The Birth of Information Theory

by Philippe Jacquet,
Ingénieur en chef des mines and Research Director at the Institute National de Recherche en Informatique et Automatique (INRIA)

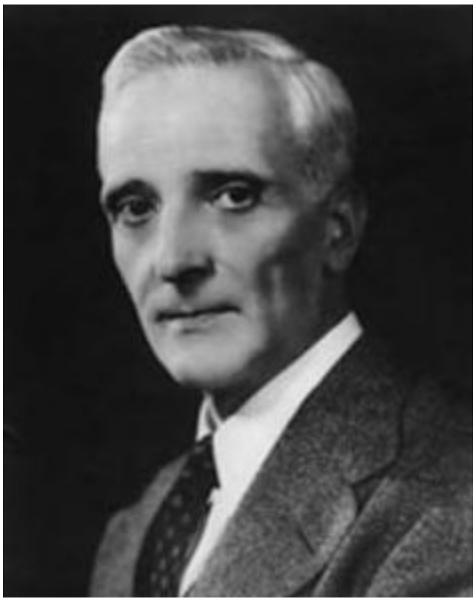


***Figure 1:* Léon Brillouin (1889-1969).**
Image American Institute of Physics.

LÉON BRILLOUIN IN THE 20TH CENTURY

Léon Brillouin was a French-American physicist and the author of major contributions in solid-state physics and in quantum physics. We owe to him the concept of "Brillouin Zones" which characterise the elementary volumes in crystal structures. Born of a scientific family, the son of physicist Marcel Brillouin[1], he entered the École Normale Supérieure in 1908. He finished his studies at the Institut de physique théorique in Munich under Arnold Sommerfeld. He returned to France in 1913 and began working on his thesis on "la théorie des solides et quanta" which he finished in 1920.

Like most young scientists at the start of the century, he had to put his work on hold for the sake of civilian mobilisation during the First World War. Thereafter, he contributed to the extraordinary development of quantum

1. Marcel Brillouin (1854-1948), student of the Ecole Normale, was a mathematician and physicist, member of the National Academy of Sciences and Professor at the Collège de France from 1900 (He was Chair of Theoretical Physics).

mechanics and also published work with Heisenberg, Bohr and Jordan. He was one of the French representatives at the famous Solvay conference of 1927 which brought together Einstein, Bohr, Marie Curie and many others[2]. In fact, his father attended the first Solvay Conference in 1911.

Owing to his expertise in wave propagation, he was appointed Director of *La Radiodiffusion française* just before World War Two. However, with the country confronted with the German Offensive, he ordered the systematic destruction of the network. At the risk of being accused of sabotage because of this action, he fled the Vichy Regime in late 1940 and immigrated to the United States; and there he began his American career. He fitted perfectly into the stimulating research environment there (University of Wisconsin, Brown University, Harvard University and Columbia University). In spite of his expatriate status, he was welcomed into the War Research Services where he worked on various radar improvements.

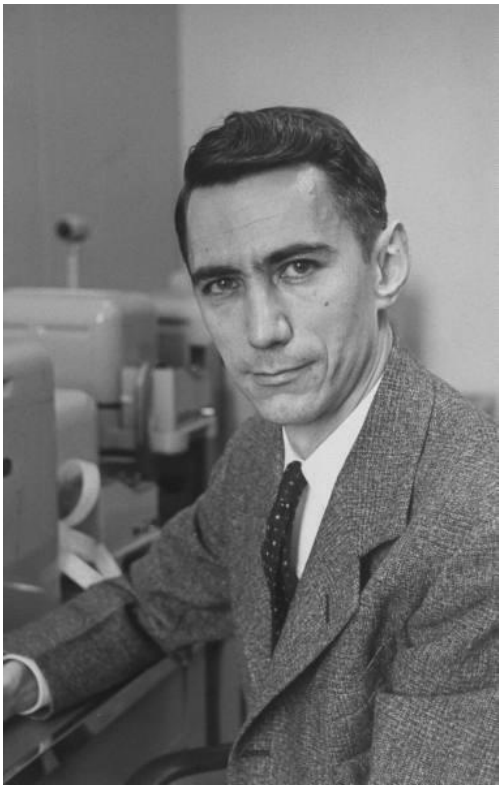


***Figure 2*: Claude Shannon (1916-2001).**
Image © Life Magazine, Alfred Eisenstaedt, 1951.

2. For the 1927 Solvay Conference and the Bohr's Complementarity Principle which was discussed there, see *BibNum* text (letter from Paul Ehrenfest) http://www.bibnum.education.fr/physique/le-congres-solvay-de-1927-petite-chronique-d%E2%80%99un-grand-evenement

