchers toward prevention and modifications of environmental conditions. At the ecosystem level, beginning with our food containers (!), the focus is on endocrine disruptors and other causes of cancer (Soto & Sonnenschein, 2010). At the organismal one, the reconstruction of the cells' microenvironment may be crucial, (Cook, 2011; Bizzarri & Cucina, 2014). In the latter case, like in the recombination experiments in (Maffini, 2005), cells inside a cancer can be normalized. The reader should consult (Bizzarri, 2013; Baker, 2014; Smithies, 2015; Pisco & Huang, 2015) for surveys. Moreover, "Thinking in terms of TOFT can spur new lines of research" (Baker, 2015), as many if not most cancer "conundrums" are made understandable along these new lines of thought, (Kato, 2016). For a closer comparison SMT vs TOFT, see (Montévil & Pocheville, 2018).

6. From TOFT to Working Hypothesis in Biology of Organisms, a short synthesis

Since its origin, TOFT was meant to be framed (and could only make sense) within a sound theory of organismal development. The connections to Darwin's evolution is also crucial to us.

The Darwinian principle of selection has been extensively used in the literature on cancer, see for example (Gatenby, 2010; Shiffman. 2013). While also referring to Darwin, we begin by focusing on his first and fundamental principle, "descent with modification", largely ignored in those analyses. Then the canalizing constraints to "reproduction with variation" may be considered as a form of selection within an organism, (Montévil, 2016), a theme still to be further worked out. Within this theoretical frame, we may better understand TOFT as positing that cancer is a tissue-based problem, not a cell based one. Typically, carcinogens affect the reciprocal interactions between stroma and parenchyma, which manifests as altered tissue architecture. Constraints to proliferation and motility are weakened or modified, tissue complexity increases (see above) and functional organization decreases.

This is consistent with the role of closure of constraints in biology (Montévil &Mossio, 2015; Mossio, 2016) and with an ecosystem-based approach to evolution: genes and their expression are more followers than promoters of changes in evolution (West-Eberhard, 2003). We claim that this is so in development as well, which does not exclude, in either theory, the occasional diver's role of DNA changes. Biology of cancer thus becomes part of a sound "theory of organisms" in correlation to evolutionary theories, see also (Montévil, 2016; Sonnenschein & Soto, 2016). A stringent example is provided by (Gatenby, 2011), where the lesson from the cavefish is learned: hybridization of close but different species of this eyeless fish, which evolved from a "normal" fish, yield fishes with functional eyes.

As for the relation to physics, in (Bailly & Longo, 2011; Longo & Montevil, 2014), we tried to articulate physical and mathematical knowledge (methods and concepts) with phenomena that are specific to life, and worked on some specific "perspectives" on organisms (rhythms, biological time, criticality ...). A tentative "theory of organisms", (Longo, 2015), is further developed in the volume (Soto, 2016v), see in particular the papers (Soto, 2016; Soto, 2016a). In that perspective, DNA is considered as a fundamental, internal "constraint" to cellular and biological activity, where constraints are given the sense described in (Montévil & Mossio, 2015; Mossio, 2016). That is, DNA is a physico-chemical trace of an entire history, punctuated by rare events (Longo, 2017), continually used by the cell dynamics, and thus constraining it to certain proteomics, according to the context. It sets boundaries to the proteome's Brownian motion, under the physical constraints and active canalization mentioned in sect. 3.2, possibly enhanced also by the quantum effects we quoted.

The role of strong, explicit principles in mathematics and physics is crucial. Their role also in biology may be motivated by the reference we made above to fundamental principles in those disciplines. These happen also to be "limit" principles, a status proper to the Darwinian "default state" of organisms: reproduction with variation and motility, which is under ecosystemic and, even more, organismal constraints (Longo, 2015; Soto, 2016a). Mathematical modeling in biology should follow a sound theoretical approach, as in (Montévil, 2016b), and not be based on the passive transfer of tools from mathematical physics. This is how physics produced new mathematics, from Newton to A. Connes, in recent years, as it never happened in mathematical biology.

