

Information of the chassis and information of the program in synthetic cells

Antoine Danchin

Received: 10 May 2009 / Revised: 8 July 2009 / Accepted: 27 July 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract Synthetic biology aims at reconstructing life to put to the test the limits of our understanding. It is based on premises similar to those which permitted invention of computers, where a machine, which reproduces over time, runs a program, which replicates. The underlying heuristics explored here is that an authentic category of reality, information, must be coupled with the standard categories, matter, energy, space and time to account for what life is. The use of this still elusive category permits us to interact with reality via construction of self-consistent models producing predictions which can be instantiated into experiments. While the present theory of information has much to say about the program, with the creative properties of recursivity at its heart, we almost entirely lack a theory of the information supporting the machine. We suggest that the program of life codes for processes meant to trap information which comes from the context provided by the environment of the machine.

Keywords Algorithmic complexity · Logical depth · Physical entropy · Self-organisation · Field · Context

Abbreviations

3D Three-dimensional
OS Operating system

Reconstructing life is central to synthetic biology's efforts, as a means to try and understand what life is. I explore the consequences of the model of the cell-as-a-computer, where the "chassis" is explicitly separated from the program, as in a computer (Danchin 2009a). As a heuristics, information is viewed here as an authentic category of reality. I organise in what follows a tentative philosophical reflection on constraints met by synthetic biology around four themes which I see as a true revolution of human thinking, the shift from a mechanistic view of the world to an algorithmic view, with the result that living organisms can be understood as information traps.

The modern reflection on information began with Hilbert's problems at the beginning of the XXth century. One of his questions was whether arithmetics, the mathematics of whole numbers, was simply a tautology, i.e. its conclusions could be automatically drawn and reached from its premises. The response brought about by Gödel and successors after 1931 recognised that arithmetic was incomplete, in that it could bring about conclusions which were understandable only when taking a point of view from the outside of arithmetics. Arithmetics incompleteness establishes that whole numbers theory must be separated from its *meaning* for the human creator and observer. Briefly, arithmetics is associated to two levels of information: the self-sufficient information carried by strings of symbols, and the information carried by the context: language and civilisation, or more generally, by the biological entities we name *Homo sapiens*. The latter provides interpretations of the demonstrations and theorems created by the axioms and definitions of number theory but the corresponding information has not yet been theorised.

Based as is arithmetics on strings of symbols, the alphabetic metaphor of the genetic program sits at the centre of several theses which I try to make explicit, via a

A. Danchin (✉)
CNRS URA2171/Institut Pasteur, 28 rue du Docteur Roux,
Paris Cedex 15 75724, France
e-mail: antoine.danchin@normalesup.org
URL: www.normalesup.org/~adanchin/index_en.html

rapid travel through history with the aim to place the category information within synthetic biology: 1. A thousand-year-old metaphysical/ontological thesis, where I discuss the relationships between shape, form and the process of in-formation; 2. An epistemological thesis exploring how information links models of phenomena to reality in a situation identifying two levels of information, the information of the model, and the meaning of the model; 3. The exploration of extant theories of information as a prerequisite to understand the concept of genetic program in synthetic biology; 4. A conjecture proposing the need to create a new theory, that of information of the chassis (machine) or, why the brain is not a computer.

Metaphysical thesis: shape, form and information

When searching for life outside Earth we look for “unusual” *shapes*, not commonly associated with standard chemistry and mineralogy. We restrict our identification of forms, looking first for the 3D architecture of the “chassis” which compartmentalises the living entity (see for example the beginning of Monod’s *Chance and Necessity* (Monod 1971)). Typically we draw aside crystalline shapes, and look for more complex shapes such as those of spheroidal or tubular objects. Yet, we need more to recognise life, as drawing our conclusions from the geometry of shapes can be misleading (this happens when an artefact is interpreted as a biological entity’s signature (Clemmer and Beebe 1991)). Furthermore, the form of living organisms does not reduce to their static shapes, it implies dynamic processes.

From the early times of philosophy, life was identified as a phenomenon connected to recognisable autonomous but not independent categories. To account for all phenomena, Aristotle recognised ten categories: οὐσία, ποσότης, ποιότης, πρὸς τι, κείσθαι, ἔξις, τόπος, χρόνος, πράττειν, παθεῖν. An essential step in understanding reality required construction of some entanglement of these categories, a process which progressively reduced them to four: matter, space, time, and subsequently energy. A remarkable achievement was reached when Einstein combined them together in a surprisingly concise equation, $E = mc^2$. Yet, it was obvious that these universal categories do not account for many phenomena: no one has been able, for example, to derive the crystal lattice of a mineral as simple as sodium chloride from the equations of microscopic physics (Grandy 1992). Imagine the challenge for synthetic biology!

Even understanding what became the modern category *matter* was never simple: matter displayed itself as an immense variety of entities (shapes, processes, phase transitions and even transmutations...). One could think about *substance* (οὐσία, ὕλη and the like, with all kinds of

idiosyncrasies) but something more was required. Analysing movement of matter, the Atomists invented a process permitting matter to take form, to be *in-formed*, via the *interaction* of indivisible Parmenidian tiny material particles, the atoms. This process had the excellent property to account for an infinity of forms, but it asked for a process *causing* interaction. The core of the Atomists’ thought was that necessity (ἀνάγκη), associated to some consistency, λόγος, *not* chance (which is neither a Greek nor a scientific concept (Grandy 1992)), was the ultimate cause (Danchin 1986). In short, the problem of *creation of form, in-formation*, superimposed on the general problem of movement. In addition to phenomenology, understanding in-formation required a process of *synthesis*. In this context, synthesis became a direct exploration of the concept of creation, uncovering a link between in-formation and creation. This was understood soon after chemistry was born—chemical synthesis is at the heart of modern chemistry—and it is therefore not unexpected that biology, where form is apparent everywhere, should develop into synthetic biology.