In America, he was immediately impressed by the newfound Information Theory, formulated in 1948 by the mathematician Claude Shannon at Bell Labs. Being the good physicist that he was, he took delight in resuming and modelling the theory based on his own research. In 1956, he published *Science and information theory* which was translated into French in 1958 by Masson under the title "La Science et la théorie de l'information"[3]. This publication greatly expands Shannon's work towards electronics and physics. Brillouin was the first to really make the link between Shannon Entropy (which was for the latter, above all a mathematical definition) and Boltzmann's Statistical Entropy which is defined within physics. Furthermore, Brillouin laid the groundwork for a real "information theory" for matter and objects which he extended to other applications other than those on which Shannon had already worked (symbol encoding or symbol processing), in computer science itself, but also in other fields of science like physics, which he used to solve Maxwell's demon, or biology.

The title of the book – which seems to oppose "science" and "theory" – sets the tone of Brillouin's contribution. He comments on this in his preface:

> *The title of this book has puzzled some readers. Its meaning is twofold; Scientific Theory of Information on the one hand, but also, the application of Information Theory to pure science problems. In other words, there is action and reaction between Science and Information. Both of these reciprocal aspects are both essential.*

THE HISTORICAL CONTEXT

Let's look back at the historical context. Mankind had just emerged from the biggest armed conflict it had ever known. Spurred on by the U.S war effort, new telecommunication technologies rocketed such as the invention of the transistor[4] and, thanks to new automatic switches, we also saw the proliferation of the telephone.

The telephone, television, and radar were not new concepts in 1950.The surprise arrived in the form of an unexpected newcomer: the computer.

---

3. Léon Brillouin is described at the beginning of the work as "Honourary Professor at the Collège de France and Member of the National Academy of Sciences" (Washington, U.S.A.).

4. The first transistors were designed in 1947 at *Bell Telephone* labs by three researches who received the Nobel Prize in Physics in 1956.

Completely different from the automatic humanoids of the 18th century and robots that passed themselves off as humans such as those in *Métropolis*[5], the computer was a machine without form.

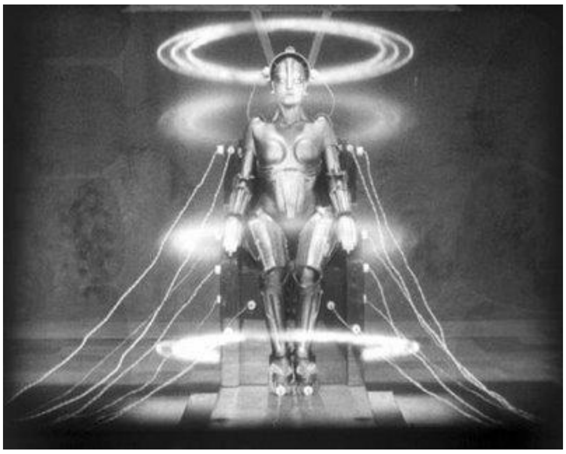


Figure 3:* The robot from Metropolis, *Fritz Lang's film (1927)

The very idea of this object not having any particular form and which could even be dispersed across a network in several indistinct forms was a cause for concern in a century where objects, such as aeroplanes, cars, ships and rockets were, above all else, forms and symbols of modernity. In 1956, American cars were heavily decorated with chrome whose only purpose was to evoke the ailerons of the US Air Force's new fighter jets which were, at the time, the ultimate of the avant-garde. In "The City and the Stars", published in 1956, Arthur C. Clarke[6] goes against public opinion at the time by describing an ultra-powerful central computer, assembled in a large room in a collection of silent white cubes which don't contain any moving parts. This latter idea also runs counter to the familiar memory of mechanical computers with loud vibrations emanating from their mobile parts. It also prefigures, 40 years in advance, multiprocessor computers using chlorofluorocarbon cooling and flash memory which are now replacing rotating hard disk drives.

