The notion of information in biology, an appraisal

Jérôme Segal

University Paris-IV, 2014/2015 on sabbatical, France

Abstract. Developed during the first half of the 20th century, in three different fields, theoretical physics, statistics applied to agronomy and telecommunication engineering, the notion of information has become a scientific concept in the context of the Second War World. It is in this highly interdisciplinary environment that "information theory" emerged, combining the mathematical theory of communication and cybernetics. This theory has grown exponentially in many disciplines, including biology. The discovery of the genetic "code" has benefited from the development of a common language based on information theory and has fostered a almost imperialist development of molecular genetics, which culminated in the Human Genome Project. This project however could not fill all the raised expectations and epigenetics have shown the limits of this approach. Still, the theory of information continues to be applied in the current research, whether the application of the self-correcting coding theory to explain the conservation of genomes on a geological scale or aspects the theory of evolution.

In a comment published in October 2014 in the well-known journal Nature, specialists of evolutionary theory discussed with "point" and "counterpoint" whether their field needed a complete rethink or not [1]. The advocates of a theoretical change stated: "We hold that organisms are constructed in development, not simply "programmed" to develop by genes. Living things do not evolve to fit into pre-existing environments, but co-construct and coevolve with their environments, in the process changing the structure of ecosystems." On the online version of the paper, David Tyler from the Cell Biology Department of the University of Valencia made this insightful remark: "It seems to me that the key issues relate to biological information and the word 'evolution' needs to be defined in a way that does justice to the origin (or loss) of information." In the same vein, a geneticist like Antoine Danchin worked on the links between the reproduction of cells and the replication of genomes, in relation to ageing [2]. He came to the conclusion that "making young structures from aged ones implies creating information" and added "I revisit Information Theory, showing that the laws of physics permit de novo creation of information, provided an energy-dependent process preserving functional entities makes room for entities accumulating information." Information seems to appear a central notion, next to matter, energy, time and mass. In the last couple of years scientists like Vedral or Battail respectively tried to reinterpret physics or biology putting the notion of information at the core or their approach [3, 4].

What can be the place of historians in such a contemporary scientific discussion? The lack of hindsight make it from the beginning a very difficult task to undertake. Most of those who recognized themselves as historians will officially consider that assessing the value of current scientific research does definitively not belong to their goals. Marc Bloch, the father of the famous Annales school,

criticized sixty years ago the tendency to judge in his *Apologie pour l'histoire ou Métier d'historien* (The Historian's Craft): "Unfortunately, by judging one finishes almost fatally by losing even the liking to explain" [5]. If the understanding matters more than the judging, Marc Bloch nevertheless accepted the idea of defining history as a science concerned with truth: "History might be proud to have (...) by a precise elaboration of its technique, opened for the mankind a new way towards the truth, and consequently the right."

The purpose of the following pages is thus to comment on the early uses of information theory in biology in order to be able to understand more recent scientific work in an historical perspective. A few considerations on the history of the use of this theory in physics might also give hints in this direction.

1. The early use of information theory in biology

First of all, what is meant by "information theory"? For many people the expression is associated to the name of a mathematician and engineer, Claude Elwood Shannon (1916–2001). Working in the Bell Telephone Laboratories, he drew upon the work of his fellow colleague Ralph Hartley who had already defined information in 1927 for the case of a message made of selection of elements of equal probability. His 1948 publications in the Bell System Technical Journal essentially rely on three main ideas [6]:
 – the description of a general schema of a communication in a noisy channel
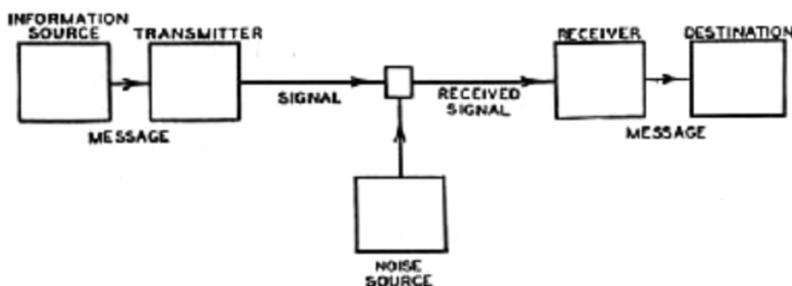
 – the definition of the quantity of information in bits, $H = -\Sigma\ p_i \log p_i$ where pi is the probability to select each element that may constitute the message (an analogous formula if offered for the continuous case)
 – the definition of the capacity of channel in bits per second

$$C = B \log_2 \left(1 + \frac{S}{N} \right)$$

where B is the bandwidth of the channel; S is the average received signal power; N is the average noise; and S/N is the signal-to-noise ratio. Shannon proved that a code exist which allows to obtain a capacity as close as one wish from this limit value.