Already asked by the presocratic philosophers the question of the nature and origin of form was renewed by Aristotle. After him, the question kept developing with Greeks in Alexandria and southern Italy, Arab and scholastic philosophy from the fall of the roman empire till the fourteenth century. Between Aristotle and the present time, I retain John Scotus Eriugena because of the way he tackled the problem of creation (Erigène trans. 1995). A neo-platonist, Eriugena divided Nature into four species: (1) Nature which creates and is not created; (2) Nature which creates and is created; (3) Nature which does not create and is created; (4) Nature which neither creates nor is created. To make a long argument short, asking questions this way leads to five modes of opposition, which introduce a hierarchy in natural entities, and in particular in living beings. As in the platonistic tradition, the material world of our experience is composed of ideas clothed in matter. However, Eriugena attempted to reconcile Plato with Aristotle, discussing Aristotle’s ten categories. Time and space were discussed as central for human perception of phenomena, matter is without form or limit, but it needs an external agent to take form, it needs to be in-formed. Interestingly, God, as defined by scriptures, escapes all categories except one, *relatio* (πρὸς τι, *ad aliquid*), which I retain here as it lies behind what we now name *information*. This elusive category remains central today (relationships appears typically in the god-like *self-organisation* (Grandy 1992)).

The second name I keep in this sequence is Averroës. His commentaries of the metaphysics of Aristotle had immediate and lasting success. I retain sentences of his *Tahafut al Tahafut*: «Matter only becomes in so far as it is combined with form. Everything that comes into being

comes into being from something else, and this must either give rise to an infinite regress and lead directly to infinite matter which is impossible, even if we assume an eternal mover, for there is no actual infinite; or the forms must be interchangeable in the ingenerable and incorruptible substratum, eternally and in rotation.» (Averroës trans. 1497). That substratum, substance, was difficult to make explicit. First split into the four elements, fire, air, water and earth, and, subsequently seen as atoms (which, despite their name, recently split into further particles), it needed association with something which makes the root of variety in the world, form (εἶδος).

How could form combine with matter? A variety of animas (souls and spirits) were invented to account for the birth, development and conservation of movement, until *energy* came in. This category permitted some entanglement of matter with space and time, and long took the role of the animating principle needed to account for life (see mesmerism and its “animal magnetism” and “positive energy” in small talk or the vocabulary of sects today). Many further categories required to account for life were discussed for centuries in the western world, most often based on the assumption that reality had to fit with explicit revelation by God of its characters, as written in the Scriptures. Thomas Aquinas used Averroës’ Grand Commentary of Aristotle’s *Metaphysics* as his model. In his *Summa theologica* he analysed the questions of Trinity and of creation, showing that standard reasoning tells us that creation is related with in-formation, placing *relationships* between all kinds of entities (including abstract entities) at a central position.

These analyses may be condensed in the question asked by the Pythia in Delphi: “I have a boat made of planks of wood. The planks are progressively replaced as they rot away. After some time, all have been replaced, none of the original ones remain: is it the same boat?” (Danchin 2003). To understand what life is, we need to understand the relationships between entities recognised as belonging to life, whether they are material, processes, or abstract, such as language. And we need to try and understand the process of *in-formation* (which we would in the modern terms name *creation of information*, noting that in the evolution of languages *redundancy* is a ubiquitous trend (Livingstone 2003)).

Information links models with reality

Belonging to reality we cannot behave as outsiders contemplating the world. Understanding information asks us to investigate the way science is constructed. Presocratic Greek philosophers recognised that our limitations in understanding truth (ἀλήθεια) only allowed us to give our

views (δόξα) on reality. “And for a certain truth, no man has seen it nor will there ever be a man who knows about the gods and about all the things I mention. For if he succeeds in the end in saying what is completely true, he himself is nevertheless unaware of it; and opinion is fixed by fate upon all things” said Xenophanes (Diels and Kranz 1935). Approaching truth is to place a fragment of reality in a particular perspective, where we can understand its relationships with human beings. The central tenet of science—often ignored, as many scientists behave as priests of a revealed religion when interacting with mass media—is that we construct models distinct from reality. We match models with phenomena, expressing local instances of reality in a particular context. Models may display a certain degree of adequacy, if not truth, with reality. The model/reality separation is so significant that several concomitant models may express our knowledge about a particular side of reality.

The importance of models to understand reality triggered the creation of axiomatics, mathematics and logics. This effort was well fitted to the Renaissance trend to replace Aristotle by Plato, removing the thought of the former to the “dark centuries” of Middle Ages. At the heart of platonistic philosophy, the shadows of mathematical archetypes had to be discovered by persons who were illuminated by their truth. This attitude placed mathematics in the world of idealities, suggesting that mathematical certainties existed separately. The medieval reflection on in-formation was soon replaced by a geometrical view of combinatorial creation of forms associated to the general structure of space, initially studied in one, two or three dimensions and later generalised to all kinds of dimensions. In parallel, and following a medieval trend of arabic mathematics, arithmetics and number theory slowly emerged as algebraic equations. Models recognised as of the highest quality were mathematical models, developing on their own, independently of reality with their in-built consistency (information). Trying to match models with reality allowed scientists to progress by producing better and better adequation with reality (Danchin 1992; Putnam 1988). However, the match between models and reality could never be direct (a mathematical model of an aeroplane does not fly). It rested on *interpretations* (processes rooted in culture and language, thus associated to a property that we might name *context* and linked to a research programme (Lakatos 1976, 1980)).