In (Longo, 2015a), Euclid's "line with no thickness" (a definitional principle made explicit in definition β, book I) and Galileo's principle of inertia are extensively discussed. They are limit constructions, that is they are the infinite limit of decreasing thickness and a limit movement, respectively, as well as founding principles for knowledge construction, far away from common sense. Our quest for principles in biology follows these examples and Darwin's, while acknowledging that the principles specific to physics—grounded on invariance, conservation properties as symmetries, and optimal trajectories—are necessary but not sufficient for the analysis of the proper observables of living beings, organisms and phenotypes. Living systems are in a permanent state of critical transition: their symmetries are continually breaking and being reconstituted, at least at each cell reproduction (Bailly& Longo, 2011; Longo & Montévil, 2014). In our perspective, Darwin's principle of reproduction-with-variation may be seen as a principle of non-conservation of phenotypes, opposed to and symmetric with the principles of conservation and invariance in mathematics and physics, but at the level of the appropriate biological observables, that is, of organisms. Within an organism, it yields the extended state of continual critical transition, at each cell reproduction.

The adequate theorization of the biological field therefore demands extensions and sums of various physical theories— such as the ones due to the coexistence of both classical and quantum random phenomena in the cell (Buiatti & Longo, 2013), of far from equilibrium dynamics (Nicolis & Prigogine, 1977; Kauffmann, 1993), of extended criticality (Bailly & Longo, 2011; Longo & Montévil, 2014). These operations rely on physical

theories and extend their methods, while remaining irreducible to their mathematical techniques. They propose proper biological principles as well as "points of view," and "perspectives" on the organism, whose unity furnishes the guiding thread through these different theoretical aspects. The intelligibility of the biological field is only possible through intersections and partial integrations that aim to construct objects-of-knowledge in dialectical relation with the constraints of experience. In biology, experiences play a singular role, beginning with the difference in vitro vs. in vivo, unknown to physics, and the peculiar role of historical knowledge and, thus, of diachronic measurement in theory building (Longo, 2017). Unity with physical theories (classical, quantum?) may be a long-term goal, surely not a reduction, as hinted at the end of sect. 3.3.

Thanks to mathematization, theorizing in physics extracts generic objects and properties, out of intentional observations and measurement, as conceptual and mathematical invariants. Their objectivity as invariance depends entirely on the theoretical framework: a falling stone or an electron are "generic", that is they are invariant in the theory and for experiences – the analysis of one is sufficient for understanding all cases. In biology, instead, objects are always historic singularities, they are "specific". Objects are grasped by conceptual models, which are qualitative, provisional, and over-determined by their history. The centrality of each singular organism, with its own historicity, implies the primacy of variation and symmetry breaking that overthrow the current mathematical primacy of invariance in physics. This primacy has had very powerful knowledge effects, but it may prove an obstacle to understanding life. It has been further disfigured by the genocentric approach to DNA and the myth of the "program", an informational invariant. For example, the radical materiality of organisms that we mentioned, its historical thickness, and the density of its material internal and external relations, rule out any dualism between "software" and "hardware" and the associated one-dimensionality of digital information, discussed above; these are further mathematical invariants, even more remote from biological phenomenality.

Finally, one of the very conditions of possibility for physical knowledge, the space of phases (the observables and the parameters), is overthrown in biology. In physics, the (phase) space is fixed a priori, a proper one for each physical theory: in classical, quantum, hydrodynamics, thermodynamics … we first pre-define their spaces of analysis, as the Kantian condition of possibility

and immanent norm of physical "trajectories", in the broadest sense. In biological processes, by contrast, the phylogenetic trajectories constitute and constantly reorganize the space of possible dynamics (of phases), the ecosystem. The observables (phenotypes and organisms) are the results of the processes. The historicity of life is grounded on these changes of observables and parameters along evolution (phenotypes and pertinent parameters change), and on the key role of rare events, a peculiarity of historical processes, (Longo, 2017).

If our analysis of living dynamics is pertinent, it poses the problem of how to test the limits of traditional scientific objectivities, of which physics and mathematics represent the paradigms, when facing biological theorization – well beyond the computational parody. Overcoming sound and powerful theoretical practices that are rooted in old, deep metaphysical and theological ideas, (Longo, 2011b), is a radical challenge, but some constructive attempts are seeing the light of day, ours is one of them.

Acknowledgments

References

Aceto, L, Longo, G & Victor B (Editors) 2003 The Difference between Concurrent and Sequential Computations, Special issue, Mathematical Structures in Computer Science, Cambridge U. P., vol.13, n.4 - 5.

