Randomness increases biological organization: a mathematical understanding of Gould's critique of evolutionary progress[1]

Giuseppe Longo[2] Maël Montévil[3]

July 28, 2012

Abstract

In this text, we informally expose the mathematical analysis in [BL09] of Gould's ideas on the increase of "complexity" along biological evolution as a result of random paths. We also revisit some related theoretical investigations on randomness (contingency) and symmetry breakings in biology, following [LM12]. Gould made several fundamental observations on how phenotypic complexity increases on average, in a random evolution, without a bias towards an increase. Technically, we understand complexity as anti-entropy, a proper biological observable. Its increase, by symmetry changes, involves a strong form of randomness. In more usual biological terms, an increase of complexity involves new elements of biological organization and the latter are "contingent" because they are not determined by the current state of the life dynamics.

1 Randomness and Complexification in Evolution.

Available energy consumption and transformation, thus entropy production, are the unavoidable physical processes underlying all biological activities, including reproduction with variation. At the origin of life, bacterial exponential proliferation was (relatively) free, as other forms of life did not contrast it. Diversity, even in bacteria, by random differentiation, produced competition and a slow down of the exponential growth (see diagram 3). Simultaneously, though, this started the early variety of live, a process never to stop.

Gould, in several papers and in two books [Gou89, Gou97], uses this idea of random diversification in order to understand a blatant but too often denied fact: the "complexification" of life. The increasing complexity of biological structures has been often denied in order to oppose finalistic and anthropocentric perspectives, which viewed life as *aiming* at *Homo sapiens* as the "highest" result of the (possibly intelligent) evolutionary path (or design).

Yet, it is a fact that, under many reasonable measures, an eukaryotic cell is more "complex" than a bacterium; a metazoan, with its differentiated cells, tissues and organs, is more "complex" than a cell … and that, by counting neurons and their connections, cell networks in mammals are more complex that in early triploblast (which have three tissues layers) and these have more complex networks of all sorts than diplobasts (like jellyfish, a very ancient life form). This non-linear increase can be quantified by counting tissue differentiations, networks and more, as hinted by Gould and more precisely proposed in [BL09]. We will first summarize and comment this latter paper here. The point is: how are we to understand this change towards complexity without invoking global aims?

Gould provides a remarkable answer based on the analysis of the *asymmetric* random diffusion of life. Asymmetric because, by principle, life cannot be less complex than bacterial life[4]. So, reproduction with variability, along evolutionary time and in an abstract space, randomly produces more complex

1 Invited lecture, to appear in the proceedings of the Conference, **Stephen J. Gould heritage: Nature, History, Society**, *Venice (It.)*, May 10 - 12 , 2012

2 Centre Cavaillès, CIRPHLES, CNRS et Ecole Normale Sup., Paris http://www.di.ens.fr/users/longo/

3 Tufts University, Department of Anatomy and Cellular Biology, Boston

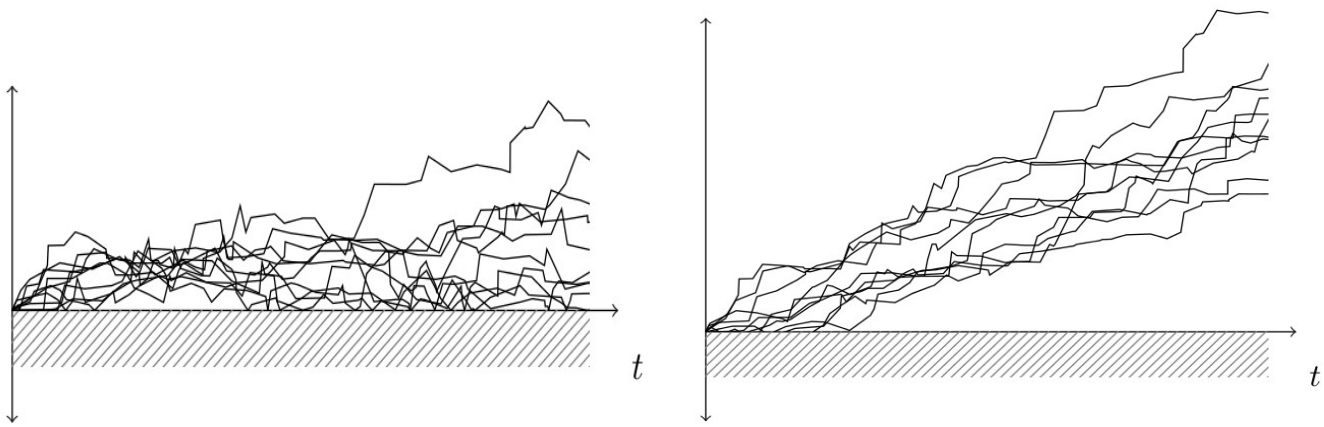
4 Some may prefer to consider viruses as the least form of life. The issue is controversial, but it would not change at all Gould's and our perspective: we only need a minimum biological complexity which differs from inert matter

individuals just as *possible paths*. Some happen to be compatible with the environment, resist and proliferate (a few even very successfully) and keep going further and randomly producing *also* more complex forms of life. *Also*, since the random exploration of possibilities may, of course, decrease the complexity, no matter how this is measured. Yet, there is a key general principle by which we will understand Gould's analysis:

Any asymmetric random diffusion propagates, by local interactions, the original symmetry breaking along the diffusion.

Typically, in a liquid, a drop of dye against a (left) wall, diffuses (towards the right) by local bumps of the particles against each other. That is, particles transitively inherit the original (left) wall asymmetry and propagate it *globally* by *local* random interactions.

Thus there is no need for a global design or aim: the random paths that compose *any* diffusion, also in this case help to understand a random growth of complexity, *on average*. On average, as, of course, there may be local inversion in complexity; yet, the asymmetry randomly forces to "higher complexity" (a notion yet to be defined). This is beautifully made visible by figure 1, in [Gou89], page 205. The image explains the difference between a random, but oriented development (the right figure, 1b, along time, the abscissa axis), and the non-biased, purely random diffusive bouncing of life expansion on the lower wall of complexity, which is "quantified" on the ordinate axis, the left figure 1a.



(a) Passive trend, there are more trajectories near 0.

(b) Driven trend, the trajectories have a drift towards an increased mean.

Figure 1: *Passive and driven trends*. In one case, the *boundary condition*, materialized by a lower wall for complexity, is the only reason why the mean increases over time. As a result, this increase is slow. However, in the case of a driven trend or biased evolution, it is the *rule* of the random walk that leads to an increase of the mean over time (there would be an intrinsic trend in evolution), and the increase of the mean is linear as a function of time. Gould's and our approach are based on passive trends, which means that we do not need any intrinsic bias for increasing complexity in the process of evolution. Our work has been to justify and frame mathematically Gould's beautiful intuition, in short here, extensively in [BL09].