If constructing models while confronting them to reality defines science (Popper 1959), then the effort to establish an explicit demarcation between science and non-science is dominated by a particular category of reality, information again, using the word with all its fuzzy connotations (Popper 1963). Defining what science is emphasises two types of information, *information of the (mathematical)*

model and information of the context. Both types place the old category, *relatio*, at their heart, but only the former has yet been theorised, in the chaining of axioms and definitions, demonstrations and theorems (Danchin 2003).

The way synthetic biology is developing illuminates these points. Starting from preconceived biological views, it *abstracts* specific features into axioms and definitions, and builds up models, whether mathematical or experimental (e.g. engineering models) (Endy 2005). The models unfold with their own rules of consistency: a *demonstration* in mathematics, yielding a *theorem*, a computer output in a *simulation*, a genetically modified cell in an experiment... Subsequently one goes back to reality by proposing a concrete *instantiation* of the output, predicting a particular phenomenon. This prediction is of two major types: Either the prediction of a novel, previously unknown or unrecognised entity (a structure, a process, a metabolite...), or that of a particular behaviour of reality, which should manifest itself along lines predicted by the model (Fig. 1). A model is (temporarily!) valid when all its predictions are recognised in actualisations of reality. Typically, in synthetic biology bacteria have been constructed which display, as expected, some type of multistable behaviour or oscillations (Elowitz and Leibler 2000) or phages with artificial regulatory regions have been shown to display the ability to grow on cells (Chan et al. 2005). In neurosciences the basis of neuromimetic networks rests on a vast number of works where selective processes play a central role (Changeux et al. 1973; Edelman 1987).

However, because the model is not reality, this ideal outcome never develops for a long time. Even when we produce a new entity not recognised before the model's

construction—a great success, comes a time when a phenomenon does not fit the model's predictions. To proceed with our example: in bacteria, bistability is not stable in time (Veening et al. 2008). Initial attempts to solve the contradictions between model predictions and observed phenomena do *not* immediately discard the model. The common practice we witness in synthetic biology is *re-interpretation* of the instantiation process that matched the model to reality. Typically: “exceptions make the rule”, or “this is not exactly what we meant, we need to focus more on this or that feature”... This polishing step permits the context of the model and its associated phenomena to be defined as accurately as possible. It marks the moment when technically arid efforts such as normalisation, defining a proper nomenclature, a database data schema have a central role. We witness this today in synthetic biology in the *standardisation* effort of the community (Endy 2005). Despite all efforts to reconcile predictions and phenomena, the inadequacy between the model and reality becomes insoluble. This contradiction implies that we need to reconsider the axioms and definitions upon which the model has been constructed, triggering a spiral of further models, making science as we know it. As always with exploration, this exploratory attitude meets resistance: most of our contemporaries would be happy to be *believers*, and forget about the impossible but necessary quest of truth (Danchin 1992). This may explain both the hype and the reluctance to accept synthetic biology.

In the subsequent inflation of models there is a hierarchy. A mathematical demonstration is perceived as the ultimate proof (Popper 1963). This justifies the huge number of mathematical models published in systems and synthetic biology. Do they result in non-trivial predictions? I am afraid that, more often than not, most models are “retrodictions”, finding what is already well known (how often do metabolic models “discover” the Krebs cycle?), rather than predictions. Indeed, assessing the interpretation of postulates which have not been expressed in a precise way has deep consequences, including in mathematics, which illustrates the importance of the category *information*, connecting it with the standard categories of reality (time in particular). Deep features of axiomatics were understood when we discovered that something taken for granted was overlooked. Zermelo's axiom of choice (given any two sets, one set is in one-to-one correspondence with some subset of the other: this looks trivial, but is not) is a famous example of this situation. Similarly, and in line with the Pythagorean/Platonistic tradition, we accept *synchro*ny in the way we use mathematics, making it independent of time: when reasoning by recurrence, if we show that something is true for $n+1$, knowing it is true for n , this will be valid, whatever the size of n . We take for granted very large numbers even though it will be

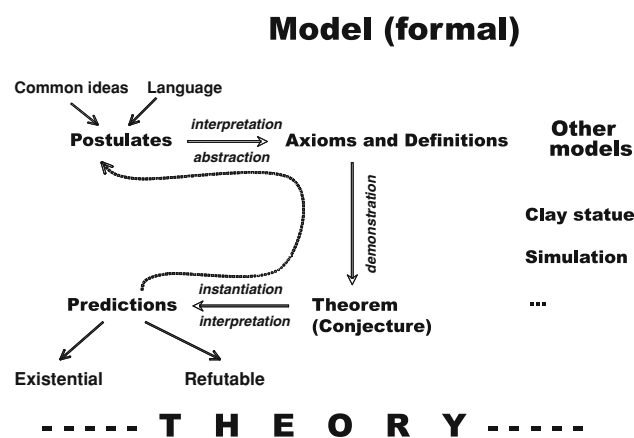


Fig. 1 Schematic representation of the dialogue between models and reality. Note that the context is essential in isolating postulates. Also, many models can co-exist, and, beside mathematical models approaches using analogies and simulations can behave as models. Contrary to Karl Popper's wish there is no clearcut link between models and reality, precluding universal processes to define the exact contours of Science

impossible to reach them in any realistic time frame. This implies that there is no time involved (no computation) to access them, assuming that the nature of mathematics does not change if n is very large. What would happen if we modified this axiom? Non-standard analysis explores our limitations if we accept that the behaviour of mathematics changes for infinitely small or infinitely large numbers (an effort that permitted Leibniz to invent differential equations) (Robinson 1996). This is mentioned here as another example of the fact, well established by Ruelle (2000), that mathematics do not exist outside reality (Desanti 1968), but belongs to it.