Alshaya, W, Mehta, V, Wilson, BA, Chafe S, Aronyk, KE & Lu, JQ 2015, "Low-grade ependymoma with late metastasis: autopsy case study and literature review". Childs Nerv Syst. 31(9) · May.

Anderson, C 2008 "The end of theory: The data deluge makes the scientific method obsolete", WIRED.

Arani, R, Bono, I, Del Giudice, E. & Preparata, G. 1995 "QED coherence and the thermodynamics of water", International Journal Physics B9, 1813.

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

Baccelli, F, Mir-Omid, HM & Khezeli, A 2016a "Dynamics on Unimodular Random Graphs", on arXiv:1608.05940v1 [math.PR].

Baccelli, F & Mir-Omid, HM 2016b "Point-Map-Probabilities of a Point Process and Mecke's Invariant Measure Equation", on arXiv:1312.0287v3 [math.PR].

Bachelard, G 1940 La Philosophie du non, PUF.

Bailly, F & Longo, G 2011 Mathematics and the natural sciences: the physical singularity of life. London: Imperial College Press, (original French version, Hermann, 2006).

Baker, S 2014 "Recognizing Paradigm Instability in Theories of Carcinogenesis", British Journal of Medicine & Medical Research, 4(5): 1149-1163.

Baker, S 2015 "A cancer theory kerfuffle can lead to new lines of research". J. Natl. Cancer Inst. 107, dju405.

Barbaresco, F &Mohammad-Djafari, A (Eds.) 2015 Information, Entropy and Their Geometric Structures, MDPI, Basel & Beijing.

Bezem, M, Klop, JW, Roelde Vrijer, R 2013 Term Rewriting Systems. Cambridge: Cambridge U. Press.

Bizzarri, M, 2014 "System Biology for Understanding Cancer Biology", Curr Synthetic Sys Biol, 2:1.

Bizzarri, M, Cucina, A 2014 "Tumor and the microenvironment: a chance to reframe the paradigm of carcinogenesis". Biomed Res Intl.:934038.

Boveri, T. 1914 Zur Frage der Entstehung maligner Tumoren. Jena: Gustov Fischer.

Buiatti M., Longo, G 2013 "Randomness and Multi-level Interactions in Biology" Theory in Biosciences, vol. 132, n. 3:139-158.

Brading, K., and Castellani, E., eds. 2003 Symmetries in Physics: Philosophical Reflections. Cambridge Univ. Press.

Bravi, B & Longo, G 2015 "The Unconventionality of Nature: Biology, from Noise to Functional Randomness" Unconventional Computation and Natural Computation, Springer LNCS 9252, Calude, Dinneen (Eds.), pp 3-34.

Cairns, J 1981 The origin of human cancers. Nature 289:353–357.

Calude, C 2002 Information and randomness. Springer-Verlag, Berlin, second edition.

Calude, C & Longo, G 2016a "Classical, Quantum and Biological Randomness as Relative Unpredictability". Invited Paper, special issue of Natural Computing, vol. 15, 2, 263–278, Springer, June.

Calude, C &Longo, G 2016 "The Deluge of Spurious Correlations in Big Data". In Found. of Science, 1-18, March.

Chibbaro, S, Rondoni, L & Vulpiani, A 2015 Reductionism, Emergence and Levels of Reality: The Importance of Being Borderline, Springer, Berlin.

Collins, F 1999 "Medical And Societal Consequences Of The Human Genome Project", The New England J. of Medicine, July 1.

Cook, LM, Hurst, DR & Welch, DR 2011 "Metastasis suppressors and the tumor micro-environment." Semin Cancer Biol. 21(2):113–122.

Cortini, R, Barbi, M, Caré, B, Lavelle, C, Lesne, A, Mozziconacci, J & Victor, JM 2016 "The physics of epigenetics", Reviews of Modern Physics, vol. 88, April-June.

Creager, A N & Gaudillière, JP 1996 "Meanings in Search of Experiments and Vice-versa : the invention of allosteric regulation in Paris and Berkeley, 1889-1968", Histor. Studies in the Phisical and Bio. Sciences, 27, 1-89.