Of course, time runs on the horizontal axis, but … what is in the vertical one? Anything or, more precisely, anywhere the random diffusion takes place or the intended phenomenon "diffuses in". In particular, the vertical axis may quantify "biological complexity" whatever this may mean. The point Gould wants to clarify is the difference between a fully random vs. a random *and* biased evolution. The biased right image does not apply to evolution: bacteria are still on Earth and very successfully. Any

finalistic bias would instead separate the average random complexification from the lower wall.

In either cases, as we said, complexity may *locally* decrease: tetrapodes may go back to the sea and lose their podia (the number of folding decreases, the overall body structure simplifies). Some cavern fishes may loose their eyes, in their new dark habitat; others, may lose their red blood cells [Ruu54]. Thus, the local propagation of the original asymmetry may be biologically understood as follows: on average, variation by simplification may lead towards a biological niches that has *more chances* to be already occupied. Thus, *global* complexity increases *as a purely random consequence of variability* and on the grounds of *local effects*: the greater chances, for a "simpler" organism, to bump against an already occupied niche. Thus, more complex variants have just slightly more chances to survive and reproduce — but this slight difference is enough to produce, in the long run, very complex biological organisms[5].

Note that, if variability and, thus, diversity are grounded on randomness, then randomness contributes to structural stability, in biology: diversity is a component of the stability of a species, a population, even an organism (e. g. the irregularities in the fractal structure of an organ may contribute to the organism adaptivity, a concept to be developed elsewhere).

That is, by a sound analysis of randomness, as "contingency" in Gould's sense, complexity and diversity increase with no need for finalism nor a priori "global aim" nor "design" at all. They increase just a consequence of an original symmetry breaking in a random diffusion on a very peculiar phase space:

bio-mass × complexity × time (see figure 3 for a complete diagram).

Consider now that both in embryogenesis and in evolution, increasing complexity is a form of local reversal of entropy. The global entropy of the Universe, as energy dispersal (or disorder, in biology) increases (or does not decrease). However, locally, by using energy of course, life inverses the entropic trend and creates (new) organization or increases complexity. Of course, embryogenesis is a more canalized process, while evolution seems to explore a diversity of "possible" paths, within the ecosystem-to-be. Most turn out to be incompatible with the environment, thus they are eliminated by selection. In embryogenesis increasing complexity seems to follow an expected path and it is partly so. But only in part as failures, in mammals say, reach 50% or more: the constraints imposed, at least, by the inherited DNA and zygote (and by the ecosystem as well), limit the random exploration due to cell proliferation. Yet, their variability, jointly to the many (variable) constraints present in development (first, a major one: DNA), is an essential component of cell differentiation. Tissue differentiation is, from our point of view, a form of (strongly) regulated/canalized variability along cell reproduction.

In conclusion, by different but correlated effects, complexity increases, on average, and reverts entropy locally. In [BL09], we called *anti-entropy* this observable opposing entropy, both in evolution and embryogenesis; its peculiar nature is based on reproduction with random variation, submitted to constraints. As observed in [LM12], anti-entropy differs from negentropy, which is just entropy with a negative sign. First when anti-entropy is added to entropy, the sum never gives 0. Moreover, anti-entropy is realized in a very peculiar singularity (different from 0): a non-null interval of extended criticality [BL11, LM11]. In the next section, we will use this notion to provide a mathematical frame for a further insight by Gould.

2 (Anti-)Entropy in Evolution.

2.1 The diffusion of Bio-mass over Complexity.

In yet another apparently naive drawing, Gould proposes a further visualization of the increasing complexity of organisms along evolution. It is just a qualitative image that the paleontologist draws on the

5 This approach does not exclude adaptive effects, which may lead *both* to greater or even lower complexity, but are locally successful.

grounds of his experience. It contains though a further remarkable idea: it suggests the “phase space” (the space of description: observables and parameters) where one can analyse complexification. It is *bio-mass density* that diffuses over *complexity*, that is, figure 2 qualitatively describes the diffusion of the frequency of occurrences of individual organisms per unity of complexity.

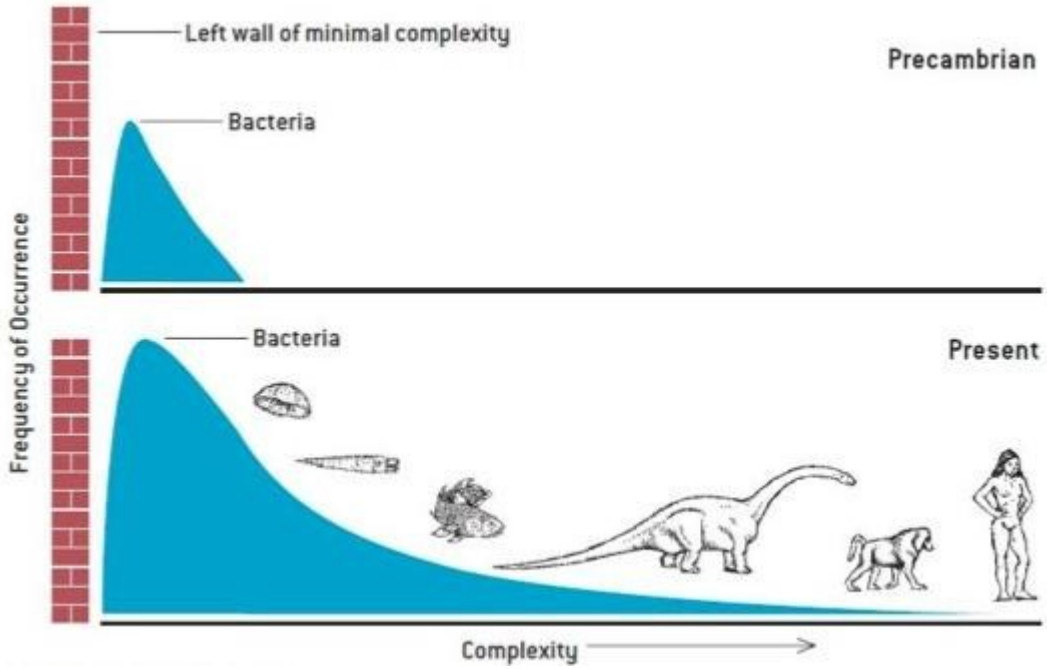


Figure 2: *Evolution of complexity as understood by Gould.* This illustration is taken from [Gou97], page 171.

