

The logico-causal *incompleteness* of Programming Theory in Molecular Biology

Giuseppe Longo

LIENS, CNRS – ENS, Paris

<http://www.di.ens.fr/users/longo>

Models in Biology, Paris, January 23, 2007

Genova, 20 aprile 2007

- **Part I:**

The logico-causal incompleteness of the programming “model” (metaphor?) in Molecular Biology: a critique from the point of view of Physics and Logic.

(With P.E. Tendero)

- **Part II (hints):**

A systemic approach: Extended Criticality

*(With **Francis Bailly** [1], Chap. 6)*

(downloadable: <http://www.di.ens.fr/users/longo>)

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

Current *causal* analysis in Molecular Biology (a common empirical tool)

- In Biology, a **differential analysis**:
(induced/observed) genomic variations (mutation) and their
phenotypic “consequences”
- *A comparative analysis with Physics*: the case of
Thermodynamics
- Differential methods in *Logic* or *Programming Theory*?

The Differential Method: General Principles from a physicalist perspective

A paradigmatic example

- *Perfect gas*: $pV = KT$ equilibrium law, first obtained by the differential, but later *derived* within a **systematic-theoretical determination** (**general construction principles**):

The Differential Method: General Principles from a physicalist perspective

A paradigmatic example

- *Perfect gas*: $pV = KT$ equilibrium law, first obtained by the differential, but later *derived* within a **systematic-theoretical determination** (**general construction principles**):
 7. Least polynomial (a mathematical principle)
 - Symmetries and geodetics (gas: kinetic analysis, *next...*):

$pV = KT$ as a *causal* model

It describes **general causal** dependencies *and*

it **may be uniquely derived** from the *principles* of **Statistical Physics** (jointly to the principles of Thermodynamics):

- an analysis of particles' kinetic energy (T) and shocks on walls (p)
- take the sum (integral - the *thermodynamic limit*) of individual trajectories, as **geodetics** (a paradigmatic reduction)

A **systematic-theoretical determination** which *justifies* the differential method, grounded on major **conservation principles**.

(That is: symmetries and least action principles).

In Molecular Biology:

Is the “programming paradigm” a
systematic-theoretical determination ?

Does it rely on

- a similar paradigm as in Physics or
- on a similar theoretical approach as in Computer Science (e.g. within the *semantics* of Programming Languages) ?

Claim:

“Per se”, a differential analysis (teratogenic) **does not provide a direct causal link (for the normal processes - the wild gene)** in **absence** of a *systematic-theoretical determination as in Physics* (**geodetics** and **symmetries** as underlying the causal structure)

***Very Relevant* differential information**

- **Pathologies** (no physical sense):

Robust evidence for the origin as *mutations* of some
(**rare!**) *genetic* diseases:

Tay-Sachs, Huntington, cystic fibrosis, thalassemy...

But... how to go from the mutation to the *wild gene*?

An example of *improper use*: **phenyl-cetonury**:

Mutation (two alleles mutated) \Rightarrow idiot child...

- *For decades: **there*** is the gene of intelligence, it encodes the structure of brain!

An example of *improper use*: **phenyl-cetonury**:

Mutation (two alleles mutated) \Rightarrow idiot child...

- *For decades: **there*** is the gene of intelligence, it encodes the structure of brain!
- *More recently:* the mutation produces an *inactive enzyme* which affects *brain metabolism* during early stages of life

([Dumaret, Stewart, 1989], [Stewart, 2004])

From mutation to the wild gene ? (*more*)

- *Another example-abuse:*

The *gene-program of marital fidelity*:

can it be derived from the differential analysis, as claimed in

Young et al., **Nature**, 400, 766-788, 1999 ?

(*within Theory of Programming...*)

From mutation to the wild gene ? (*more*)

- *Another example-abuse:*

The *gene-program of marital fidelity*:

can it be derived from the differential analysis, as claimed in

Young et al., **Nature**, 400, 766-788, 1999 ?

(*within Theory of Programming...*)

Predestination in Protestantism...

A short list of “genes-as-programs” in the literature

The gene of:

- Longevity
- Obesity (N.Y. Rockefeller Center, 2000: 20 million dollars?)
- Teaching skill
- Sociability
- Seduction
- Mathematics
- **Academic freedom (??)**

[Charles Auffray, Qu'est-ce qu'un gène?, 2004]

Entangled causality

[Goldschmidt, 1938]:

Consequences of the displacement of a **homeotic gene** in *drosophyla* : eyes → legs, wings → legs... (and viceversa)

Later observed:

the **pressure** or the local intensity of the **electromagnetic field**, as parameters of the ecosystem (“norms of reaction”) *may induce the same phenotypic consequences*

(in particular: *pressure on embryo* shifts wings ↔ legs)

[Stewart, 2004; Fox Keller, 2004]

What causes what, in absence of general theoretical principles?