However, a machine that could store and process much more information and at a faster speed than the human brain and whose language was that of pure mathematics was a real revolution for many people. If the start of the century had seen the triumph over matter and energy – both in transport, thanks to internal combustion engine, and in telecommunications, thanks to an

5. Metropolis is (1926) a science-fiction book by the German writer Thea von Harbou which was adapted to a silent film in 1927 by Fritz Lang.

6. Arthur Clarke (1917-2008) was a British author and inventor who also wrote "2001: A Space Odyssey" (1968).

understanding of radio waves –, the 1950s saw the emergence of complex electronic systems and the resurgence of mathematics in order to master them. Mathematics suddenly came back to the forefront after being reduced, for a long time, to a simple aid to gaining knowledge or something used by a mechanical engineer to perfect his work.

Scientific definition of information

Brillouin defines information. He avoids falling into the trap of giving a circular definition.

> *First of all, what is information? Let us look at Webster's dictionary: "Communication or reception of knowledge or intelligence. Facts, ready for communication, as distinguished from those incorporated in a body of thought or knowledge. Data, news, intelligence, knowledge obtained from study or observation…". We may state that information is the raw material and consists of a mere collection of data, while knowledge supposes a certain amount of thinking and a discussion organising the data by comparison and classification. Another further step leads to scientific knowledge and the formulation of scientific laws.*

Information Theory is not concerned with the contents or with the means of understanding information. At this level, Brillouin turns his back on programming theory which he had, nevertheless, gotten well underway with Turing. For the latter, the concept of information context, which contains the mechanisms for understanding information, also plays an essential role in compiling and compressing the information. Without recourse to the context and the "programme" object, it would be impossible to understand the source code of an image and the computer would not be able to display the image on the screen[7]. This view is not, however, inconsistent with Brillouin's main goal which is to eliminate any human factor in the concept of information to make it a physical quantity, just like mass and electric charge.

> *The methods of this theory can be successfully applied to all technical problems concerning information: coding, telecommunication, mechanical computers, etc. In all of these problems, we are actually processing information or transmitting it from one place to another, and the present theory is extremely useful in setting up rules and stating exact limits for what can and cannot be done. But we are in no position to investigate the*

7. Beyond Information Theory is "Complexity Theory" which was thrown open by Turing's work and then that of Kolmogorov on programming.

process of thought, and we cannot for the moment, introduce into our theory any element involving the human value of the information. The elimination of the human element *is a very serious limitation, but this is the price we have so far had to pay for being able to set up this body of scientific knowledge. The restrictions that we have introduced enable us to give a quantitative definition of information and to treat information as a physically measurable quantity. This definition cannot distinguish between information of great importance and a piece of news of now great value for the person who receives it.*

Brillouin also offers an impressive range of the scope of Information Theory. He mentions mechanical computers and, in fact, computers which he elaborates on later in the book. Univac, the first computer, completed in 1944 and computer science were taking off. In the chapter devoted to mechanical computers, Brillouin envisions the huge prospects of computing by quoting Poe.

> *(…) we can without difficulty conceive the possibility of so arranging a piece of mechanism, that upon starting it in accordance with the data of the question to be solved, it should continue its movements, regularly, progressively and undeviatingly toward the required solution (…)*[8]

By proposing to incorporate the computer within electronic circuits, assigning a "negative entropy" value to it in Information Theory, he prefigures the emergence of the digital age.

***Figure 4:* Univac, the first commercial computer (released in 1951)**

8. Edgar Poe, *Le Joueur d'échecs de Maelzel* (1836) in *Œuvres complètes,* (quoted by Brillouin in his work, page 269)

THE MATHEMATICS OF INFORMATION

Brillouin outlines some of the outcomes of the mathematical definition of information, in particular, the additive property of logarithms when two pieces of information are independent of one another (see the box below). Recently, some physicists have called into question the use of logarithms to measure information. At the level of elementary particles, they instead suggest a representation based on the square of probabilities which in their opinion is easier to make the hidden variables in paradoxical quantum systems visible, such as those of the intricate photons of the EPR paradox. However, the truth is that this quadratic representation has not been very successful because of its failure to move to the macroscopic scale.