Crick, FHC 1956 "On Protein Synthesis" Symp. Soc. Exp. Biol. XII, 139-163.

Danchin, A 2003 The Delphic Boat. What genomes tell us, Harvard University Press.

Danchin, A 2009 "Bacteria as computers making computers", Microbiology Review, 33: 3–26.

Del Giudice, E, Doglia, S, Milani, M & Vitiello, G 1983 "Spontaneous symmetry breakdown and boson condensation in biology", Phys. Lett., 95A, 508.

Del Giudice, E, Doglia, S, Milani, M & Vitiello, G 1986 "Electromagnetic field and spontaneous symmetry breakdown in biological matter", Nucl. Phys., B275, 185.

Di Cosmo, R 1998 Le hold-up planétaire, Calmann-Lévy, Paris.

Duncan, A 2013 "Aneuploidy, polyploidy and ploidy reversal in the liver" Semin Cell Dev Biol. Apr; 24(4):347-56.

Elowitz, MB,Levine, AJ, Siggia, E & Swain, PS 2002 "Stochastic Gene Expression in a Single Cell". Science, 297.

von Eschenbach, A 2003 "NCI Sets Goal Of Eliminating Suffering And Death Due To Cancer By 2015", J Natl Med Assoc., 95:637-639.

Fox Keller, E 2000 The century of the gene, Harvard U. P..

Fromion, P, Leoncini, E & Robert, P 2013 "Stochastic gene expression in cells: A point process approach". SIAM Journal on Applied Mathematics, 73(1):195–211.

Gatenby, RA, Gilles, R & Brown J 2010 "The evolutionary dynamics of cancer prevention", Nat Rev Cancer. Aug; 10(8): 526–527.

Gatenby RA 2011, Of cancer and cave fish, Nature Reviews, Cancer, vol 11, April.

Gatenby, RA 2017 "Is the Genetic Paradigm of Cancer Complete?"Radiology, 284:1–3.

Gerlinger, M et al (22 authors) 2012 "Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing", Engl J Med 366; 10, march 8.

Gilbert, W 1992 "A vision of the Grail" in The Code of Codes: Scientific and Social Issues in the Human Genome Project (Daniel J. Kevles, Leroy E. eds) Harvard U.P..

Gillespie, D 1977 "Exact stochastic simulation of coupled chemical reactions". J. physical chemistry, 81(25):2340–61.

Giuliani, A 2010 "Collective motions and specific effectors: a statistical mechanics perspective on biological regulation" BMC Genomics, 11(suppl 1):S2.

Gouyon, PH, Henry, JP & Arnoud, J 2002 Gene Avatars, The Neo-Darwinian Theory of Evolution, Kluwer.

Gorria, C, Allejo, M &Vega, L 2013 "Discrete conservation laws and the convergence of long time simulations of the mkdv equation", Journal of Computational Physics 235, 274–285.

Griffiths, PE 2001, "Genetic Information: A Metaphor in Search of a Theory", Philosophy of Science, 68: 394–412.

Hanahan, D & Weinberg RA 2000 "The hallmarks of cancer". Cell,100, 57–70.

Herrenschmidt, C 2007, Les trois écritures. Gallimard, Paris.

Huang, S 2014 "The war on cancer: lessons from the war on terror". Front Oncol. 4:293.

Huang, S, Jaurand, MC, Kamp, D, Whysner, J & Heil T 2011 "Role of Mutagenicity in Asbestos Fiber-Induced Carcinogenicity and Other Diseases", J Toxicol Environ Health B Crit Rev. Jan-Jun; 14(1-4): 179–245.

Kato, M, Lippman, SC, Keith, T, Flaherty, H & Razelle, K 2016 "The Conundrum of Genetic "Drivers" in Benign Conditions", J Natl Cancer Inst, 108 (8).

Karp, RD, Johnson, KH, Buoen, LC, Ghobrial, HK, Brand, I & Brand, KG 1973 "Tumorigenesis by Millipore filters in mice: histology and ultrastructure of tissue reactions as related to pore size". J Natl Cancer Inst. 51(4):1275-85.

Kauffman, S. 1993 The origins of order. Oxford U. P.

Kosman-Schwarback, Y 2010 The Noether theorems: Invariance and conservation laws in the twentieth century. Springer-Verlag, Berlin.