The theoretical(-causal) frame of Programming

Physical Determination (Classical)

Laplace's view:

1) determination \Rightarrow predictability

and

2) determination \neq randomness

[Laplace, Philosophie des Probabilités, 1786]

[J. Monod, Le hasard et la nécessité, 1970]

Poincaré (1890):

classical randomness = unpredictable determinism (classical physics) by “sensitivity to border conditions”...

but determination $\not\Rightarrow$ predictability

(physical measure is an interval + non-linearity!)

The theoretical(-causal) frame of Programming

The originality of the machine

Turing Machine:

- a **Logic Computing Machine**: a man in the least act of computing-thinking (1936)
- a *Discrete State Machine*, since 1948
(a physical perspective: *what about determination and causality?*)

The laplacian nature of **Turing's Discret State Machine**

- [Turing, 1950; p. 47]:
“ In a **Discret State Machine** (DSM)... it is always **possible to predict** all future state ...” by the *discrete data types* and computational *environment*
“ This is reminiscent of **Laplace's** view ... ”
- [Turing, 1952]:
A **non-linear (continuous) system for Morphogenesis**: main property
“*exponential drift*” (i.e. sensitivity to initial conditions, a *theoretical property*), which yeilds unpredictibility.
- [Turing, 1950]:
"It is an essential property of ...[DSMs] that **this phenomenon does not occur**. Even when we consider the actual physical machines instead of the idealized machines,..., **prediction is possible**".
- Reason: **discrete** data types \Rightarrow perfect **iteration** (a translation symmetry)
AND exact measure

[Longo, 2007: “Laplace, Turing... the impossible geometry ...”]

Shroedinger, What is life? (1944)

- « In calling the structure of the chromosomes a *code-script*, we mean that the all-penetrating mind, **once conceived by Laplace**... could tell from their structure how the egg would develop... . » p. 22-23.
- Today, the code-script has been fully decoded...

Where is the program?

The changing notion of ‘gene’

- 1 - [Mendel, 1907]: *no attempted structural definition nor location; a **functional unit of recombination** (of assortment and segregation)*
- 2 - Morgan (1905) and Muller [Morgan, 1926]: *a **structural unit of mutation**; hypothetical location in **chromosomes** of the determination of phenotypic (observable) **characters**.*
- 3 - [Watson and Crick, 1953, Crick, 1957]: *characterization in terms of **molecular biology** (the chemical structure of DNA: stable, organised, “rigid”); loci of determination of **proteins**.*

And then: the “**central dogma**” :

*a **direct-causal and one-way** relationship from genes,
as **portions of DNA**, to proteins.*

Weakening the central dogma

Reduced to the

“one gene one enzyme hypothesis” [Perutz, 1987]

Later shown to be false [Brett et al. 2001; Bartel, 2004]:

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

Massive polymorphisms and pleiotropy:

- ***one gene – many proteins :***

- the same portion of DNA may engender different proteins;

- almost isomorphic portions of DNA may produce similar effects.

- ***one protein - no genes :***

the primary structure of proteins may depend on the transfer activity of RNA within cytoplasm eukaryotic cells (e. g. by “*alternative splicing*” %)

But... where are genes?

- “... the *transcripts* come from regions of DNA previously identified as holding protein-coding genes. *But many do not.*”
- ... the descriptions of proteins encoded in DNA *know no borders* — each sequence reaches into the next and beyond.”
- “The more scientists become in molecular genetics, the less easy it is to be sure about *what, if anything, a gene actually is.*”

[Helen Pearson, **Nature**, May 2006]

[Fox Keller, 2007]

[D. Brett; H. Pospisil; J. Valcárcel; J. Reich; P. Bork "[Alternative splicing and genome complexity](#)". *Nature Genetics* **30**, 2001].

In summary, up to now:

- No structural determination of what a gene is (*part of DNA? It includes RNA transfer activities?...*)

Reasons (from Physics) for the abuse:

- No way to deduce from a *differential analysis* a **direct causal** link, *without general theoretical principles*

Can these principles be derived from Logic or Computing?