This is just a mathematically naive, global drawing of the paleontologist on the basis of data. Yet, it poses a conceptual challenge. The diffusion, here, is not along a spatial dimension. Physical observables usually diffuse over space in time; or, within other physical matter (which also amounts to diffusing in space). Here, the diffusion of bio-mass density (or number of occurrences of individual organisms weighted by their individual mass, the ordinate axis) takes place over an abstract dimension, “complexity”, on the abscissa axis. But what does biological complexity exactly mean? Hints are given in [Gou97]: the addition of a cellular nucleus (from bacteria to eukaryotes), the formation of metazoa, the increase in body size, the formation of fractal structures (usually — new — organs) and a few more…. In a sense, any added novelty, provided by random “bricolage” and Gould’s “exaptation” along evolution contributes to complexity when they are at least for some time compatible with the environment. Only a few organisms become more complex over time, but, because of the original symmetry breaking mentioned above (and now represented by the *left* wall of complexity, where bacteria are), this is enough to increase the global complexity.

Of course, the figure above is highly unsatisfactory. It gives two slices over time where the second one is somewhat inconsistent: where are dinosaurs at present time? It is just a sketch, but an audacious one if analyzed closely. Mathematics may help us to consistently add the third missing dimension: time.

A simple form of diffusion equation of q in time t over space x is:

$$\frac{\partial q}{\partial t} = D\frac{\partial^2 q}{\partial x^2} + Q(t,x) \qquad (1)$$

where $Q(t,x)$ is a source term describing the increase of biomass.

Yet, in our case, the diffusion of this strange quantity, m, a *bio-mass density*, takes place over an even

more unusual “space”, biological complexity, whatever the latter may mean. In [BL09], we dared to further specify Gould’s hints for biological complexity, as a quantity

$$K = \alpha K_c + \beta K_m + \gamma K_f \tag{2}$$

where α, β, and γ are the respective “weights” of the different types of complexity within the total complexity. The details are in [BL09], let’s just summarize the basic ideas.

K_c (“combinatorial” complexity) corresponds to the possible cellular combinatoric (number of cell types or tissues in a metazoan); K_m (“morphological” complexity) is associated to the forms which arise (connexity and fractal structures); K_f (“functional” complexity) is associated to the relational structures supporting biological functions (metabolic and neuronal relations). We will discuss further this approach in section 4 below.

K is a tentative quantification of complexity, in particular in biological evolution: the increase of each of its components (more cellular differentiation, more or higher dimensional fractal structures, richer networks …) yield a more “complex” individual. Note that K is meant to give a purely static account of complexity: it is anatomy on a fixed body, no functions are considered. This may allow a distinction between biological complexity and “biological organization[6]”. Of course, many more observables and parameters may be taken into account in order to evaluate the complexity of an organism: [BL09] provides just a mathematical basis and a biological core for a preliminary analysis (an application to ontogenesis as an analysis of *C. Elegans* development is also presented in [BL09]). They suffice though for a qualitative (geometric) reconstruction of Gould’s curve, with a sound extension to the time dimension.

As mentioned above, K opposes, locally, to entropy: we called it anti-entropy since it has the same physical dimension as entropy, yet it differs from negentropy, since anti-entropy and entropy do not sum up to 0, when their amount are equal. It differs also from information theoretic frame, where negentropy has been largely used, as negentropy (= information, in Shannon-Brillouin’s approach) is *independent from coding and Cartesian dimensions*. This is crucial for Shannon Brillouin’s as well as for Kolmogorof-Chaitin’s information theories. Anti-entropy, instead, as defined above, depends on foldings, singularities, fractality … it is a *geometric* notion, thus, by definition, it is *sensitive to codings* (and to dimension).

The next step is to adapt eq. 1 to these new dimensions. Just use Gould’s observables and parameters, m, bio-mass density, and K, that we specified some more, and write:

$$\frac{\partial m}{\partial t} = D\frac{\partial^2 m}{\partial K^2} + Q(t, K) \tag{3}$$

But what is here $Q(t,K)$, the source term? In order to instantiate Q by a specific function, but also in order to see the biological system from a different perspective (and get to the equation also by an “operatorial approach”), we then gave a central role, as an observable, to the “global entropy production”.

In biology, constitutive processes, such as anti-entropy growth (the construction and reconstruction of organization), also *produce entropy*, since they also produce some (new) disorder. Proliferation goes with variation: even in a single mitosis, at least the proteome is non-uniformly and randomly distributed in the new cells, see for example [LMD+08, CFX06]. Thus increasing order, from one to two cells, produces also some new disorder: mitosis is never perfectly symmetric, like in crystal formation, at least in principle[7].

6 We were very interested in seeing this distinction of ours recently applied by two biologists of cancer, Carlos Sonnenschein and Ana Soto: in general, in an organism, cancer increases complexity, but decreases organization.

7 Note the difference: crystallography views symmetry and invariance as key principles, where noise is a side ``friction" with respect to the mathematical analysis. For us, following Darwin and Gould, *reproduction with modification*, thus variation and symmetry changes (even very small changes: in a mitosis), is the key principle of intelligibility for evolution

One more change of perspective. In physics, energy, *E*, is the "main" observable, from Galileo's inertia, a principle of energy (or momentum) conservation, to Noether's theorems and Schrödinger's equation. Equilibria, conservation properties, geodetic principles … directly or indirectly refer to energy and are understood in terms of symmetry principles (see [BL11]). At least since Schrödinger and his equation, in (quantum) physics, one may view energy as an operator and time as a parameter[8].

In these far from equilibrium, dissipative (possibly even non-stationary) processes, such as evolution and ontogenesis, energy turns out to be just one (very important) parameter. Typically it is a parameter in allometric equations: the bigger we are, the slower we get, see [LCI81]. Of course, production and maintenance of organization requires energy, but it yields a different observable, one that has a different dimension, tentatively defined by *K* above, anti-entropy. Thus, in our approach, the key observable is complexity that is formed or renewed (anti-entropy formation).

Note now that *entropy production* is associated to all irreversible processes, from energy flows to anti-entropy growth: as mentioned above, increasing complexity also, in evolution and embryogenesis, is accompanied by a (slight) production of disorder, thus of entropy. Entropy (some new disorder) results also from the asymmetry of two cells after a mitosis, as we said above, to, more generally, "reproduction with modification". Of course, entropy also decreases locally, when one organized being gets two, but some new entropy is produced as the two are slightly disorder (a modification happened). Thus, entropy, both as the result of energy flows (purely physical) and of anti-entropy production (essentially biological), is the observable which summarizes all ongoing phenomena; by its irreversibility, it is strongly linked (technically: conjugated) to time.