Kupiec, JJ 1983 "A probabilistic theory for cell differentiation, embryonic mortality and DNA C-value paradox". Specul Sci Technol 6, 471-478.

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

Kuznetsov, VA, Knott, GD & Bonner, RF 2002 "General statistics of stochastic process of gene expression in eukaryotic cells. Genetics", 161(3):1321–1332.

Kreisel, G 1982 "Four letters to G. Longo" http://www.di.ens.fr/users/longo/files/FourLettersKreisel.pdf

Jacob, F 1965 "Leçon inaugurale", Collège de France, 7 mai.

Jacob, F 1974 "Le modèle linguistique en biologie", Critique, vol. XXX, n. 322.

Jaeger, G 2009 Entanglement, information, and the interpretation of quantum mechanics, Heildelberg: Springer.

Lesne, A 2007 "The discrete vs continuous controversy in physics", Mathematical Structures in Computer Science, vol 17, 2, pp. 185-223.

Lesne, A & Victor JM 2006 "Chromatin fiber functional organization: some plausible models", Eur Phys J E Soft Matter, Mar;19(3):279-90.

Longo, G 2011 "Reflections on Concrete Incompleteness," in Philosophia Mathematica, 19(3): 255-280.

Longo, G 2011b "Mathematical Infinity "in prospettiva" and the Spaces of Possibilities" Visible, Semiotics J., n. 9.

Longo, G, Montévil, M & Kauffman, S 2012 "No entailing laws, but enablement in the evolution of the biosphere". Invited Paper, ACM proceedings of the Genetic and Evolutionary Computation Conference, GECCO'12, Philadelphia (PA, USA), July 7-11.

Longo, G 2014 "Science, Problem Solving and Bibliometrics" Use and Abuse of Bibliometrics, Blockmans (eds), Portland Press.

Longo, G 2015a "The consequences of Philosophy", Glass-Bead, Web Journal, http://www.glass-bead.org/article/the-consequences-of-philosophy/?lang=enview (extended version "Le conseguenze della filosofia", in A Plea for Balance in Philosophy, Lanfredini, R (ed.), ETS, Pisa, 2015).

Longo, G 2016 "A review-essay on reductionism: some reasons for reading "Reductionism, Emergence and Levels of Reality. The Importance of Being Borderline", a book by S. Chibbaro, L. Rondoni, A. Vulpiani. Urbanomic, London, https://www.urbanomic.com/document/on-the-borderline/ , May 8.

Longo, G 2017 "How Future Depends on Past Histories and Rare Events in Systems of Life", Foundations of Science, pp. 1-32 (versione preliminare in italiano in Paradigmi, n. XXXIII, Agosto, 2015, Gagliasso & Sterpetti eds).

Longo, G 2018 "Interfaces of Incompleteness" in Minati, G, Abram, M & Pessa, E (Eds.) Systemics of Incompleteness and Quasi-systems, Springer, New York, NY, to appear (preliminary version: "Incompletezza" per La Matematica, vol. 4, Einaudi (English version in print, downloadable)).

Longo, G, Miquel, PA, Sonnenschein C & Soto A 2012 "Is Information a proper observable for biological organization?" Prog. Biophys. Mol. Biol., Vol. 109, Issue 3, pp. 108-114, August.

Longo, G & Montévil M 2014 Perspectives on Organisms: Biological Time, Symmetries and Singularities. Dordrecht: Springer.

Longo, G & Montévil, M 2017 "Comparing Symmetries in Models and Simulations", Springer Handbook of Model-Based Science, (L. Magnani and T. Bertolotti. Eds), Springer.

Longo, G, Montévil, M., Sonnenschein C & Soto A 2015 "In Search of Principles for a Theory of Organisms", Journal of Biosciences, Springer, pp. 955–968, 40(5), December.

Longo, G & Soto, A 2016 "Why do we need theories?" Prog. Biophys. Mol. Biol., 122, 1, 4-10, Soto, Longo & Noble eds.

Longo, G & Tendero, PE 2007 "The differential method and the causal incompleteness of Programming Theory in Molecular Biology" Foundations of Science, n. 12, pp. 337-366.

Maffini, MV, Calabro, JM, Soto & AM, Sonnenschein, C 2005 "Stromal regulation of neoplastic development: Age-dependent normalization of neoplastic mammary cells by mammary stroma". Am. J. Pathol. 167, 1405-1410.