Differential Methods in Logic or Programming Theory?

- **Theory of Programs:**

Given a program: understand its “semantics” (that is the *computed function*) by inducing “mutations”?

Differential Methods in Logic or Programming Theory?

- **Theory of Programs:**

Given a program: understand its “semantics” (that is the *computed function*) by inducing “mutations”? **No way!**

- *Analogy:* the “**Böhm-out**” technique

[Hyland, 1976; Barendregt, Longo, 1980; Barendregt, 1984].

Böhm’s theorem (1967):

“Two different terms of lambda-calculus in normal form cannot be consistently equated”

Once given the operational semantics!! (the “compiler” and... Church-Rosser) or certain mathematical (denotational) domains of interpretation (a *direct* “causal/deductive structure”).

More on **Differential Methods in Logic and Computing**

- Formally: given M and N *terms* in normal form (programs),

$$M \neq N \Rightarrow [M]_E \neq [N]_E$$

in any already given semantic environment $[-]_E$,

i.e. in any mathematical meaning of programs.

- This result, which may be considered of the “differential” type, *does not* provide the semantics of a program (or of a lambda-term), its “phenotype”.
- It may instead help to characterizes syntactically the terms having *the same* (or different) *semantics*, within an *already given* mathematical structure.

[Hyland, 1976; Barendregt, Longo, 1980]

New paradigms in Programming ?

(*concurrency in networks*)

- **Message-passing**, *doesn't model protein cascades*, as it doesn't support *re-combination* and *decay* (moreover, no distinction seems possible between a message and a process, in genomics)
- **Data-flow** doesn't support loops (proteins'-enzymes cycles)
- **Synchronous and asynchronous automata** badly handle the omnipresent *feedbacks* in the cell
- Proteins processes are *inhibited* and this is not described by term-rewriting or **Petri Nets**
- **In operator algebras** the *gates' output is predetermined*, it cannot be a function of the input, as it seems to be in molecular dynamics.

(this is only in reference to *biochemical cascades*;

yet, the “**compiler**” is the the *global coherence structure* of the cell - and of the organism, their *organisation* and *stability*...)

The Differential Method from a physicalist perspective

*summary**

The causal **incompleteness** of the Programming Paradigm
(*model/metaphor*) as **physical incompatibility** with respect to

- the context (dynamical, thermodynamical... *to say the least*)
- the differential method:
 - in Physics: **finite number of experiences** *interpolated* by geodetics in an *intended continua*:
 - **isolate** one *parameter* (or develop a close analysis of correlations)
 - the *conceptual frame* (optimality criteria, e.g. simplest polynomial)
 - the *mathematical frame* (geodetics) are **continuous** - at least in variational contexts.
 - large numbers treated as **integrals** (sums: e.g. perfect gas)

Finalism and the Program

- The need for *finalism* in biologists:
 - a program is *made for*, it is the result of an explicit *aim* and an *intelligent design*;
 - it **follows the law** : Galileo, Descartes, Newton... concept of *normative* physical law - written by God: we are God to the machine *and* the machine is the result of an **intelligent design**.

About terminology:

Consider a slope, a cavity, a (water) turbulence....

A further cavity engendered by the first....

A program? The first cavity contains the information for the second?? *Modern Physics* goes otherwise.

Conclusion of Part I

Molecular Biology:

- An **extraordinary experimental richness**
(remarkable empirical correlations between the *modifications* of genotype and of phenotype)
- No *causal theory* (genotype \Rightarrow phenotype)
- Highly unsatisfactory use of notions from Programming.

Hints to Mathematics in Biology

Remarkable mathematical work at

different levels of organization:

- **Morphogenesis, phyllotaxis** ... since D'Arcy Thompson, Waddington, Turing, Thom... : mostly organs (loci of *energy exchange* - e. g. geodetics : fractal structures...)
- **Networks** (Von Neumann, Hopfield... G. Parisi...): neural, cellular... (intercellular exchange - *gradient of energy*)
- Different mathematical methods for *different levels of organisation* (no dialogue...)

Issue: *the systemic unity* of an organism (and its structural stability)

Can they only depend on the **organisation** and **stability** of DNA?

One of the many issues:

In a cell (an organism) almost everything is *causally correlated* to almost everything (how to de-correlate variables?),

The structural stability (coherence structure) **of the cell** (and/or the organism), (*causally*) *contributes* the **organisation** and **stability** of proteins' cascades from DNA.