In summary, we proposed to change the conceptual frame and the conceptual priorities: we associated the global entropy production σ to the differential operator given by time, $\partial/\partial t$ (Schrödinger does this for energy, which is conjugated to time, in quantum physics). Thus, our approach allows to consider biological time as an "operator", both in this technical sense and in the global perspective of attributing to time a key constitutive role in biological phenomena, from evolution to ontogenesis. But how to express this global observable?

2.2 A Balance Equation.

In a footnote to [Sch44], Schrödinger proposes to analyse his notion of negative entropy as a form of Gibbs free energy *G*. In order to have a rigorous use of dimensions and use anti-entropy in equations, we applied Schrödinger's idea by setting $S^- = -kK$ (k is a positive dimensional constant and K is our phenotypic complexity). Now, Gibbs free energy may be given as $G = H - TS$, where T is temperature, S is entropy and $H = U + PV$ is the system's enthalpy (U is the internal energy, P and V are respectively pressure and volume).

By definition, the *metabolism R* has the physical dimension of a power and corresponds to the difference between the fluxes of *generalized free energy G* through the surface Σ:

$$R = \sum [J_G(x) - J_G(x + dx)] = - \sum dx (\mathrm{Div}\, J_G) \qquad (4)$$

Locally the conservation (or balance) equation is expressed in the general form:

$$R = - \mathrm{Div}\, J_G = \frac{dG}{dt} + T\sigma \qquad (5)$$

where σ represents the speed of global production of entropy, that is σ is the entropy produced by *all*

(plus *selection*, of course).

8 In short, Schrödinger transforms an equation with the structure $E = \frac{p^2}{2m} + V(x)$, where $V(x)$ is a potential, by associating E and p to the differential *operators* $\partial/\partial t$ and $\partial/\partial x$, respectively, see[BL09].

irreversible processes, including the production of biological complexity or anti-entropy. Thus, the global balance of metabolism for the "system of life" (the biosphere) has the following form, where S^- and S^+ are anti-entropy and entropy, respectively:

$$R = \frac{dH}{dt} - T\left(\frac{dS^-}{dt} + \frac{dS^+}{dt}\right) + T\sigma \simeq a\frac{dM}{dt} - T\left(\frac{dS^-}{dt} + \frac{dS^+}{dt}\right) + T \tag{6}$$

where $H \simeq aM$, for a mass M and a coefficient a, which has the magnitude of a speed squared.

$T\sigma$ is a crucial quantity: it contains entropy production σ, modulo the temperature T, since R is a power. $T\sigma$ corresponds to the product of forces by fluxes (of matter, of energy — chemical energy, for instance — etc.). Now, a flux is proportional to a force, thus to a mass, and hence $T\sigma$ is proportional to a mass squared. It can then be written, up to a coefficient ζ_b and a constant term $T\sigma_0$ as:

$$T\sigma \approx \zeta_b M^2 + T\sigma_0 \tag{7}$$

ζ_b is a constant that depends only on the global nature of the biological system considered and it is 0 in absence of living matter.

Without entering into further details (see [BL09]), by using as "state function" a *bio-mass diffusion function* over complexity K, that is the bio-mass density $m(t,K)$ as a function of t and K, the operatorial approach applied to equation 7 gave us the equation, with a linear source function $\alpha_b m$:

$$\frac{\partial m}{\partial t} = D_b \frac{\partial^2 m}{\partial K^2} + \alpha_b m \tag{8}$$

Its solution, 9, yields the diagram in figure 3.

$$m(t,K) = \frac{A}{\sqrt{t}} \exp(at) \exp(-K^2/4Dt) \tag{9}$$

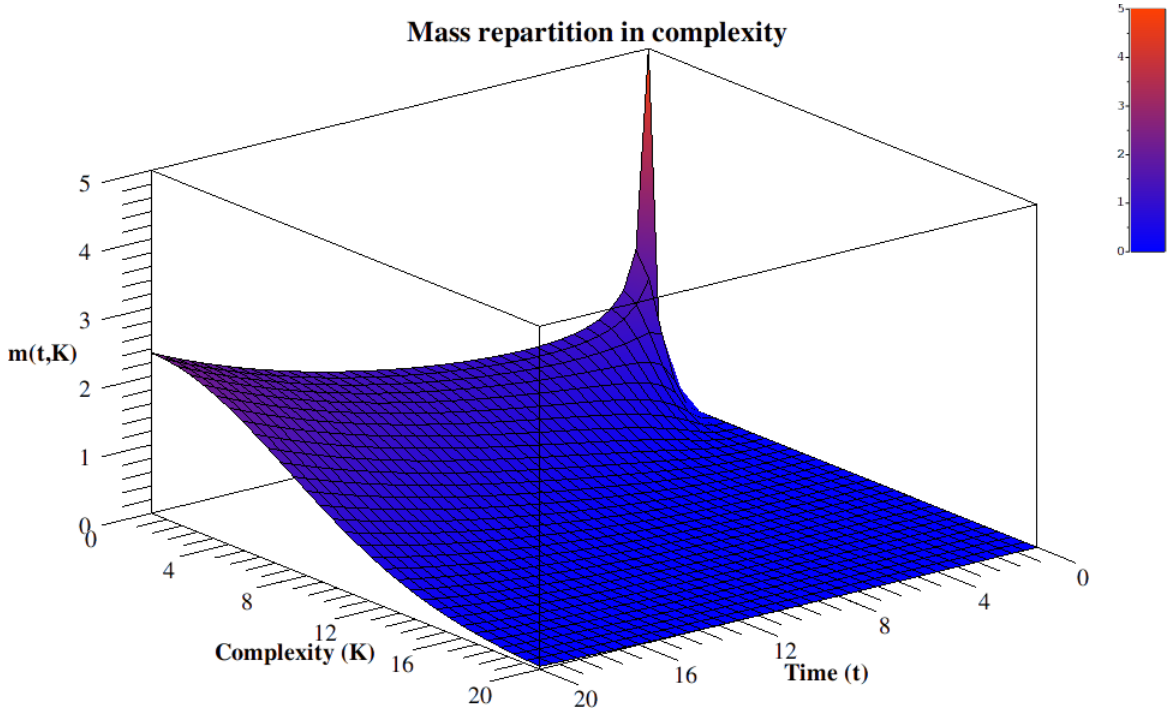


Figure 3: *Time evolution of mass repartition over anti-entropy*. The initial condition is a finite mass at almost 0 anti-entropy, leading to very high concentration of biomass in our phase space.

In summary, by skipping all the technical details in [BL09], we could derive, by mathematics and starting

from Gould's informal hints, a general understanding as well as the behavior of the "evolutionary complexity function" w. r. to time. And this fits data: at the beginning the linear source term gives an exponential growth of free bacteria. Then, they diversify, complexify and compete.

