Progress in Biophysics and Molecular Biology 109 (2012) 108–114

Contents lists available at SciVerse ScienceDirect

Progress in Biophysics and Molecular Biology

journal homepage: www.elsevier.com/locate/pbiomolbio

Review

Is information a proper observable for biological organization?

G. Longo[a,*], P.-A. Miquel[b], C. Sonnenschein[c], A.M. Soto[c]

[a] CREA, Ecole Polytechnique, 32 Boulevard Victor, 75015 Paris, France
[b] Laboratoire CEPERC, UMR CNRS, 6059, Université de Provence, France
[c] Department of Anatomy and Cellular Biology, Tufts University School of Medicine, 136 Harrison Avenue, Boston, MA 02111, USA

ARTICLE INFO

Article history:
Available online 13 July 2012

Keywords:
Entropy
Complexity
Program
Determinism
Code
Signal

ABSTRACT

In the last century, jointly with the advent of computers, mathematical theories of information were developed. Shortly thereafter, during the ascent of molecular biology, the concept of *information* was rapidly transferred into biology at large. Several philosophers and biologists have argued against adopting this concept based on epistemological and ontological arguments, and also, because it encouraged genetic determinism. While the theories of elaboration and transmission of information are valid mathematical theories, their own logic and implicit causal structure make them inimical to biology, and because of it, their applications have and are hindering the development of a sound theory of organisms. Our analysis concentrates on the development of information theories in mathematics and on the differences between these theories regarding the relationship among complexity, information and entropy.

Contents

1. Introduction

The concept of information has dominated biological discourse particularly in genetics and molecular biology. Although biologists have for the most part embraced this notion, others have raised objections "on the ground that enthusiasm for information in biology has been a serious theoretical wrong turn", and because "…it fosters naive genetic determinism" (Godfrey-Smith and Sterelny, 2008; Hacking, 1999). Our current analysis stems from information theories in mathematics and addresses the differences between these theories regarding the relationship among complexity, information and entropy. These differences have led to intrinsic inconsistencies and ambiguities when metaphorically applied to biology. Beyond the observables[1] known from physics,

* Corresponding author. Tel.: +33 (0) 145526420.
E-mail address: Giuseppe.Longo@ens.fr (G. Longo).

[1] A measurable property of a physical system, such as mass or momentum. In quantum mechanics, observables correspond to mathematical operators used in the calculation of measurable quantities (from http://www.thefreedictionary.com/observable).

0079-6107/$ – see front matter
http://dx.doi.org/10.1016/j.pbiomolbio.2012.06.004

there is a need for new observables in biology that would increase its intelligibility and facilitate the quantification of biological organization. *Information* was erroneously thought to be the key observable. Thus we will argue that the concept of information is not applicable to biology and point out which have been the unintended consequences of the information metaphor.

2. The *elaboration* of information, or how the story began

In 1931, Kurt Gödel, a major logician, invented the number-theoretic coding of everything. With an idea that now may seem obvious, he associated one-to-one letters to numbers and by a clever use of arithmetic properties he coded all sentences of formalized mathematics as numbers. A few years later, Alan Turing made out of this idea the fantastic Logical Computing Machine for manipulating all sentences of formal theories (Turing, 1936). He accomplished this by providing instructions, later called *programs*, also subject to being coded by numbers, that is, by sequences of zeros and ones. And thus, he invented the modern theory of *elaboration* of information. The program was encoded by the same structure as the data, a sequence of 0s and 1s. A Universal Machine could then compute with any program by using part of its memory to code for the program and another for the data. This "architect", as Schrödinger called it in 1945, uses a string of numbers as a program (Schrödinger's "plan") to carry on the computation over data, what is now called an operating system and a "compiler".

Under the form of code-script, Schrödinger first hinted at *information*[2] as the observable proposed by the newly invented Theory of Computation and of Coding (Schrödinger, 1945). This new discrete observable would be straight-forwardly found in chromosomes, the "aperiodic crystals" envisioned by Schrödinger. The base pair complementarity of DNA discovered in 1953 strengthened the force of the information metaphors and provided an uncomplicated conceptual frame for the analysis of the transmission of hereditary traits (Mendelian inheritance) and for the "reading" of a (coded) gene sequence into its corresponding protein.