The logarithmic law of information

In the first chapter of his book, Brillouin outlines the connection between information as defined by Shannon and probability:

We consider a problem involving a certain number of possible answers, if we have no special information on the actual situation. When we happen to be in possession of some information on the problem, the number of possible answers is reduced, and complete information may even leave us with only one possible answer. Information is a function of the ratio of possible answers before and after, and we choose a logarithmic law in order to insure additivity of the information contained in independent situations.

He tackles the problem, as we have said, primarily as a physicist:
Let us consider a situation in which P0 different possible things might happen, but with the condition that these P0 possible outcomes are equally probable a priori (…)

- *Initial situation:* $I_0 = 0$*, with* P_0 *equally probable outcomes.*
- *Final situation :* $I_1 \neq 0$*, with* $P_1 = 1$*, i.e. one single outcome selected*

He defines information I as logarithm of the probability modulo, a constant: $I_1 = K\ \mathrm{Log}(P_0/P_1) = K\ \mathrm{Log}\ P_0$. $I_1 = K\ \mathrm{Log}(P_0/P_1) = K\ \mathrm{Log}\ P_0$. They are two independent problems, one with P_0 equally probable outcomes, and the other with Q_0 equally probable outcomes. The number of possible initial states[9] coming from these two sets of data is P_0Q_0. So we have $I = K\ \mathrm{Log}\ (P_0Q_0) = K\ \mathrm{Log}\ P_0 + K\ \mathrm{Log}\ Q_0 = I$ (system p) + I (system q). The independence of the two initial

9. Don't forget, to get a better understanding of the way in which Brillouin sets out his problems, that he is a quantum physicist. Also recall the definition given by Boltzmann of entropy as proportional to the logarithm of the number Z of microscopic states of the system S = k log Z, where k is Boltzmann's constant.

systems leads to multiplicative probabilities and to the additivity of information.

For the contained information, Brillouin gives the striking example of card games: if we choose a card from a pack of 32, and we choose the constant adequately to use the binary system, we can write $I = Log_2 P$. For 32 cards it is, $I = Log_2 32 = 5$ bits. He uses the example of two packs of 32 cards, one card drawn from each pack, the number of states is $PQ = (2^5)^2 = 2^{10}$. So we have I = 10 bits.

The choice of one card from 32 cards can be "coded" on 5 bits, so the choice of two cards, one from each of the two packs, is coded on 10 bits. Note that this last problem is different from that of choosing one card from two packs of mixed blue and red cards which is coded on 6 bits (5 bits from the card and one bit from the colour of the pack).

THE PHYSICS OF INFORMATION AND MAXWELL'S DEMON

From a physics point of view, Brillouin also highlights the connection between Boltzmann's entropy and an amount of information. If Shannon used the term entropy to measure an amount of information, it is simply because the two concepts coincide in their mathematical definition[10]. There is a great difference between considering the statistics of words in a message and considering that of atomic states of a gas. However, Brillouin overcomes this and he forces both concepts together all the way through the book. For him, there is no difference between physical entropy and entropy of information as to obtain information on a system, it must be measured. This measurement increases the entropy of the universe by an amount exactly equal to the amount of information obtained. In this way, Brillouin establishes an absolute scale for the measurement of information and to this end he creates a new quantity: the negentropy, or negative entropy. His reasoning is reciprocal; information creates entropy, but entropy creates information. The universe provides information on itself every time that it makes entropy, which it does continuously given the second principle.

10. Léon Brillouin pays tribute, many times in fact, to the pioneering work of the physicist Léo Szilard in his work [L. Szilard, *Z. Physik,* 53, 840 (1929)] ; the work of Shannon and Weaver was published in 1940 [*The mathematical theory of information,* University of Illinois Press].

Every physical system is incompletely defined. We only know the values of some macroscopic variables, and we are unable to specify the exact positions and velocities of all the molecules contained in a system. We only have scanty, partial information on the system, and most of the information on the detailed structure is missing. Entropy measures the lack of information; it gives us the total amount of missing information on the ultramicroscopic structure of the system.
This point of view is defined as the negentropy principle of information, and it leads directly to a generalisation of the second principle of thermodynamics, since entropy and information must be discussed together and cannot be treated separately. This negentropy principle of information will be justified by a variety of examples ranging from theoretical physics to everyday life. The essential point is to show that any observation or experiment made on a physical system automatically results in an increase of the entropy of the laboratory. It is then possible to compare the loss of negentropy (increase of entropy) with the amount of information obtained to the associated increase in entropy. The efficiency of an experiment can be defined as the ratio of information obtained to the associated increase in entropy. This efficiency is always smaller than unity, according to the generalised Carnot principle.