The present status of information in synthetic biology

Most developments of synthetic biology consider the genetic program as an algorithm, implicitly assuming that the cell behaves as a computer, a machine manipulating information. I will not repeat the argument meant to justify the model of the cell as a Turing Machine (Danchin 2009a). Suffices it to say that this implies the existence of two entities, associated via a read/write process. A *machine* is moving a device that carries a support with a *linear string of symbols written in a finite alphabet*; the data of the string of symbols, read by the machine, triggers its future actions. The focal point of representing the atom of life, the cell, as a Turing Machine, assumes the *physical separation* between machine (“chassis”) and *data/program*, represented by one or several linear strings of symbols. The crux of the model is that one should be able to isolate the entity carrying the program, put it back in a recipient host, and observe that the program in its new location displays phenomena specific of the information it carries. Beside experiments showing that pieces of program can be handled by cells (viruses and horizontal gene transfer), experiments produce results consistent with the model: 1. Animal cloning (Wilmut et al. 1997) is now commonplace; 2. The genome of a *Mycoplasma* species *M. mycoides*, was transplanted into another species, *M. capricolum*, and after several rounds of reproduction (reproduction of the machine and replication of the program, see below) the host species was replaced by a colony of the *donor* genome (Lartigue et al. 2007). This latter experiment is so important conceptually that it is essential for synthetic biology that it is reproduced in many laboratories. Yet, this might not satisfy us that the model is adequate to represent reality. Three main lines of reasoning argue against the cell-as-a-computer model.

1. The first counterargument explores the concept of Operating System (OS) (von Neumann 1958). Because the machine is separated from the program, a subset of

the program must be devoted to the interaction with the machine and its “users” (in the most general sense) (Danchin 2009a). If a particular routine is meant to reproduce the machine, then a subset of the program must be somehow linked to the architecture of the machine. Analysis of the genes giving bacteria their shape showed that there is indeed an unexpected coincidence between gene clustering in genomes and shape of bacteria (Tamames et al. 2001). In multicellular organisms, the distribution of control genes, the homeogenes, parallels the body plan: changing the order of some homeogenes in the chromosomes changed the shape of the organism, putting organs in the place of others (Gaunt 1991). Rather than an objection, the existence of a correlation between the organisation of the program and the architecture of the organism fits a prediction of the model.

2. The second counterargument is that the program is carried by some material structure, bringing about contextual information. However, this is true in computers as well: the material support of the program has its saying in permitting the machine to run properly. Different machines may be driven by the same program on different supports. Thus, even the cloning experiment, which does not involve naked DNA but a whole nucleus, with its envelope, its proteins and its RNAs, is not different from a material support of a program in a computer. Indeed, nocturnal animals use chromatin in the nuclei of neurons using the retina in an extraordinary way. Their retina can detect one unique photon. Yet, the photon receptors are located behind neurons, which absorb or diffuse photons rather than preciously conserve them. When light is dimmed, the chromatin changes transcription and reorganises in such a way that its material behaves as a lens, focusing photons on receptors located behind the neurons (Solovei et al. 2009)! This novel function for DNA, which has nothing to do with its role in carrying the genetic program, shows that another type of information has to be taken into account. In the same way, in many computers the support of the OS belongs the casing part of the chassis.
3. A third counterargument is that many rules prescribe the organisation of the cell soma, reflecting a large amount of information unrelated to the information in the program. Quite true, but this is true again for computers as well. The design of the interfaces, the microprocessors and the energy supply of the machine require much information.

In summary, two types of information (coupling of a particular form—not simply shape—with matter, energy, space and time), *information of the chassis* (casing +

metabolism) and *information of the program* are associated together in a cell (Tanaka 1984). A synthetic cell needs the association of a chassis *developing metabolism* (not a simple 3D casing) and a program similar to that found in computers. The conclusions of Dyson's argument on the double origin of life, with reproducing metabolism pre-dating replication are therefore a pre-requisite for synthesis of life (Dyson 1985). This dichotomy is visible in present synthetic biology, with a fairly clear separation between those who study the chassis (and are often also interested in the origin of life) (Kuruma et al. 2009; Shenhav and Lancet 2004) and those who think that life is essentially due to the genetic program, organising their activity around construction of program biobricks, or even as complete genomes (Gibson et al. 2008).

Information of the program

The study of the genetic program as a text, applying accepted rules of the theory of information (Shannon and Weaver 1949; Cover and Thomas 1991) to its analysis (Hénaut et al. 1996) resulted in the emphasis placed on DNA in synthetic biology. Schneider created his famous “logo” representation of sequences (Schneider and Stephens 1990) in a model of molecular machines based on Shannon's information (Schneider 1991a, b). His work was based on the intuition that creation of information was consuming energy (Schneider 1991b). Furthermore, it assumed that the data has no meaning (hence no “value”), and could be characterised purely by analysing the probability of presence of a given symbol in the sequence, generating its logo (Schneider and Stephens 1990). A similar trend is visible in the way information is used in the mass media. It is current writing—because all kinds of signals can be digitised—that everything has an information coded in sequences of (0,1), restricting the concept of information to that particular view of sequences of symbols, and forgetting about in-formation (creation and accumulation of information, or a *value* associated to an information). The common feature of this conceptualisation is dematerialisation: the corresponding information becomes an abstract entity, which can be manipulated using mathematic tools.