Maltoni, C, Lodi, P, Masina, A, 1986 "Mesoteliomi negli operai di officine di grandi riparazioni (OGR) delle Ferrovie dello Stato italiane, esposti ad asbesto". Primo resoconto. Acta Onco:., 7: 159-86.

Marinov, G.K., Williams, B.A., McCue, K., Schroth, G.P., Gertz, J., Myers, R.M. & Wold, B.J. 2014 "From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing". Genome Res. 24, 496-510.

Martincorena, I (15 more authors) 2015 "High burden and pervasive positive selection of somatic mutations in normal human skin", Science, Vol. 348 no. 6237 pp. 880-886, May.

Maynard-Smith, J 1999 "The idea of Information in Biology" The Quarter Rev of Biology 74: 495-400.

Mroue, R & Bissel, MJ 2013 "Three-dimensional cultures of mouse mammary epithelial cells" Methods Mol Biol.; 945:221-50. doi: 10.1007/978-1-62703-125-7_14.

Microsoft 2016 http://news.microsoft.com/stories/computing-cancer/ accessed on October, 31.

Monod, J 1970 Le Hasard et la Nécessité, PUF.

Montévil, M & Mossio, M 2015 "Closure of constraints in biological organisation", Journal of Theoretical Biology, vol. 372: 179-191.

Montévil, M, Mossio, M, Pocheville, A & Longo, G 2016 "Theoretical principles for biology: Variation", Prog. Biophys. Mol. Biol., 122, 1, 36-50, Soto, Longo & Noble eds.

Montévil, M & Pocheville, A 2018 "The Hitchhiker's Guide to the Cancer Galaxy. How two critics missed their destination", Organisms, this issue.

Montévil, M, Speroni, L, Sonnenschein, C & Soto AM 2016b "Modeling mammary organogenesis from biological first principles: cells and their physical constraints", Prog. Biophys. Mol. Biol., 122, 1, 58-69, Soto, Longo & Noble eds., 2016.

Mossio, M, Montévil, M & Longo, G 2016 "Theoretical principles for biology: Organization", Prog. Biophys. Mol. Biol., 122, 1, 24-35, Soto, Longo & Noble eds.

Nicolis, G & Prigogine I. 1977 Self-organization in non-equilibrium systems. New York, Wiley.

Nowell, PC 1976 "The clonal evolution of tumor cell populations". Science, 194(4260):23–28.

Onuchic, J, Luthey-Schulten, Z &Wolynes, P 1997 "Theory of protein folding: The Energy Landscape Perspective", Annual Review of Physical Chemistry, Vol. 48: 545-600.

Paldi, A 2003 "Stochastic gene expression during cell differentiation: order from disorder?" Cell Mol. Life Sci., 60, 1775-1779.

Paloma AM 2012 Molecular and Cellular Endocrinology 355; 201–207.

Perret, N & Longo, G 2016 "Reductionist perspectives and the notion of information", Prog. Biophys. Mol. Biol., 122, 1, Soto, Longo & Noble eds.

Pauling, L 1987 "Schrödinger contribution to Chemistry and Biology", Schrödinger: Centenary Celebration of a Polymath (Kilmister ed.) Cambridge U. P..

Perutz, MF 1987 "E. Schrödinger's What is Life? and molecular biology", Schrödinger: Centenary Celebration of a Polymath (Kilmister ed.) Cambridge U. P..

Pisco, O & Huang, S 2015 "Non-genetic cancer cell plasticity and therapy-induced stemness in tumour relapse: 'What does not kill me strengthens me'", British Journal of Cancer, 112, 1725–1732.

Pocheville, A & Montévil, M 2017 "Intervening in a continuous world: when Achilles meets his tortoise'' submitted.

Richard E. (15 authors) 2016 "Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Proces". Plos Biology, December 27.

Raj, A, Peskin, CS, Tranchina, D, Vargas, DY & Tyagi, S 2006 "Stochastic mRNA synthesis in mammalian cells". PLoS Biol 4:e309.

Raj, A, van Oudenaarden R 2008 "Stochastic Gene Expression and its Consequences". Cell, 135(2): 216–226.