Our perspective: this coherence structure is a “**physically critical state**”, yet not pointwise, but *extended... not treated by current physical **Theories**.*

Some Philosophy

A **theoretical/epistemic** separation (**not ontological**)

Example: Quantum Mechanics:

- Non-separability
- Non-locality (entanglement)
- No trajectories in space-time

Bohr, 1932 (*Light and Life*, revised, 1962):

« ... no way to understand the **stability of the atom** in terms of classical or relativistic physics... similarly, there is no way to understand the **stability of the cell** in terms of current physical theories »

Within Physics: unification not reduction (of theories) !

(sole case:

Thermodynamics → Statistical Physics as *well constructed theories*)

Logic in Biology: challenges

- Part of Mathematics as a **science of invariants** and *invariance preserving transformations*.
- Computer Science: *also* a science of **iteration** (*from primitive recursion to portability of software*); different symmetries and symmetry breaking.
- Biology: a science of **structural stability** *and variability* (the main invariant?).

The physical singularity of life

Extended Critical Situations

F. Bailly

CNRS, Meudon

G. Longo

CNRS – ENS, Paris

<http://www.di.ens.fr/users/longo>

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

(chapter 6: downloadable english translation)

%

Some references

<http://www.di.ens.fr/users/longo>

- Bailly F., Longo G. **Mathématiques et sciences de la nature. La singularité physique du vivant.** *Hermann, Paris, 2006.*
- Longo G., Tendero P.-E. . *The causal incompleteness of Programming Theory in Molecular Biology.* Downloadable (A preliminary and longer french version will appear aux actes du colloque "**Logique, informatique et biologie**", Nice, à paraître chez DeBoeck, Paris, 2008.)
- Longo G.. *Laplace, Turing and the "imitation game" impossible geometry: randomness, determinism and programs in Turing's test.* In Epstein, R., Roberts, G., & Beber, G. (Eds.). **The Turing Test Sourcebook**, 2007.
- Bailly F., Longo G. *Schèmes géométriques pour le temps biologique.* Draft.

Project: life phenomena as *singularities* of physics
or as *extended critical situation*

A theoretical/epistemic separation (**not ontological**)

A methodological paradigm: Quantum Mechanics:

6. Non-separability
7. Non-locality (entanglement)
8. No trajectories in space-time

Physical Singularity of Life Phenomena

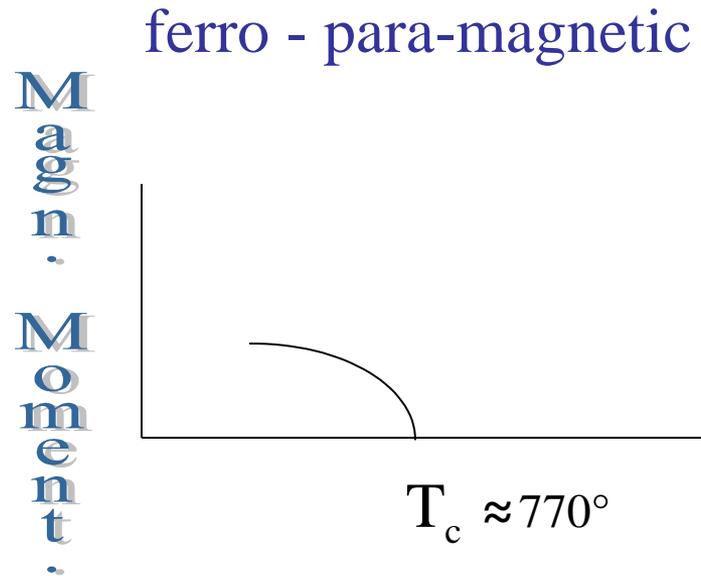
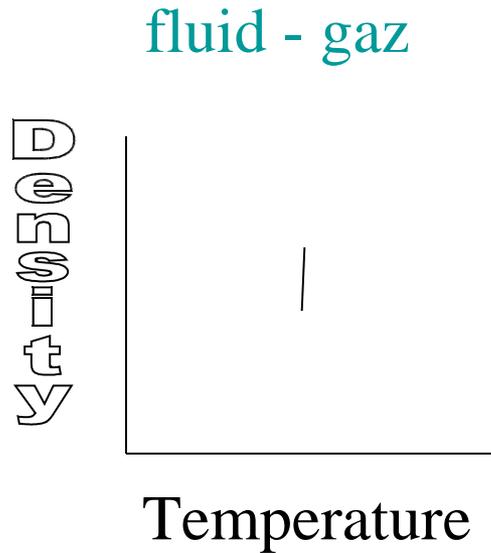
- Physical variation vs. **Variability** (individuation)
- Physics: *generic* (invariant) objects and *specific* trajectoires (geodetics)
Biology: *generic trajectories* (compatible) and *specific objects*
- Physics: deterministic unpredictability or quantum indetermination within a given phase space (entangled probabilities)
Biology: **Intrinsic indetermination** due to change of the phase space - phylogenesis
- Physics: **energy** as operator, **time** as parameter;
Biology: *time as operator* (Prigogine), *energy as parameter*
- **Biology**: **Extended critical situation** (from physical criticality)