Yet pure abstraction is obviously inaccurate in terms of what we would like to name information. Messages without meaning (random messages) are without value. “O singe fort” in German has a meaning totally different from that in French (Yockey 1992). Can we see, even within the digitisation (or binarisation) paradigm, whether we should go further? The soviet school of electronics following Andronov, Kolmogorov, and the Americans Chaitin and Solomonoff constructed formal models of information *vs* chance by considering sequences of symbols as the result of

an algorithm. Any sequence of symbols has some *algorithmic complexity*: the length of the shortest program generating the sequence. A repeated sequence of $2n$ bits 0101010... is coded by a simple program of the type: BEGIN DO [1,n] PRINT 01 RETURN END. For n large, the program is much shorter than the sequence. In contrast, if the sequence is random (this is proposed as a definition of randomness), the only way to get the sequence is BEGIN PRINT <sequence> END, i.e. a program with a length similar to that of the sequence.

Algorithmic complexity has been related to Shannon's information (Cover and Thomas 1991) and to physical entropy: “Algorithmic randomness provides a rigorous, entropy-like measure of disorder of an individual, microscopic, definite state of a physical system. It is defined by the size (in binary digits) of the shortest message specifying the microstate uniquely up to the assumed resolution. Equivalently, algorithmic randomness can be expressed as the number of bits in the smallest program for a universal computer that can reproduce the state in question (for instance, by plotting it with the assumed accuracy). In contrast to the traditional definitions of entropy, algorithmic randomness can be used to measure disorder without any recourse to probabilities” (Zurek 1989). This success led many to think that we had a final Theory of Information, which could tell us what information is.

However, we can point out a first difficulty here. We know of an infinite set of transcendent numbers, such a π , whose digits are generated by fairly short algorithms while their succession cannot be predicted. They are therefore of limited algorithmic complexity. Yet, they are much more interesting than repeated sequences with the same complexity. Knowing the exact value of a digit placed in the digits of π , very far away in the digitisation might be interesting, but the only way to reach that value is to actualise the process of computation. Bennett named *logical depth* the time needed to reach that value and related it to physical complexity (Bennett 1988b). This is a first indication that we are far from having a thorough theory of information.

How do we have access to the information of the genetic program? Practice of computation is fairly old (well before al'Khawarizmi algorithms, Erasthones' sieve is a familiar example) but we had to wait for Pascal's computing machine, and for Lovelace and Babbage Analytical Engine to reach today's situation, with the basic concepts proposed by Turing, von Neumann and others, coupling the machine and the program, via an OS managing the “housekeeping” functions of the machine. The functions coded by the genetic program are the result of a very long evolution. And if we keep the algorithmic metaphor, because DNA comes from DNA comes from DNA... in an endless replication process, the nucleotides in the sequence have considerable logical depth.

As with computer OSs, the housekeeping program is abstract and general, yet its concrete implementation, resulting from billions of years of evolution, makes that several OSs may coexist, revealing again two kinds of information, information of the program and information of the context in which the program is expressed. This has considerable consequences for synthetic biology: cellular functions can be general and ubiquitous, whereas there is no reason why they should always be performed by structurally related objects. Overall, living cells display similar abstract features, and the genetic code argues for universality. Yet, Woese uncovered a significant discrepancy between two unicellular classes, the Archaea and the Bacteria (Woese et al. 1978). To identify ubiquitous functions operated by non-ubiquitous structures one had to devise an operational strategy, based on the concept of gene *persistence* (tendency of a given gene to be present in a quorum of species) (Fang et al. 2005). Different structural entities with common functions in different bacterial clades were indeed characterised (Danchin 2009b; Woese 2002). A structure is therefore recruited for a particular function, dependent on the *context* in which it operates. The context creates the function.

A way forward: the information of the machine/chassis

Emphasis on the idea of information as meaningless strings of symbols (Shannon and Weaver 1949), restricted our thought to that very limited feature of reality. In the Turing Machine, there is a *machine*. While its actions are explicit, nothing is said about its innards, at least when mathematicians analyse its behaviour. This is no longer so when engineers build up computers. The same is true for the chassis in synthetic biology. Not only does one need to make a machine that performs the actions of the cell/Turing Machine but this machine needs to be implemented in the real world (Tanaka 1984). It must be made of explicit matter, its actions need to be energised and there must be an Environment/Machine interaction with sensors, transporters, adhesins, safety valves (Danchin 2009c)...

The *information of the chassis* provides the relevant context which allows it to read the program and interpret it into actions (including modifying the program). Even systems which “self-organise” do not organise by themselves, but do so only when placed in proper context, which drives organisation (the DNA double helix does not form in dimethylformamide) (Grandy 1992). This type of information has a huge variety of properties: shapes, dynamics and fluxes. It displays relationships between components of the machine, and between the machine and the environment. It expresses a *situation*. It has characteristics which are somewhat similar to those of a *field* but also of a *graph*.