Of course, this diagram, similarly to Gould's, is a global one: it only gives a qualitative, geometric, understanding of the process. It is like looking at life on Earth from Sirius. Analogously to Gould's diagram, the "punctuated equilibria", say, and the major extinctions are not visible: the insight is from too far and too synthetic to appreciate them. It only theoretically justifies Gould's proposal and soundly extends it to time dependence, by mathematically deriving it from general principles: the dynamics of a diffusion by random paths, with an asymmetric origin, yields an increasing phenotypic complexity. Its source is given by an exponential growth. Life expansion, as reproduction with modification, is then constrained, canalized, selected in the interaction with the ever changing, co-constituted ecosystem. The core random complexification persists, while its "tail" exponentially decreases, see equation 9 and figure 3. In that tail, some neotenic big primates, with a huge neural network, turn out to be the random complexification of bacteria, a result of variability and of the immense massacres imposed by selection.

Another (important) analogy can be made with Schrödinger's approach (his famous equation, not his book on life) and further justifies the reference to it for the analysis of this (rather ordinary) diffusion equation. Schrödinger dared to describe the state of quantum systems as a wave function following a deterministic evolution and interpreted this as the *dynamic of a probability density* (and the intrinsic indetermination of the quantum system follows). We synthetically represented evolutionary dynamics of *bio-mass density* as the *dynamics of a potential of variability*, under the left wall asymmetric constraint (minimal or bacterial complexity). Again, this idea is essentially Gould's idea in his 1996 book: he sees evolution just as an asymmetric diffusion of random variability. We just made this point explicit and developed some computations as a consequences of the analogy with Schrödinger's equational determination and the operatorial approach in Quantum Mechanics[9].

In particular, we looked at bio-mass abstractly, as a potential of variability, whose random diffusion over complexity led, among others, to our contingent human being.

3 Theoretical consequences of this interpretation

Gould's understanding of evolution stresses Darwin's two key principles: *reproduction with modification* and *selection of the incompatible*. Moreover, following Gould, sudden (*punctuated*) "bursts" of change radically accelerate the global evolutionary modifications, while *exaptations* provide a further path towards modifications. All these principles rely on contingent dynamics, without excluding the structural stability of organisms, species, ecosystem, which are also essential to life, of course. It is stability via change, or, more exactly, randomness produces variability, which yields diversity (of populations, species, ecosystems) which is an essential component of biological stability. More generally, as mentioned in [LMK], evolutionary trajectories are based on a "non-conservation principle" for phenotypes: reproduction with modification, its punctuated dynamical equilibria[10], and selection, as a largely contingent result of interactions (organism/ecosystem). This theoretical frame is in strong contrast to the main physical theories, which are largely based on conservation principles (energy, momentum ... conservation), and the related theorems on symmetries in the equations, since Noether's, see [BL11]).

---

9 By our approach, we provide a theoretical/mathematical justification of the ZFEL principle in [MB10], at the core of their very interesting biological analysis: "ZFEL (Zero Force Evolutionary Law, general formulation): In any evolutionary system in which there is variation and heredity, there is a tendency for diversity and complexity to increase, one that is always present but may be opposed or augmented by natural selection, other forces, or constraints acting on diversity or complexity." That is, this principle is derivable from our analysis, but one should look at [BL09] for more details.

10 From the point of view of physics, organisms, species, ecosystems … are always far from equilibrium, as they need a flow of energy. The relative equilibrium may be referred to phenotypes, as observables: little changes of anti-entropy only.

In [LM11], we proposed to understand biological systems as characterized by a cascade of symmetry changes. Now, our understanding of a "biological trajectory", a phylogenetic and ontogenetic path, as a cascade of symmetry changes yields a proper form of randomness to be associated to the construction and maintenance of biological complexity and organization. This perspective is particularly relevant for us, since it links the two theoretical approaches of the living state of matter that our team has introduced: anti-entropy [BL09] and extended criticality [BL08, LM11].

More precisely, in phylogenesis, randomness is associated to the "choice" of different organizational forms, which occurs even when the biological objects are confronted with remarkably similar physical environment and physiological constraints. For example, the lungs of birds and mammals have the same function in similar conditions; but they have phylogenetic histories which diverged long ago and, extremely different structures.

This example is particularly prone to lead to approximate common symmetries, since it relates to a vital function (respiration and therefore gas exchanges) shared by a wide class of organisms. It is noteworthy that numerous theoretical studies have analysed lungs by optimality principles [Hor77, WBE99, GKPC05]. However, the optimality principles differ in these studies (minimum entropy production, maximum energetic efficiency, maximum surface/volume ratio, …). Accordingly, even among mammals, structural variability remains high. For example, [NWG90] describe the differences in the geometrical scaling properties of human lungs on one side, and of rats, dogs and hamsters lungs on the other side. [MFWS04] show that the criteria of energetic optimality and of robustness for the gas exchanges, with respect to geometric variations, are incompatible. More generally, optimization criteria are not particularly theoretically stable. In particular robustness is a relative notion: it depends on the property considered as well as on the transformations with respect to which we expect it to be robust [Les08]. As a result, the symmetry of the objects considered are loose at most.

Similarly, the theoretical symmetries constituted in ontogenesis are the result of the interactions with the environment and of the developmental trajectory already followed at a given time. In our perspective, this trajectory must then be understood as a history of symmetry changes. And, of course, the situation at a given moment does not "determine" the symmetry changes that the object will undergo. This is a crucial component of the randomness of the biological dynamics, as we consider that random events are associated to symmetry changes. These events are given by the interplay of the organism with its own physiology (and internal milieu) and with its environment, the latter being partially co-constituted by the theoretical symmetries of the organism, since many of the relevant aspects of the environment depend also on the organism.