The posterity of these publications is due to Warren Weaver, heading from 1932 to 1955 the Division of Natural Sciences at the Rockefeller Foundation. It is his credit to have clearly shown in his introducing paper to The mathematical Theory of Communication the limits of Shannon's formalism [7]. He distinguished three levels with different questions: Level A, "How accurately can the symbols of communication be transmitted?" (The technical problem); Level B, "How precisely do the transmitted symbols convey the desired meaning?" (The semantic problem); and Level C, "How effectively does the received meaning affect conduct in the desired way?" (The effectiveness problem). The meaning is clearly excluded from the first level and the history of information theory may be seen as different temptations to introduce meaning in Shannon's theory.

Shannon himself, the father of the mathematical theory of communication almost never used the expression information theory (see Shannon on the "Bandwagon" that represents this theory [8]).

Information theory must be understood in its historical context, as a woolly theory closed to cybernetics, a term coined by the mathematician Norbert Wiener to describe a general theory of control and communication which applies in living or technical systems as well as in the analysis of the society [9]. Significantly, the first bibliography published in the Institute of Radio Engineers (IRE) Transactions on Information Theory is entitled "A Bibliography on Information Theory (Communication Theory - Cybernetics)" [10].

The identification of different kinds of feedback loops represents one of the key ideas of cybernetics. It is in this general frame that information is defined as a scientific notion, namely the quantity that circulate in these loops. Biology was present from the very beginning since Wiener start working on cybernetics after his discussion with the cardiologist Arturo Rosenblueth who told him about biological feedback problems [11].

Therefore, there are many kinds of applications concerning the use of information theory in biology. Some directly stem from Shannon's work, some others are closer to cybernetics. As for the first case, one might think of Quastler's symposium, organized in 1952. This Viennese radiologist émigré invited scientists from different horizons to discuss and apply Shannon's model. Among them, Herman Branson, from the Harvard physics department, who calculated the information quantity (H) contained in a man with the expression "H(food and environment) = H(biological function) + H(maintenance and repair) + H(growth, differentiation, memory)" ([12, p. 39]). Quastler came with Sidney M. Dancoff to the conclusion that "H(man)" was about 2.1028 bits (p. 167). Fifty years later, such calculus might look pretty useless for the development of biology but they are intertwined with early reflections on the notion of information which are probably worth a quote, thinking of today's discussions. Quastler declared for instance:

> "Information Theory" is a name remarkably apt to be misunderstood. The theory deals, in a quantitative way, with something called "information", which, however, has nothing to do with meaning. On the other hand, the "information" of the theory is related to such diverse activities as arranging, constraining, designing, determining, differentiating, messaging, ordering, organizing, planning, restricting, selecting, specializing, specifying, and systematizing; (...) (p. 41)

Quastler clearly explained that specification can also be analyzed by information theory but that the information which one refers to in the information theory has nothing to do with its meaning (a point which was clearly stated by Shannon and before him Ralph Hartley).

In 1952, at the time of this symposium, the DNA structure was still unknown. Quastler and Dancoff were however working under the assumption that "somewhere in the chromosomes [there] is a system of determiners, called the genome" and that "it contains a substantial fraction of all information about an adult organism" (p. 269). The amazing enthusiasm which accompanied the rapid development of information theory explains how a strong determinism arose in modern genetics. A kind of "fetichisation" of genetic information occurred, which was later reinforced by the identification of a "genetic code". This lead in the 1990s to a strong belief in a genetic determinism as the titles of many publications witness (alcoholism gene, homosexuality gene...)

The "RNA Tie Club", this informal group of scientists who worked from 1954 onwards with Georges Gamow to "solve the riddle of RNA structure, and to understand the way it builds proteins", basically applied information theory and more precisely the mathematical theory of communication. As Lily Kay has brilliantly shown, if they did not succeed on the scientific level, this work enabled the constitution of a scientific community of people sharing the same vocabulary [13]. Information theory was at the origin of the creation of what Galison called a "trading zone" for an area where a new language is developed to permit commerce between cultures that lack a common language.