Typically, what we name *epigenetics* carries over chassis-type information. A great many works dealing with the study of the brain (Edelman 1987), or of cognition (Clark 1998; Ryle 1949) has taken into account this type of information. It is also at the root of much work on artificial life, learning and memory where reproduction, rather than replication is the explicit goal (see e.g. Bullock et al. 2008). But there is not yet, despite many advances, explicit consistent theories of the corresponding information (Tanaka 1984).

That we might code this information after digitisation does not place it automatically within the realm of understandable or valuable sequence information, as the very process of digitisation is only efficient knowing which type of Turing Machine would read out the corresponding sequence. This can be shown as follows. Algorithmic complexity was meant to define what chance is, because chance is the reference that permits definition of physical information: a random sequence displays the highest complexity (Cover and Thomas 1991; Zurek 1989). However this definition does not hold, as it is context-dependent (Grandy 1992). Here is an open conjecture (a preliminary version has been proposed in a different context by Wolfram (1985) and it is interesting to follow the analysis of π by Simon Plouffe: <http://www.lacim.ugam.ca/~plouffe/>). Take once again a transcendent number like π . Its digits are *pseudo-random*: for any sequence, there is a place in its digital development where one can find the sequence, whatever it is (this conjecture holds for infinite many real numbers and we would need to have a short algorithm to chose the one where the position of the sequence can be readily identified). Now, the digits can be generated by an algorithm of length N . Let us choose a putative algorithmically random sequence of length N plus an appropriate constant. The sequence can be generated by an algorithm shorter than the sequence, giving the algorithm generating the number and the position of the supposedly random sequence. Hence the sequence is not random (QED). Note that the value of the information (logical depth) is not fixed. It depends on the algorithm, as it *differs* in π and in any other transcendent algorithmically generated real number. Said otherwise, the complexity and depth of the sequence *depends on the algorithm* i.e. on the context. Provided that we can prove the conjecture, this fits exactly the objection raised against the cell-as-a-computer model: beside information in the program, there must be information in the machine (providing the context). Hence, the information of the machine is not described by our present theory of information.

In the definition of logical depth, we have implicitly used a property of algorithms, *recursivity*. In 1931 Gödel constructed a recursive algorithm, which, when decoded, translated into a particular proposition, which, briefly, stated: “I am impossible to prove”. Moving from one context to another one, *recursivity created a novel information*, the

statement of a fact of which the world was previously unaware. The genetic code, which enables nucleic acids to be translated into proteins, which in turn manipulate nucleic acids, behaves exactly as Gödel's procedure does (Danchin 2003, 2009a). The consequence of this demonstration is that a purely deterministic system, with known initial conditions, may have an entirely unpredictable outcome. By contrast with the mechanistic philosophy, even with its more modern appendices such as feedback and feedforward loops, recursivity brings about novelty not because we fail to grasp all initial conditions of a particular phenomenon, but because it can be only understood a posteriori, after it has unfolded in space and time. This implies that synthetic biology, *when it takes recursivity into account*, develops in a world that is totally irreducible to the world of systems biology, which remains an elaborate episode of the study of mechanistic automata. An important consequence is that what we commonly term the "genetic program" because it unfolds through time in a consistent manner is not a programme with an aim (we would not be able to predict any aim) - it is merely there, and functions because it cannot do otherwise. We only perceive a design because the end result is familiar to us, and thus seems more "right" than any other possible result (Danchin 2009a).

If creation of information depends heavily on the context, we must identify in living organisms functions which permit it to accumulate, and relate it to the material world. How are particular structures or processes recruited to this aim? A reflection on the coupling between accumulation of information and energy based on work developed by Landauer and Bennett, showed that the relationship between energy and information is that there exists degradative processes which "make room" using energy to *prevent* degradation of what is functional ((Bennett 1988a; Landauer 1961), detailed analysis in (Danchin 2009a, d), see also a very recent development (Sagawa and Ueda 2009)). In such a situation it is the context that determines which gene product is functional and which is not. The consequence is that, if the context does not vary too rapidly, then the functions which will be selectively retained are sculpting an image of the environment within, creating adaptation. This is exactly how an information can get a meaning. In terms of synthetic biology, this orients research towards learning and memory, rather than towards fixed mechanical engineering.

In guise of conclusion: the brain is not a computer, yet it manipulates information

Trying to put vitalism to an end, Claude Bernard placed biology within the realm of physics and chemistry (Bernard 1865). This led his followers to ask the question: what are the relevant entities (material objects and processes) which

make a cell alive? The biochemical inventory stage started well, with the discovery of the ribosomes, of the structure of the DNA double helix, of the sequence of the polypeptide chain of insulin, and, rapidly of messenger RNA (Judson 1979). Yet, many features of biological entities resisted the classical analysis of chemistry and physics. This was apparent in the laws of genetics, where linear arrangements of the elusive genes was central (Gayon 2007). Even in biochemistry the shape of molecules posed an enigma: *La dissymétrie, c'est la vie*, insisted Pasteur. But the involvement of shape was deeper than usual: the very process of replication placed the concept of *form* in a world quite different from the simple arrangement of a particular setup in 3D as shape would suggest. Replication shifted the idea of a chemical as the *substrate* of a recognition process to that, abstract, of a *template*, in this case for a duplication process doubling the number of the initial molecule. Subsequently, the discovery of transcription, translation, and associated control and coding processes, continued to shift emphasis from shape to form in an abstract way, commonplace in mathematics.