In other terms, the conservation, in biology, is not associated to the biological *proper observables*, the phenotype, and the same (physical) interface (e.g. energy exchange) with the environment may yield very different phenotypes; thus, there is no need to preserve a specific phenotype. In short, the symmetry changes occurring in an organism can only be analysed in terms of the previous theoretical symmetries (biology is, first, an historical science) and the differences of the possible changes can be associated to different forms of randomness:

- In the cases of a symmetry breakings *stricto sensu*, the symmetry change corresponds to the passage to a subgroup of the original symmetry group. As a result, the theoretical possibilities are, in principle predefined (as the set of subgroups of the original group). This typically occurs in the case of physical phase transitions, and the result is then a randomness associated to the choice of how the symmetry gets broken: symmetric elements have to become different. For example, if an organism has an approximate rotational symmetry, this symmetry can be broken in a subgroup, by providing a particular oriented direction. We then have a rotational symmetry along an axis. This can again be broken, for example into a discrete subgroup of order 5 (the spatial symmetry of a starfish). Note that such an analysis can also be performed retroactively, so that the initial symmetry is actually understood as such when one witnesses its breaking.
- Another situation corresponds to the case where the symmetry changes are constituted on the basis

of already determined theoretical symmetries (which can be altered in the process). This situation can be analyzed as the formation of additional observables which are attached to or are partially the result of already existing ones. Then these symmetry changes are associated with already determined properties, but their specific form is nevertheless not predetermined. A typical example is the case of physically non-generic behaviours that can be found in the physical analysis of some biological situations, see [LV06]. From the point of view of the theoretical structure of determination, it is then a situation where there are predetermined attachment points, prone to lead the biological system to develop its further organization on them. The form of the biological response to these organizational opportunities of complexification is not, however, predetermined and therefore generates an original form of randomness. This theoretical account is close to the notion of next adjacent niche, proposed in [Kau02]; however, we emphasize that the theoretical determination of these next organizational possibilities is only partially defined. For example let's consider that a biological dynamic has approximately certain symmetries, which leads to a non-generic singular point; then it is possible (and maybe probable) that this point will be stabilized in evolution, in an unpredictable way (which specific forms will actually implement those symmetries?).

- The former case is constituted, in a sense, by a *specific* organizational opportunity. We can, however, consider cases where such opportunities are not theoretical predetermined. Now, the constitution of symmetry changes should be understood as even more random, and the associated predictability is extremely low. Gould's most quoted example of "exaptation', the formation of the bones of the internal hear from the double jaw of some vertebrates, some two hundred million years ago, can fit in this category.

We have seen that the symmetry changes are related to a strong form of randomness. This randomness and its iterative accumulation are, however, the very fabric of biological organization. Therefore, we have a theoretical situation where order (biological organization) is a direct consequence of randomness, as contingency, jointly to a (changing) structural stability — for example, the need for an internal coherence of organisms (but of a niche or an ecosystem, as well). The global analysis allowed us to give mathematical sense to Gould's complexification along evolution, as a consequence of the random paths of an asymmetric diffusion (sections 1 and 2). In ongoing work, a finer (or local) analysis may suggest a way to understand ontogenetic changes also in these terms, that is as a random dynamic of symmetry changes (see the preliminary, longer version of this paper for some hints [LM12]) . This situation should be not confused with the cases of order by fluctuations or statistical stabilization (for example, by the central limit theorem). In our case, indeed, order is not the result of a statistical regularization of random dynamics into a stable form, which would transform them into a deterministic frame. On the contrary, the random path of a cascade of symmetry changes yields the theoretical symmetries of the object (its specific phenotypes), which also contribute to its ontogenesis.

In this context, the irreversibility of these random processes is taken into account by entropy production. The latter, or more precisely a part of the latter, is then associated to the ability of biological objects to generate variability, thus diversity and adaptability. In ontogenesis, this point confirms our analysis of the contribution of anti-entropy to entropy production, in association with variability, including cellular differentiation. This situation is also consistent with our analysis of anti-entropy as a measure of symmetry changes, as hinted in the preliminary version of this paper. Notice that the symmetry changes, considered as relevant with respect to anti-entropy, may be taken into account, for example, in the coefficients corresponding to the different components of anti-entropy.

4 APPENDIX: Anti-entropy as a measure of symmetry changes

In [LM11], by comparison and contrast with physical theories, we proposed to understand biological phenomena as a conceptual frame where the theoretical symmetries are "constantly" broken. We will now show that such considerations allows us to interpret anti-entropy, somewhat in the spirit of Boltzmann's approach of physical entropy. In [BL09], premises of these aspects are considered from a strictly combinatorial point of view, leading to a "constructive" definition of the three components of anti-entropy that we recalled in section 2.

In order to show how symmetries come into play we will more closely analyze now these components. As anti-entropy, as biological complexity, is our main observable, it must be quantifiable.

Combinatorial complexity, K_c:

For a total number of cells N and for a number n_j of cells of cell type j, the combinatorial complexity is defined as:

$$K_c = \log\left(\frac{N!}{\prod_j n_j!}\right) \tag{10}$$

The classical combinatorial point of view is that the number of ways to classify N cells in j categories each one of size n_j. More precisely, we recognize, inside the logarithm, the cardinal, $N!$, of the symmetry group S_N, that is the group of transformations, called permutations, that exchange the labels of N elements. Similarly, $n_j!$ is the number of permutations among n_j units, which has the biological meaning of permutations of cells within a cell type. In other words, permuting cells *within the same cell type* is a combinatorial invariant of the complexity of an organism. Thus, the group of permutations leaving the cell types invariants is the group $G_{type} = \prod S_{nj}$, that is the group obtained as direct product of the symmetries corresponding to permutations within each cell type. Formally, this group corresponds to the change of labels in each cell type, which can all be performed independently and conserve the classification by cell types. The cardinal of this group is $\prod_j n_j!$.

Then, the number of cell type configurations is the number of orbits generated by the right action of G_{type} on S_N. In other words, a cell type configuration is first given by a permutation of $[\![1,N]\!]$, which gives the random determination for N cells. Moreover, these transformations must be computed modulo any transformation of G_{type} that gives the same configuration (as we said, cells within each cell type are combinatorially equivalent — we will discuss below this hypothesis, in more biological terms). Lagrange theorem then gives the number of remaining transformations $N!/\prod_j n_j!$, which is the number of possible configurations. Clearly, an organism with just one cell type (typically, a unicellular being) has combinatorial complexity $\log(1) = 0$. As a result, this measure of combinatorial complexity depends on the total number N of cells, but is actually *a measure of the symmetry breaking induced by the differentiation in cell types*.

Let's compare the situation with Boltzmann approach of entropy. If one has a number of microscopic phase space states Ω having the same energy, the corresponding entropy is defined as $S = k_b \log(\Omega)$. In the case of gases, one considers that the particles are indiscernible. This means that one does not count twice situations which differ only by permuting particles. In other words one formally understands the situation by saying that labels attached to particles are arbitrary. Thus, more soundly, S is defined by $S = k_b \log(\Omega) - k_b \log(N!) > 0$. This symmetry by permutation reduces the size of the microscopic possibility space, and, as a result, entropy.