The cybernetics parts of information theory has also been used in biology, for instance by Jacques Monod in the operon model which included positive and negative feedback regulation (to represent

enzyme induction and coordinate repression). A whole chapter of Chance and Necessity is even entitled "microscopic Cybernetics" [14].

Nowadays, many different parts of information theory are used in biology: algorithmic complexity, quantum information, informatics among others. In order to gain a better understanding of this situation, a short excursion in the realm of physics might be useful.

2. Considering the role of information theory in physics

During the 1920s, about twenty years before what we know today as the "information theory", information has been defined independently as a scientific notion in three different domains: statistics (with the work of Fisher), communication engineering (mostly at the Bell Labs) and physics (in the German speaking part of Europe). The Polish physicist Marjan von Smoluchowski (1872–1917) was the first in 1912 to tackle this subject, almost indirectly. Connecting the problem of Maxwell's Demon with that of Brownian motion, he wrote that in order to violate the second principle of thermodynamics, the Demon had to be "taught" (in German "unterrichtet") regarding the speed of molecules. Szilard referred seventeen years later to this result in a paper he published in German and it is only in the 1964 translation in English that we read that the Demon must be "informed". Meanwhile, other physicist clearly recognised the importance of information. As early as 1930, Gilbert Lewis wrote "Gain in entropy always means loss of information, and nothing more" [15–17].

The relation between entropy and information started being of the most importance after Shannon decided to call entropy the quantity of information defined in the discontinuous case by $H = -\Sigma p_i \log p_i$ where all the p_i stand for the different selection probabilities. It seems that it is the mathematician Johannes von Neumann who advised him, aware of the possible ambiguity. Since the physical quantity, entropy, is wrongly connected to disorder (an eminent subjective notion), this formal analogy has been tremendously misleading, among others in the development of the so-called "second cybernetics" based on the idea that noise could create order. Norbert Wiener also played a decisive role for the development of information theory in physics when he wrote in Cybernetics that "Information is information, not matter or energy". From the early 1950s, information started being used as a physical notion.

In the introduction of the proceedings of the 1952 symposium on the use of information theory in biology, Quastler recognized that information as a new quantity. He started with these words:

> One of the basic tools in natural science is the energy concept. In recent years, another concept has begun to attain comparable dignity. It is something more subtle and elusive than energy; it is derived from a desire for dealing methodically with problems of complexity, order, organization, specificity...

The French physicist Léon Brillouin decided at this time to reinterpret whole chapters of physics considering information as the main notion. He published his first paper on this topic in 1949 and seven years later a book entitled in Science and information theory [18].

Nowadays, the scientific notion of information has still an important heuristic power. Let us consider two examples. Firstly, G. Cohen-Tannoudji, professor of theoretical particle physics, recently proposed to introduce a fifth fundamental constant after G (gravity), c (light speed), h (Planck's action quantum), and k (Boltzmann's constant), namely b for a Brillouin constant, which should enable a new approach of complexity in physics [19]. Secondly, a professor in optical sciences, Roy Frieden, published a book in 1998 with the title Physics from Fisher Information to show that it is the scientific definition of information introduced in statistics in the 1920s by Ronald Fisher that could help deriving physics. Regarding the connection to entropy, he wrote:

> This book defines and develops a unifying principle of physics, that of "extreme physical information". The information in question is, perhaps surprisingly, not Shannon or Boltzmann entropy but, rather, Fisher information, a simple concept little known to physicists. Both statistical and physical properties of Fisher information are developed. This information is shown to be a physical measure of disorder,

sharing with entropy the property of monotonic change with time. The information concept is applied "phenomenally" to derive most known physics, from statistical mechanics and thermodynamics to quantum mechanics, the Einstein field equations, and quantum gravity. Many new physical relations and concepts are developed, including new definitions of disorder, time and temperature. The information principle is based upon a new theory of measurement, one which incorporates the observer into the phenomenon that he/she observes. The "request" for data creates the law that, ultimately, gives rise to the data. The observer creates his or her local reality [20].