Some concepts from Physics of Criticality

- *Example: crystal and a gas*: opposite of the scale, as for order
- *Phase transition*:
 - along a state change (e.g. gas \Rightarrow liquid \Rightarrow crystal...):
 - passage through a *critical state*
 - (sudden) change of the correlation length
- *Correlation length*:
 - a coherence structure (in some cases: distance of (possible) causal relations):
 - in a perfect gas, particles interactions
 - order of magnitude of the crystalline structure
- *State change*:
 - the *global* structure (may be) involved in the activities of the single particle
 - the *local* situation depends on (correlates to) the global.

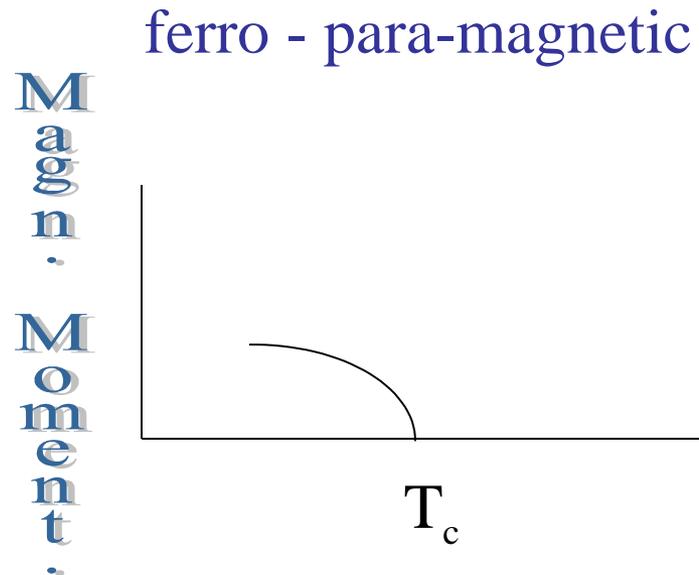
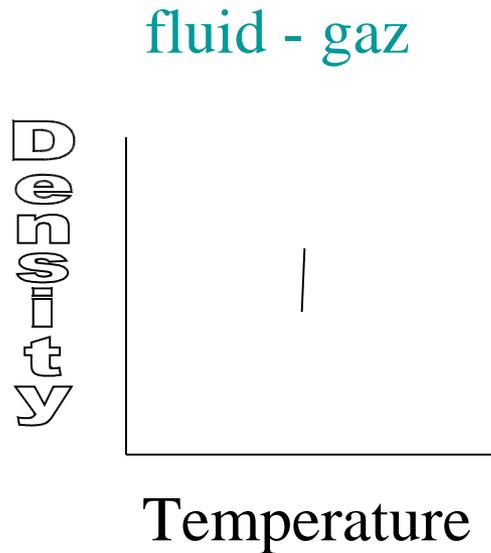
First and second order singularities

- Singularities *in* Physics: critical states: phase transitions...
Examples:



First and second order singularities

- Singularities *in* Physics: **critical states; phase transitions...**
Examples:

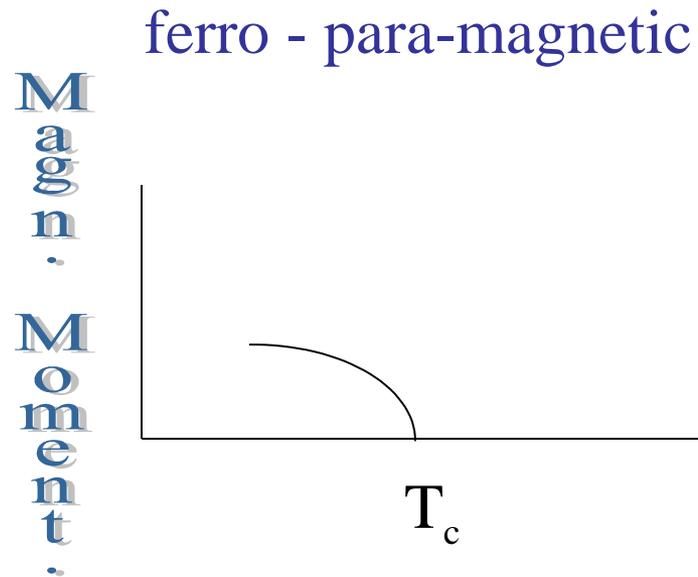
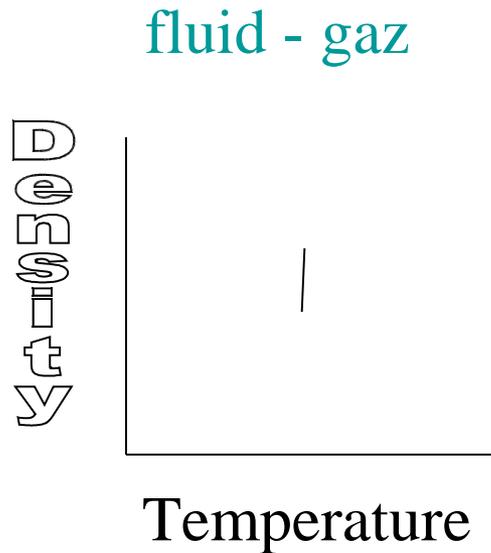


Critical phenomena (read backwards: ←)

- a passage from *statistical* to *geometric* properties, e.g. order (and viceversa: →)
- **divergence** of correlation length; from **local** to **global**:

First and second order singularities

- Singularities *in* Physics: **critical states; phase transitions...**
Examples:

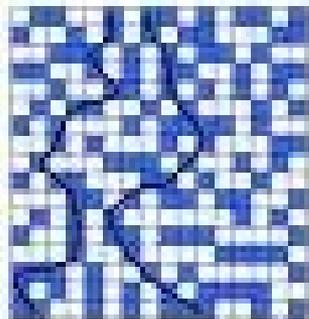
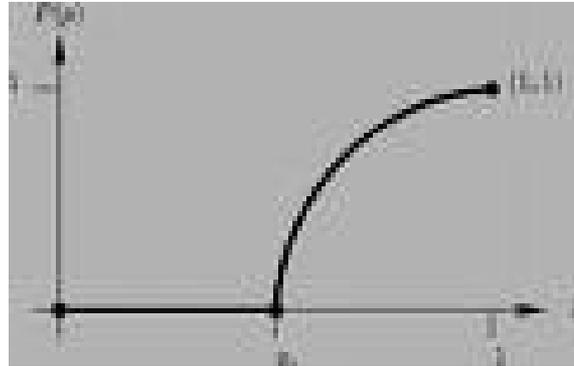


$$\text{Correlation length: } L = (T/T_c - 1)^{-\nu}$$

Local to global: L goes to infinity at T_c (one point!)

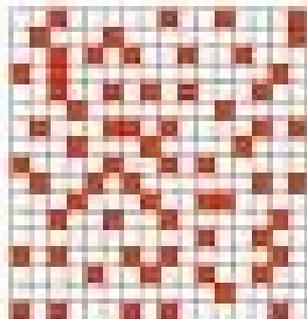
Entropy decreases (when lowering “←” the temperature)

Percolation:



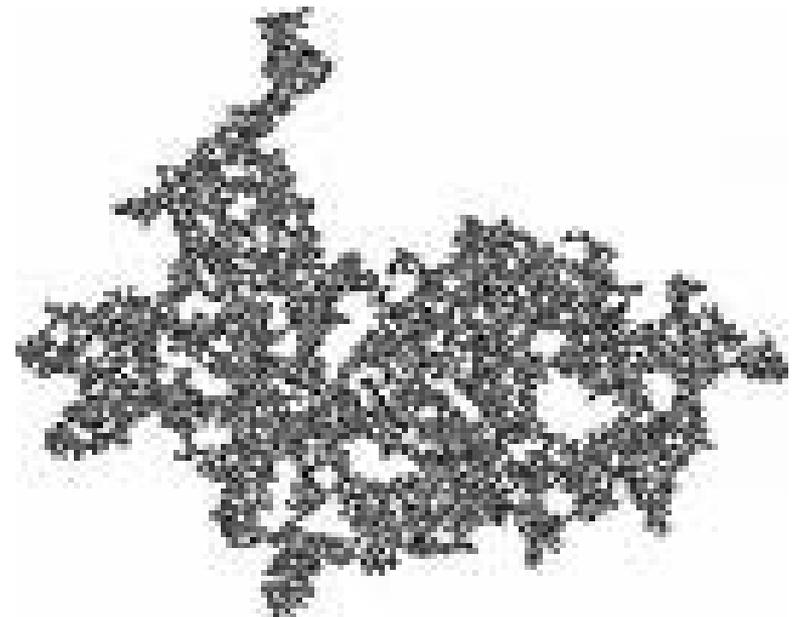
40% (P)