Biologists sensed the *zeitgeist* and followed the fashion of the time, as if these notions where "causally neutral"; however, they are not. Note that Frege's and Hilbert's mathematical logic (Barwise, 1978) were proposed as a foundation of mathematics based on Arithmetic and programmatically departing from physics and its foundations (Bailly and Longo, 2011). Gödel's and Turing's work addressed questions in this purely linguistic and abstract context. Thus, from this mathematical logic perspective, the encoding of the Aristotelian "essence" of an organism into an aperiodic crystal was not supposed to bring into biology any intended "structure of physical determination" or causal relationship. Instead, it was supposed to merely imply the "soft" symbol of manipulation, typical of coding and programming, which was later associated by Crick to the principle of genetic determination. In a famous paper, Crick claimed that the transcription of the *information* contained in DNA molecules — along with its translation — takes place "in a linear order" that starts from the "genetic material" and ends with the synthesis of proteins. Thus, genetic information could be stored in DNA (or RNA) and become known by the reading of the nucleic acid sequence, and that interactions among proteins, lipids, and other chemical entities could not change it. He stated explicitly in this sense, that: "the sequence of bases determines the sequence of amino acids" (Crick, 1958).

[2] *Information* in italics denotes the use of this concept in biology in contraposition of information in mathematics.

Schrödinger rightly acknowledged the peculiar deterministic nature of his proposal when he wrote « …In calling the structure of the chromosome fibers a code-script we mean that the all-penetrating mind, once conceived by Laplace, … could tell from their structure whether the egg would develop, under suitable conditions, into a black cock or into a speckled hen…» (Schrödinger, 1945). In code-script, regardless of whether it is a matter of programming or cryptography, determination is "Laplacian", that is, it implies causality and predictability.[3] Programming and decrypting led in principle to predictable dynamics on discrete and exact data types. From Gödel's primitive *recursive* functions to portability of software, the purpose of programming is *identical iteration*. Primitive recursion is iteration plus updating a register. Programs ought to be *portable*, that is, they should iterate identically even in slightly different, but compatible, computational environments. When re-launching a program, even in a network, or when opening a webpage in say, Japan, or South Africa, one would want both the program and the website to run identically … always! And in computer networks this is hard, as space-time continua step in. Yet, it works! This is also true in cryptography: given the key, decoding must work as expected and iterate, as all ordinary bank teller machines do.

Biological phenomena, instead, do not exhibit this predictability. For example, a relatively low percentage of fertilized eggs produce live births. Also, normal cells extracted from humans rarely can be "established" in culture conditions. In addition, only one in hundreds of "nuclear transfer embryos", like Dolly, are born, and so on. Mathematically, on the other hand, codes and programs are made to work exactly. The program may be too long or its results yet unknown in practice, but *in principle*, everything should be in it. This is why, in 1945, Schrödinger wrote that if the chromosomes are indeed a code-script, once they become decoded, deterministic predictability should fully display its power. Hopefully, then, we would know in which way "the DNA code … is the *program* for the behavioral computer of this individual" (Mayr, 1961). Today, however, this Laplacian structure of determination going from the DNA to the phenotype is highly controversial. More modestly, most biologists view the genome just as a program for producing proteins (Danchin, 2003). Yet, a process is "programmable" (i.e., modeled in a faithful way by a program on discrete data types) if, and only if, it is deterministic and predictable, and thus Laplacian (Bailly and Longo, 2011).

3. Information, entropy and negentropy in two different perspectives

3.1. The analysis of transmission of information

The word "information" is absent in Schrödinger's book. However, he writes about *code-script*, meaning the "order" coded by and in an aperiodic crystal. This was an audacious proposal into the possible structure of chromosomes, especially because at that

[3] In short, a physical system is deterministic when it has a "conceivable" determination in explicit mathematical terms – equations, evolution functions… For Laplace (1749–1827), any deterministic system is *predictable* and unpredictability, as randomness, had to be analyzed in purely probabilistic terms. This turned out to be false, since, in his Geometry of Dynamical System, born from the Three Body Problem, Poincaré (1854–1912) showed that classical randomness may be understood as "unpredictable determinism". This is typically due to the role of non-observable fluctuations (fluctuations below the best physical measure) in continuous dynamics, which may produce relevant observable effects by non-linear amplifications. This was a consequence of Poincare's proof of non-analytical solvability of the equations for three celestial bodies in their gravitational fields (1890), see (Barrow-Green, 1997).