In our approach, we have $K_c = \log(N!) - \sum_i \log(n_i!)$ which is greater than 0, as soon as there is more than one cell type. Thus, the increase of the possibility space (the diversity or the differentiations) increases the complexity. More precisely, the complexity, as absolute value of anti-entropy, is decreased by the remaining symmetries and quantified by the term $\sum_i \log(n_i!)$. We understand then that anti-entropy can be analyzed, at least in this case, as an account of how much biological symmetries are broken by the cascade of differentiations. Formally, we can sum the situation up by saying that the combinatorial complexity and its contribution to anti-entropy are based on a group of transformations, S_N, and a subgroup, G_{type}. The biologically relevant quantity is then the ratio of sizes of the groups S_N and G_{type}.

Morphological complexity, K_m:

This complexity is associated to the geometrical description of biologically relevant shapes. It is computed among others by counting the number of connex areas. Note that this number corresponds to *space symmetry breakings* for motions covering this space — or ergodic motions. Then, one has to consider the number of shape singularities, in the mathematical sense, where singularities are invariants by action of diffeomorphisms. The fractal-like structures are particularly relevant since they correspond to an exponential increase of the number of geometrical singularities with the range of scales involved. Thus, fractal-like structures lead to a linear growth of anti-entropy with the order of magnitudes where fractality is observed.

Functional complexity, K_f

(the last quantity proposed in [BL09]): This quantity is given by the number of possible graphs of interaction. As a result, the corresponding component of anti-entropy is given by the choice of one graph structure (with distinguished nodes) among the possible graphs. This involves the selection of the structure of possible graphs and, correspondingly, a hypothesis on which graphs are considered equivalent. In terms of symmetries, we first have a symmetry among the possible graphs which is reduced to a smaller symmetry, by the equivalence relation. For example, in [BL09], the case is considered where the number of edges is fixed, so the considered symmetry group is engendered by the transformations which combine the deletion of an edge and the creation of another one. The orbits preserve the total number of edges, so that the orbit of a graph with $\langle k \rangle N$ edges are the graphs with this number of edges.

We understand then that anti-entropy, or at least its proposed decomposition in [BL09], is strictly correlated to the amount of symmetry changes associated to a configuration. We will now look more closely at the case of combinatorial complexity since it involves only groups of permutations and their subgroups, but at the same time will also allow us to express crucial conceptual and mathematical points.

We indeed encounter a paradox in the case of combinatorial complexity. On one side, we have an assumption that cells of the same cell type are symmetric (interchangeable). On the other, in section 2, we stressed that each cell division consists in a symmetry change. This apparent paradox depends on the scale we use to analyze the problem, as well as on the "plasticity" of the cells in a tissue or organ, as the possibility to be interchanged and/or to modify their individual organization. Typically, one can assume that liver cells function statistically (what matters is their average contribution to the function of the organ), while neurons may have strong specific activities, yet they may also deeply modify their structure (change number, forms and functionality of synaptic connections, for example). Thus, we will next consider the individual contribution of cells to the combinatorial complexity of an organism at different scales.

If we consider an organism with a large number of cells, N, and the proportion q_j for cell type j we get two different quantities for the combinatorial complexities, K_{c1} and K_{c2}:

$$\frac{K_{c1}}{N} = \frac{\log(N!)}{N} \simeq \log(N) \quad \frac{K_{c2}}{N} = \frac{\log\left(\frac{N!}{\prod_j (q_j N)!}\right)}{N} \simeq \sum_j q_j \log(1/q_j) \tag{11}$$

We propose to understand the situation as follows. Basically, both levels of cellular individuation are valid; but they have to be arranged in the right order. Cellular differentiation is the first and main aspect of the ability of cells to individuate in a metazoan, so we can assume that the main determinant of combinatorial complexity is K_{c2}. It is only after this contribution that the further process of cellular individuation occurs. The latter leads to a mean contribution which is $\sum_j a_j^{(q_j \log(q_j N) - 1)}$ per cell, where a_j quantifies the ability of each cell type to change their organization. It seems reasonable to expect that the a_j are high in the cases, for example, of neurons or of cells of the immune system. On the contrary, the a_j should be especially low for red blood cells. The reason for this is not only their lack of DNA, but also their relatively simple and homogeneous cytoplasmic organization. Similarly, liver cells may have statistically irrelevant changes in their individual structure.

Thus, the contribution of cell types to anti-entropy derives first from the formation of new cell types, while considering the ability of cells to reproduce, with changes, within a cell type as a further important aspect of their individuation process. Note that this analysis does not assume that a cell type for a cell is irreversibly determined, but it means that the contribution of cell type changes to anti-entropy are understood as changes of K_{c2}.

We can then provide a refined version of S_c^-, where a_{ct} is the "weight" accorded to the formation of different cell types:

$$\frac{S_c^-}{-Nk_b} = a_{ct} \sum_j q_j \log(1/q_j) + \sum_j a_j \left(q_j \log(q_j N) - 1\right) \tag{12}$$

$$= \left(a_{ct} - \langle a_j \rangle\right) \langle \log(1/q_j) \rangle + \langle (\langle a_j \rangle - a_j) \log(1/q_j) \rangle + \langle a_j \rangle \log(N) \tag{13}$$

where $\langle x \rangle$ is the mean of x among all cells (so that the contribution of each cell type is proportional to its proportion in the organism). Both equations 12 and 13 are biologically meaningful. The terms in equation 12 correspond, by order of appearance, to the contribution of the categorization by cell types and to the contribution of individuation among a cell type. In equation 13, we have obtained terms that can be assimilated to K_{c1} (last term) and to K_{c2} (first term), the latter being positive only if $a_{ct} - \langle a_j \rangle > 0$, meaning that the contribution associated to cell types is positive only if it is greater than the mean cellular individuation. This is logical since cell types make a positive contribution to the complexity only if the amount of cellular diversity they introduce is greater than the one that cellular individuation alone would introduce.

Last but not least, the second term has the sign of an anti-correlation between a_j and $\log(1/q_j)$, meaning that this term is positive when there are many low complexity cell types[11] and few high complexity cell types. More precisely, using the Cauchy-Schwartz equality case, we get that maximizing (and minimizing) this term (everything else being kept constant), leads to $\langle a_j \rangle - a_j \propto \log(1/q_j) - \langle \log(1/q_j) \rangle$. Then this optimization *a priori* leads to maximizing the *variance* of information (in informational terms), at constant entropy (=mean information).