A connected path exists across the grid.

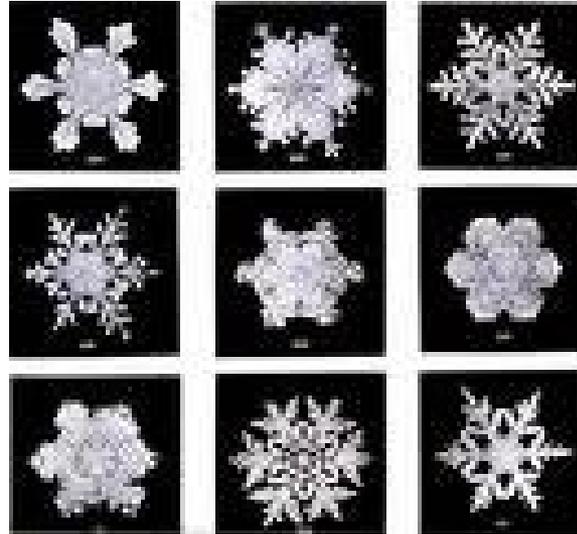


20% (P)

No connection exists across the grid.



Snowflakes



From thermodynamic potential (possible molecular links),
to a sudden and local, decrease of entropy
(in view of a slow-down of the brownian motion)

- Nicolis G., Prigogine I., **Self-organization in non-equilibrium systems: from dissipative structures to order through fluctuations**, Wiley, New York, 1977.
(cascades of bifurcation, as transitions towards chaos)
- Bak P., C. Tang C. and Wiesenfeld K. “Self-organized criticality”. **Physical Review A** 38: 364--374, 1988.
(the sand pile: a critical angle)
- Kauffman S.A., **The origins of Order**, Oxford U. P., 1993. (*“order for free”; metabolic nets*)

A critical state, in *Physics*

- An **isolated** point: a **singularity** in a process (or a bifurcation or a Thom's catastrophe, if irreversible).

- minor fluctuations**, possibly below the level of observability, may lead towards radically different evolutions.

- different* from "being far from equilibrium" (which does not imply *possibly different evolutions* of the system (bifurcations))

The *Mathematics* of criticality:

- **infinitesimal** variations yield **finite** changes (or finite variations lead to infinite changes, mathematical)
- instantaneous process as **divergence** of some functional descriptions (intended parameters);
- **maximality of complexity**, which also yields instability.

Life phenomena as *singularities* of physics or as *extended critical zone*

Physics (critical states, phase transitions...), *summary*:

- a **coherence structure** (order) may appear;
- **local to global** correlation length
- **isolated points** (0 measure) of the parameter(s);
- some **diverging** observables.

Living matter as extended critical zone:

- far from **equilibrium**
- **local to global**: internal correlation length *extended over time*:
“same” order of the system itself (critical phenomena)
- **fluctuations** within *extended* limits (several control parameters)
- **infinite physical complexity** (in view of divergences)
- Maintained into criticality by **autopoiesis** and **homeorhesis**
- Nesting and coupling of **levels of organization**.

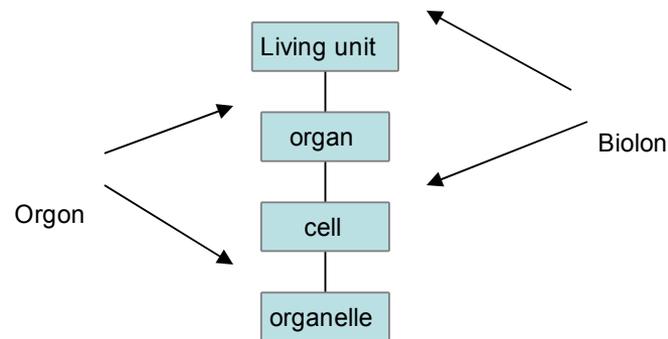
A possible meaning of a dynamics in an “extended critical situation”