Information—creating and manipulating form—was essential to account for life processes. From the world of Plato's archetypes, those who explored the basic concepts of life resorted to discussions which began with Aristotle and placed form as a central category of reality. For some time, and this is still quite visible in systems biology as well as in synthetic biology, living organisms were seen as mechanistic automata, with feedback and feedforward loops as paradigmatic entities. The purpose of the present reflection was to try and show that investigating the concept of information shifts eighteenth century's automata to modern algorithmic machines, capable of authentic creation. This implied replacing feedback by recursivity, a much deeper process. Recursivity, associated to appropriate management of energy (Bennett 1988a; Landauer 1961; Sagawa and Ueda 2009), creates information (Danchin 2009a). It does so by identifying two domains where information must be taken into account: *information of a program* and *information of a machine*. However, while the information of the program is fairly deeply explored by a vast community of investigators, this is not so of the information of the machine/chassis, which involves some kind of measurement of the context (in terms of implementation within the four categories, matter, energy, space and time) (Tanaka 1984; Sagawa and Ueda 2009).

It is perhaps in the functioning of the brain that we can make the latter type of information most prominent. Indeed, while von Neumann and others invented computers with mimicking the brain in mind (von Neumann 1958), the brain does not appear to behave as a Turing Machine (Edelman 1987). There is no "ghost in the machine" (Ryle 1949). However, nobody would doubt that brain manages

information, and in a very efficient way (Clark 1998; Bullock et al. 2008). To my view this is a strong indication that the information we describe when considering messages is a tiny part of what information is. Because we use language, built on the exchange of sequences of symbols, exactly as programs are exchanged in computers, linguists often saw the brain as a Turing Machine. But language is deeply associated to meaning: I had in 1974 at a meeting of the Centre Royaumont pour une Science de l'Homme at the MIT, a heated argument with Noam Chomsky about other features of human languages, such as rhythm (in the west african language *mòré*, a speaker may begin a rhythmic sentence, which is answered preserving rhythmic rules by somebody in the audience, triggering another ejaculation of the speaker, with related rules, etc) suggesting that beside grammatical syntactic structures, there may exist a variety of superimposed *contexts* which transmit information mediated by channels that are not those usually considered (Danchin 1987, Danchin and Marshall 1987; Marshall et al. 1987). As in Dyson's scenario of the origin of life, the basic functioning of the brain would base on reproduction, while invention of language with its linear sequences of phonemes, when spoken, and letters when written, would be, in Man, the transition moment when it would begin to discover recursivity in linear strings of symbols (phonemes) which can be propagated from brain to brain, as programs in a Turing Machine. In any event, in the few cases where it might do so, it would be an extremely slow one (Sackur and Dehaene 2009). With this view, Nature would have discovered twice the importance of coding and recursivity, in the emergence of life first, and in the emergence of language, quite recently.

Acknowledgements I wish to thank Philippe Binder for pointing out to me Wolfram's article, Jean-Baptiste Masson for pointing out Sagawa and Ueda's article and anonymous reviewers for very constructive comments. This work developed in the Stanislas Noria network is supported by the PROBACTYS and the TARPOL European Union programmes in Synthetic Biology (http://www.normale sup.org/~adanchin/causeries/causeries_en.html).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Averroës, Tahafut al Tahafut, EJW Gibb Memorial Trust, Cambridge (UK) ca 1175 (trans. latin 1497)
- Bennett C (1988a) Notes on the history of reversible computation. *IBM J Res Dev* 44:270–277
- Bennett C (1988b) Logical depth and physical complexity. In: Herken R (ed) *The universal turing machine: a half-century survey*. Oxford University Press, Oxford
- Bernard C (1865) *Introduction à l'étude de la médecine expérimentale*. Garnier-Flammarion (réed. 1966), Paris
- Bullock S, Noble J, Watson R, Bedau MA (ed) (2008) *Artificial life XI*. MIT Press, Cambridge
- Chan LY, Kosuri S, Endy D (2005) Refactoring bacteriophage T7. *Mol Syst Biol* 1:0018
- Changeux JP, Courge P, Danchin A (1973) A theory of the epigenesis of neuronal networks by selective stabilization of synapses. *Proc Natl Acad Sci USA* 70:2974–2978
- Clark A (1998) *Being there: putting brain, body, and world together again*. MIT Press, Cambridge
- Clemmer CR, Beebe TP Jr (1991) Graphite: a mimic for DNA and other biomolecules in scanning tunneling microscope studies. *Science* 251:640–642
- Cover T, Thomas J (1991) *Elements of information theory*. Wiley, New York
- Danchin A (1986) Order and necessity. In: Quagliariello E, Bernardi G, Ullmann A (eds) *From enzyme adaptation to natural philosophy: heritage from Jacques Monod*. Symposium J Monod and molecular biology, yesterday and today. Elsevier Sciences Publishers, Amsterdam, Trani, Italy: 13–15 Dec 1986, pp 187–196
- Danchin A (1987) Biological foundations of language: a comment on Noam Chomsky's approach of syntactic structures. In: Mogdil S, Mogdil C (eds) *Noam chomsky, consensus and controversy*. The Falmer Press, New York, pp 29–39
- Danchin A, Marshall JC (1987) Cross replies. In: Mogdil S, Mogdil C (eds) *Noam chomsky, consensus and controversy*. The Falmer Press, New York, pp 50–55
- Danchin A (1992) Science and technology: a western imbroglio. *Projections* 7/8:39–48
- Danchin A (2003) *The Delphic boat. What genomes tell us*. Harvard University Press, Cambridge
- Danchin A (2009a) Bacteria as computers making computers. *FEMS Microbiol Rev* 33:3–26
- Danchin A (2009b) A phylogenetic view of bacterial ribonucleases. *Prog Mol Biol Transl Sci* 85:1–41
- Danchin A (2009c) Cells need safety valves. *Bioessays* 31:769–773
- Danchin A (2009d) Natural selection and immortality. *Biogerontology* 10:503–516
- Desanti J-T (1968) *Les Idéalités Mathématiques*. Le Seuil, Paris
- Diels H, Kranz W (1935) *Die Fragmente der Vorsokratiker*. Weidmann, Berlin
- Dyson FJ (1985) *Origins of life*. Cambridge University Press, Cambridge
- Edelman G (1987) *Neural darwinism. The theory of neuronal group selection*. Basic Books, New York
- Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403:335–338
- Endy D (2005) Foundations for engineering biology. *Nature* 438: 449–453
- Erigène JS *De la division de la Nature (Periphyseon)*, Presses Universitaires de France, Paris ca 850 (trans 1995)
- Fang G, Rocha E, Danchin A (2005) How essential are nonessential genes? *Mol Biol Evol* 22:2147–2156
- Gaunt SJ (1991) Expression patterns of mouse Hox genes: clues to an understanding of developmental and evolutionary strategies. *Bioessays* 13:505–513
- Gayon J (2007) The concept of the gene in contemporary biology: continuity or dissolution? *Logic, Epistemol Unity Sci* 6: 91–95
- Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA et al (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* 319:1215–1220
- Grandy WT (1992) On randomness and thermodynamics. *Found Phys* 22:853–866