Debates are still raging regarding the use of information theory, for instance in physics with Freiden or Vedral mentioned above in the introduction. In a lot of other disciplines, information theory plays different kind of roles, scientific, epistemological or sociological [21, 22].

3. Perspectives

Information theory constitutes today a tremendous opportunity to develop interdisciplinarity. We cannot assess now the scientific value of ongoing works, but one thing is sure, information theory offers a interdisciplinary language. An communication engineer like G. Battail published for instance in the Europhysics letters with such a question like "Does information theory explain biological evolution?" using the theory of error correcting codes [23 and 4].

Almost all the main developments of information theory are used in biology, even the quite abstract theory of algorithmic complexity (developed by Kolmogorov, Solomonoff, Martin-Löf and Chaitin). The emerging field of biocomputing also uses a scientific definition of information. Established as a scientific notion in the middle of the 20th Century, information is probably a concept that is still promised to a great future in the life sciences, in their relation with other disciplines.

References

[1]   Laland, Kevin, Tobias Uller, Marc Feldman, Kim Sterelny, Gerd B. Müller, Armin Moczek, Eva Jablonka, et al. "Does Evolutionary Theory Need a Rethink?" Nature **514** (7521): 161–64 (2014)
[2]   Danchin, Antoine, "Natural Selection and Immortality." Biogerontology **10** (4): 503–16 (2009)
[3]   Vedral, Vlatko, *Decoding Reality: The Universe as Quantum Information*. Oxford [England]; New York: Oxford University Press (2010)
[4]   Battail, Geìrard, *Information and Life*. Dordrecht; New York: Springer (2014)
[5]   Bloch, Marc, *Apologie pour l'histoire ou Métier d'historien*. Paris: Armand Colin (1949, 1997)
[6]   Shannon, C. E. "A Mathematical Theory of Communication." Bell System Technical Journal 27 **3**: 379–423 (1948)
[7]   Shannon, Claude E., and Warren Weaver, *The Mathematical Theory of Communication*. First Edition. Urbana, Ill.: University of Illinois Press (1949)
[8]   Shannon, Claude E., "The Bandwagon." IRE Transactions on Information Theory, (**2**) March 2 (1956)
[9]   Wiener, Norbert, *Cybernetics: Or Control and Communication in the Animal and the Machine*. The Technology Press//Wiley (1948)
[10] Stumpers, F. "A Bibliography of Information Theory (communication Theory - Cybernetics)." *IRE Transactions on Information Theory* **1**(2): 31–47 (1955)
[11] Heims, Steve Joshua, *The Cybernetics Group*. Cambridge, Mass: The MIT Press (1991)
[12] Quastler, Henry, *Essays on the Use of Information Theory in Biology*. University of Illinois (1953)
[13] Kay, Lily, *Who Wrote the Book of Life?: A History of the Genetic Code*. 1 edition. Stanford, Calif.: Stanford University Press (2000)
[14] Monod, Jacques, *Le Hasard et la Nécessité*. Paris: Seuil (1973)

[15] Smoluchowski, Marian, "Experimentell nachweisbare, der üblichen Thermodynamik widersprechende Molekularphänomene," Physikalische Zeitschrift **13**: 1069-108 (1927)

[16] Szilard, L. "Über Die Entropieverminderung in Einem Thermodynamischen System Bei Eingriffen Intelligenter Wesen." Zeitschrift Fur Physik **53** (November): 840–56 (1929)

[17] Lewis, Gilbert N. "The Symmetry of Time in Physics." Science **71** (1849): 569–77 (1930)

[18] Brillouin, Leon, *Science and Information Theory*. Courier Dover Publications (1956)

[19] Cohen-Tannoudji, Gilles, *Universal Constants in Physics*. English language ed edition. New York: Mcgraw-Hill (1992)

[20] Frieden, B. Roy, *Physics from Fisher Information: A Unification*. Cambridge, U.K.?; New York, N.Y: Cambridge University Press (1999)

[21] Segal, Jérôme, *Le Zéro et le Un: Histoire de la notion scientifique d'information au 20e siècle*. Paris: Editions Syllepse (2003)

[22] Segal, Jérôme, "La Théorie de L'information existe-t-elle?" Pour La Science, no. fév.: 18–25 (2013)

[23] Battail, G. "Does Information Theory Explain Biological Evolution?" *EPL (Europhysics Letters)* 40(**3**) (1997)