- To be a **dissipative** process, **far from equilibrium**, as maintaining the internal organization requires energy exchanges, an **abnormally low entropy** w.r. to the equilibrium

(global flux entropy, in all dimensions = the sum of all Lyapounov exponents
(Ruelle) \neq Shannon-Kolmogorov-Sinai entropy = information flux, only positive exponents
(Lesne))

2. To form a **unity in space and time** (at least locally), where
 - **correlation lengths** at the various level of organization, at least one of the size of the system itself (the new object pertinent to the critical state)
 - **characteristic times**: within a stability (attracting) zone, exhaustive paths correspond to existence time.
 - **two dimensions of time**: a thermodynamical arrow of time, a compactified internal time (endogenous rhythms)

- To have an **internal organization**, given by the **interaction of the various levels** of organization, corresponding to dissipation constraints (dissipative structures) and to **organization loops** (e.g., autopoietic and sensory-motor loops). Alternating levels of organisation: organelles, cell, organ, living unit



(orgon: typical *fractal structures*)

4. A **volume**, not just a space of parameters (like in Physics), within a semi-permeable **membrane** (regulating energetic exchanges).

Integration, regulation and causal regimes

- Within *the borders of the extended critical state*, **phase transitions, changes in correlation length, passages through singularities ... occur continually** (e.g. pathologies), but, within within the borders of criticality, they are dealt with by the regulatory activities.

- **Integration** is the (upward-causal) presence of the local into the global structure:

Integration may be understood as the *correlations of variables* that give unity to a given system of equations; or, also, as the *organizing role of singularities* in their solutions.

- **Regulation** is the global structure that *causally* affects (downwards) local ones:

Regulation may play the role of the *initial or border conditions* on global behaviors of the solutions of equation systems (differential or at finite differences).

Causality and complexity

Physics:

Causal relations are local; global only in the sense of a field (by propagation of *local* interactions; i. e. by transitivity) or of the global determination (by equations).

Biology:

local causality may *differ* radically from **global** correlations, yet it *cannot be isolated* from the latter: integration and regulation, typically, *causally affect* local interactions (e.g. local bio-chemical exchanges may be regulated by cascades of hormones or neural signals of an entirely different nature).

Time: the product a compactified dimension (the loops) and a physical arrow.

More on the intelligibility of life phenomena

On complexity:

Elementary \Rightarrow **Simple** (the cartesian myth)

vs.

Elementary & Complex (today's challenge)

(strings (Physics); cells; cognitive units)

Our proposal: Life phenomena as **an extended critical states:**

- continual phase transitions (over an interval)
- global unity of new entity (causal entanglement)
- infinite (physical) measure of complexity

Biological units live in a state of **maximal** (infinite), physically unstable (even: *physically unsuitable*), **state of complexity:**

an extended critical situation

(a dynamical interference of global and local causality)

The **physical paradigms** helped us to formulate this notion, which is **not** internal to physical *theories*.

Summary: Physical Singularity of **Life** Phenomena

- **Biology: Extended critical situations** (from *physical criticality*)
- Physical variation vs. **Variability** (individuation)
- Physics: *Specific* trajectoires (geodetics) and *generic* objects
Biology: generic trajectories (compatible) and **specific objects**
- Physics: *deterministic unpredictability* or *indetermination* within a given phase space
Biology: Intrinsic indetermination due to change of the phase space - phylogenesis (ontogenesis?)
- Physics: **energy** as operator, **time** as parameter;
Biology: time as operator, **energy** as parameter

Some references

<http://www.di.ens.fr/users/longo>

- Bailly F., Longo G. **Mathématiques et sciences de la nature. La singularité physique du vivant.** *Hermann, Paris, 2006.*
- Longo G., Tendero P.-E. . *The causal incompleteness of Programming Theory in Molecular Biology.* Downloadable (A preliminary and longer french version will appear aux actes du colloque "**Logique, informatique et biologie**", Nice, 2005, paraître chez Vrin, Paris, 2007.)
- Longo G.. *Laplace, Turing and the "imitation game" impossible geometry: randomness, determinism and programs in Turing's test.* In Epstein, R., Roberts, G., & Beber, G. (Eds.). **The Turing Test Sourcebook**, 2007.
- Bailly F., Longo G. *Schèmes géométriques pour le temps biologique.* Draft.