Here, the issue derived from looking with an increasing finer resolution at the individuation potential. However, the reciprocal situation can also occur. Let's consider the functional complexity, understood as the possibility of interactions between cells (the paradigmatic example is neurons). Then, by assuming that there are N neurons with $\langle k \rangle$ average number of synapses for each neuron (where $\langle k \rangle$ is between 103 and 104 for humans), as presented in [BL09], we get:

$$N_G = \binom{\binom{N}{2}}{\langle k \rangle N} \frac{K_{f1}}{N} \simeq \langle k \rangle \log(N) \tag{14}$$

However, if we postulate that *any* graph of interaction is possible, then we get a total number of possible interactions which corresponds to a choice between interaction or no interaction for each entry of the interaction matrix (N^2 cells). However, the latter is symmetric; and we do not count the self-interactions (because they correspond to the complexity of the cell) so we obtain $N(N-1)/2$ binary choices, so $2^{n(n-1)/2}$ possibilities: $K_{f2}/N \simeq N/2$.

There is two main lines of reasoning we can follow to understand the situation. The first is to look at the time structure of symmetry changes. Indeed, the symmetry changes occur as a temporal cascade. As a

11

In information theory, $\log(1/q_j)$ is the information associated to j: it quantifies its scarcity. If one assume that $a_j = \langle a_j \rangle \pm a$ and that we keep constant the mean complexity of cells, the anti-correlation is typically obtained when we have more low complexity cell types, with fewer cells, than high complexity cell types (which have therefore more cells). If one consider again the a_j as a degree of freedom, the same result can be achieved high complexity cell types with very high complexity and therefore a high number of bellow average complexity cell types.

result, the temporal hierarchy of individuation is crucial. Here, we can refer to some phenomena concerning the graph of interaction of neurons. A crude description of the formation of neural networks is the following. First, a large number of "disordered" connections take place. Only after, the functional organization really increases by the decay of unused synapses (see for example [LO05]. Then, the "bigger" symmetry group involved in the description is of the form K_{f1}, with $\langle k \rangle$ mean number of connections; but then this symmetry group is reduced to obtain a smaller symmetry group with $\langle l \rangle$ mean number of connections. This operation can be seen as a change of symmetry groups, from the transformations preserving the number of connections with $\langle k \rangle N$ connections to those preserving $\langle l \rangle N$ connections.

Of course there are many other possible components for a measure of biological complexity. This proposal, defined as anti-entropy, provides just a tentative backbone for transforming the informal notion of "biological complexity" into a mathematical observable, that is into a real valued function defined over a biological phenomenon. It should be clear that, once enriched well beyond the definition and the further details given in [BL09], this is a proper (and fundamental) biological observable. It radically differs from the rarely quantified, largely informal, always discrete (informally understood as a map from topologically trivial structures to integer numbers) notion of "information", still dominating in molecular circles, see [LMSS12] for a critique of this latter notion.

References

[BL08] F. Bailly and G. Longo. Extended critical situations: the physical singularity of life phenomena. Journal of Biological Systems, 16(2):309, 2008.

[BL09] F. Bailly and G. Longo. Biological organization and anti-entropy. Journal of Biological Systems, 17(1):63–96, 2009.

[BL11] F. Bailly and G. Longo. Mathematics and the natural sciences; The Physical Singularity of Life. Imperial College Press, London, 2011. Preliminary version in French: Hermann, Vision des sciences.

[CFX06] L. Cai, N. Friedman, and X.S. Xie. Stochastic protein expression in individual cells at the single molecule level. Nature, 440(7082):358 – 362, 2006.

[GKPC05] S. Gheorghiu, S. Kjelstrup, P. Pfeifer3, and M.-O. Coppens. Is the lung an optimal gas exchanger? In Gabriele A. Losa, Danilo Merlini, Theo F. Nonnenmacher, and Ewald R. Weibel, editors, Fractals in Biology and Medicine, Mathematics and Biosciences in Interaction, pages 31–42. Birkhäuser Basel, 2005.

[Gou89] S.J. Gould. Wonderful life. Norton, 1989.

[Gou97] S.J. Gould. Full house: The spread of excellence from Plato to Darwin. Three Rivers Pr, 1997.

[Hor77] K. Horsfield. Morphology of branching trees related to entropy. Respiration Physiology, 29(2):179, 1977.

[Kau02] S.A. Kauffman. Investigations. Oxford University Press, USA, 2002.

[LCI81] S.L. Lindstedt and W.A. Calder III. Body size, physiological time, and longevity of homeothermic animals. Quarterly Review of Biology, pages 1–16, 1981.

[Les08] A. Lesne. Robustness: Confronting lessons from physics and biology. Biol Rev Camb Philos Soc, 83(4):509 – 532, 2008.

[LM11] G. Longo and M. Montévil. From physics to biology by extending criticality and symmetry breakings. Progress in Biophysics and Molecular Biology, 106(2):340 – 347, 2011. Systems Biology and Cancer.

[LM12] G. Longo and M. Montévil. Randomness increases order in biological evolution. In M. Dinneen, B. Khoussainov, and A. Nies, editors, Computation, Physics and Beyond, volume 7160 of *Lecture Notes in Computer Science*, pages 289 – 308. Springer Berlin / Heidelberg, 2012. 10.1007/978-3-642-27654-5_22.

[LMD+08] A.B. Lindner, R. Madden, A. Demarez, E.J. Stewart, and F. Taddei. Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. Proceedings of the National Academy of Sciences, 105(8):3076 – 3081, 2008.

[LMK] G. Longo, M. Montévil, and S. Kauffman. No entailing laws, but enablement in the evolution of the biosphere. Arxiv 1201.2069. To be published in GECCO proceedings 2012.

[LMSS12] G. Longo, P.-A. Miquel, C. Sonnenschein, and A.M. Soto. From information to organization in biology. to appear, 2012.

[LO05] L. Luo and D.M. O'Leary. Axon retraction and degeneration in development and disease. Annual Review of Neuroscience, 28:127–156, 2005.

[LV06] A. Lesne and J.-M. Victor. Chromatin fiber functional organization: Some plausible models. Eur Phys J E Soft Matter, 19(3):279–290, 2006.

[MB10] D.W. McShea and R.N. Brandon. Biology's first law: the tendency for diversity and complexity to increase in evolutionary systems. University of Chicago Press, 2010.

[MFWS04] B. Mauroy, M. Filoche, E.R. Weibel, and B. Sapoval. An optimal bronchial tree may be dangerous. Nature, 427:633–636, 2004.

[NWG90] T.R. Nelson, B.J. West, and A.L. Goldberger. The fractal lung: Universal and species-related scaling patterns. Cellular and Molecular Life Sciences, 46:251 – 254, 1990.

[Ruu54] J.T. Ruud. Vertebrates without erythrocytes and blood pigment. Nature, 173(4410):848, 1954.

[Sch44] E. Schrödinger. What Is Life? Cambridge U.P., 1944.

[WBE99] G.B. West, J.H. Brown, and B.J. Enquist. The fourth dimension of life: Fractal geometry and allometric scaling of organisms. Science, 284(5420):1677–1679, 1999.