- Hénaut A, Danchin A (1996) In Neidhardt F (ed) *Escherichia coli* and salmonella, cellular and molecular biology. ASM Press, Washington, pp 2047–2065
- Judson H (1979) The eight day of creation. The makers of the revolution in biology. Simon and Schuster, New York
- Kuruma Y, Stano P, Ueda T, Luisi PL (2009) A synthetic biology approach to the construction of membrane proteins in semi-synthetic minimal cells. *Biochim Biophys Acta* 1788:567–574
- Lakatos I (1976) Proofs and refutations. Cambridge University Press, Cambridge
- Lakatos I (1980) The methodology of scientific research programmes, vol 1: philosophical papers. Cambridge University Press, Cambridge
- Landauer R (1961) Irreversibility and heat generation in the computing process. *IBM J Res Dev* 3:184–191
- Lartigue C, Glass JI, Alperovich N, Pieper R et al (2007) Genome transplantation in bacteria: changing one species to another. *Science* 317:632–638
- Livingstone D (2003) School of computing. University of the West of Scotland, Paisley Campus, Scotland, p 193
- Marshall JC (1987) Language learning, language acquisition, or language growth? In: Mogdil S, Mogdil C (eds) Noam chomsky, consensus and controversy. The Falmer Press, New York, pp 41–49
- Monod J (1971) Chance and necessity: an essay on the natural philosophy of modern biology. Alfred A Knopf, New York
- Popper K (1959) The logic of scientific discovery. Hutchinson, London
- Popper K (1963) Conjectures and refutations: the growth of scientific knowledge. Routledge, London
- Putnam H (1988) Representation and reality. MIT Press, Cambridge
- Robinson A (1996) Non-standard analysis. Princeton University Press, Princeton
- Ruelle D (2000) Mathematical platonism reconsidered. *Nieuw Archief voor Wiskunde* 5:30–33
- Ryle G (1949) The concept of mind. The University of Chicago Press, Chicago
- Sackur J, Dehaene S (2009) The cognitive architecture for chaining of two mental operations. *Cognition* 111:187–211
- Sagawa T, Ueda M (2009) Minimal energy cost for thermodynamic information processing: measurement and information erasure. *Phys Rev Lett* 102:250602/250601–250602/250604
- Schneider TD (1991a) Theory of molecular machines. I. Channel capacity of molecular machines. *J Theor Biol* 148:83–123
- Schneider TD (1991b) Theory of molecular machines. II. Energy dissipation from molecular machines. *J Theor Biol* 148:125–137
- Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* 18:6097–6100
- Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana
- Shenhav B, Lancet D (2004) Prospects of a computational origin of life endeavor. *Orig Life Evol Biosph* 34:181–194
- Solovei I, Kreysing M, Lanctôt C, Kösem S et al (2009) Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* 137:356–368
- Tamames J, Gonzalez-Moreno M, Mingorance J, Valencia A, Vicente M (2001) Bringing gene order into bacterial shape. *Trends Genet* 17:124–126
- Tanaka M (1984) A physical characterization of biological information and communication system model of ecosystems. *J Theor Biol* 110:619–635
- Veening JW, Smits WK, Kuipers OP (2008) Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol* 62:193–210
- von Neumann J (1958) The computer and the brain. Yale University Press, New Haven (reed 1979)
- Wilmot I, Schnieke AE, McWhir J, Kind AJ, Campbell KH (1997) Viable offspring derived from fetal and adult mammalian cells. *Nature* 385:810–813
- Woese CR (2002) On the evolution of cells. *Proc Natl Acad Sci USA* 99:8742–8747
- Woese CR, Magrum LJ, Fox GE (1978) Archaeobacteria. *J Mol Evol* 11:245–251
- Wolfram S (1985) Origins of randomness in physical systems. *Phys Rev Lett* 55:449–452
- Yockey HP (1992) Information theory and molecular biology. Cambridge University Press, Cambridge
- Zurek W (1989) Algorithmic randomness and physical entropy. *Phys Rev A* 40:4731–4751