Rogers, H 1967 Theory of Recursive Functions and Effective Computability, McGraw Hill.

Sarpeshkar, R 1998 "Analog versus digital: extrapolating from electronics to neurobiology", Neural Comput 10, 1601-38.

Sarti, A, Citti, G, Piotrowski, D 2018 "Differential heterogenesis and the emergence of semiotic function". Submitted

Schiffman, JD, Maley, CC, Nunney, L, Hochberg, M & Breen, M, 2013 "Peto's Paradox and the promise of comparative oncology", Phil Trans Royal Society of London, B. 370: 20140177.

Smorinski, C 1978 "The incompleteness theorem", Handbook of Mathematical Logic (Barwise ed.), North Holland.

Sonnenschein, C & Soto, AM 1999 The society of cells: cancer and control of cell proliferation. Springer.

Sonnenschein, C & Soto, AM 2011 "The Death of the Cancer Cell", Cancer Res.; 71:4334-4337.

Sonnenschein, C & Soto, AM 2013 "The aging of the 2000 and 2011 hallmarks of cancer reviews: a critique", Journal of Biosciences; 38:651-63.

Sonnenschein, C & Soto, AM 2015 "Cancer Metastases: So Close and So Far" J. Natl Cancer Inst, 107(11).

Sonnenschein, C, Davis, B & Soto AM 2014 "A novel pathogenic classification of cancer", Cell Cancer Int. 14: 113.

Sonnenschein, C, Soto, AM 2016 "Carcinogenesis explained within the context of a theory of organisms". Prog Biophys Mol Biol. Oct;122(1):70-76.

Soto, AM, Longo, G & Noble, D, eds. 2016v From the century of the genome to the century of the organism: New theoretical approaches. Special issue, Prog. Biophys. Mol. Biol., 122, Issue 1, Pages 1-82,

Soto, AM, Longo, G, Montévil & M, Sonnenschein, C 2016 "The biological default state of cell proliferation with variation and motility, a fundamental principle for a theory of organisms", Prog. Biophys. Mol. Biol., Soto, Longo, Noble eds., 122(1):16-23.

Soto, AM, Longo, G, Miquel, PA, Montévil, M, Mossio, M, Perret, N, Pocheville, A & Sonnenschein, C 2016a ''Toward a theory of organisms: Three founding principles in search of a useful integration", Prog. Biophys. Mol. Biol., Soto, Longo & Noble eds, 122, 77-82.

Soto, AM & Sonnenschein, C 2010 "Environmental causes of cancer: endocrine disruptors as carcinogens", Nat Rev Endocrinol, Jul;6(7):363-70.

Soto, AM & Sonnenschein, C 2011 "The tissue organization field theory of cancer: A testable replacement for the somatic mutation theory", Bioessays 33: 332–340.

Soto, AM & Sonnenschein, C 2017 "Why is it that despite signed capitulations, the war on cancer is still on?" Organisms, Journal of Biological Sciences, Vol. 1, 1.

Sporn, MB 1996 "The war on cancer" Lancet; 347(9012):1377–1381

Stanford Encyclop. of Philosophy 2016 accessed on Oct., 31: http://plato.stanford.edu/entries/information-biological/

Straus, DS 1981 "Somatic mutation, cellular differentiation, and cancer causation". J. Natl. Cancer Inst. 67:233–241.

Venter, C 2010 "We Have Learned Nothing from the Genome'', Der Spiegel, July 29.

Versteeg, R 2014 "Tumors outside the mutation box", Nature, vol. 1.

Vogelstein, B, Papadopoulos, N, Velculescu, VE, Zhou, S, Diaz, Jr LA & Kinzler, KW 2013 "Cancer genome landscapes", Science, 339(6127):546-58.

West-Eberhard, MJ 2003 Developmental plasticity and evolution. Oxford Univ. Press, New York.

Weinberg, R 2014 "Coming Full Circle - form endless complexity to simplicity and back again'', Cell 157, March 27.

Wolfram, S 2013 "The importance of Universal Computation", in A. Turing, his work and impact, Cooper ed.,Elsevier.

Zhang, XH, Jin, X & Malladi, S 2013 "Selection of bone metastasis seeds by mesenchymal signals in the primary tumor stroma", Cell.;154(5):1060–1073.York.