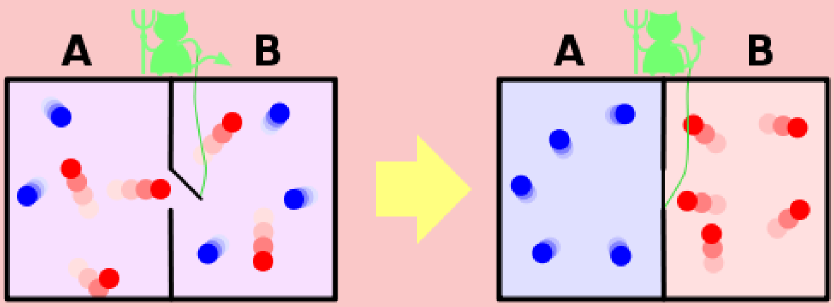
Some physicists have gone much further. However, if truth be told, they have done so with very little success. John Archibald Wheeler went as far as to compare the universe with a huge computer that is programmed to provide information to all of its conscious observers. He summarises his example by his famous phrase "it from bit[11]". According to him, the universe invents its laws of physical as they are observed and tested. However, if truth be told, these audacious views remained largely without echo in the scientific community; the idea that Nature calculates instead of reacting to physicals laws was not very convincing. In fact, new insights into the functioning of the brain, the object of nature which is most like the computer, have shown that the former is not completely organised like the latter with its processors, circuits and separate memory. The brain is said to be rather, an object containing complex but soft wiring and based on purely analogue functions. So, if humans aren't exactly computers, why would the universe be one?

11. J.A. Wheeler (1911-2008), American theoretical physicist, worked with Einstein at Princeton on the search for a "Unified Theory of Physics". With his allegory "it from bit", he assumes that the laws of physics are ultimately all derived from the theory of binary information.

For the moment, theoreticians are sticking to a fairly strict separation of the physical aspects of usable information which is considered an extremely weak part of the entropy of the universe due to being mainly composed of sources of random signals. Nevertheless, it would be very interesting to pursue further research in order to know the exact nature of the boundary between information and pure entropy. Nowadays we are starting to speak of "moles" of information (6.02×10^{23} bits which is just a bit less than 2 to the power of 79) thanks to the constant progress of memory capacity. The stakes are high but is it possible that one day we will be able to store information in every molecule of a given volume of gas? Will we be able to store information in the intimate structure of matter and of space-time?

Maxwell's demon as seen by Brillouin

Brillouin bases this conclusion on his solution to "Maxwell's demon" which he develops in his book. Maxwell's paradox involves maintaining a partition, which can be opened and closed, between a compartment of hot gas and a compartment of cold gas. There is a demon which only allows the most energetic molecules from the cold compartment to pass into the hot compartment and only the least energetic molecules of the hot compartment to pass into the cold compartment. By doing so the demon, without using up energy, helps to increase the temperature of the hot compartment while lowering the temperature of the cold compartment which contradicts the second principle of thermodynamics.



Léon Brillouin, to use his own expression, "exorcised the demon" by correctly pointing out that the sorting of molecules actually requires the acquisition of information on the energy level of each molecule and that this sorting of molecules uses up some energy because of the negentropy which has been thus created. This restores the validity of the second law. The result of the solution to Maxwell's paradox is that it is possible to give a minimum value of the entropy developed to collect and store a bit of information. It is a question of dividing The Boltzmann Constant by the logarithm of two which is roughly 2×10^{-23} joule per Kelvin. Below this value, Maxwell's demon could heat up the hot compartment and cool down the cold compartment which would violate the second law.

INFORMATION IN BIOLOGY

To conclude, Brillouin considers the promising aspects of Information Theory. He does mention biology but he places great emphasis on physics and, more particularly on Maxwell's Demon. Brillouin is vaguer regarding biology which is probably due the fact that he was a trained physicist. However, the transmission of genetic information – the genetic code – is fundamental in the organisation of life and had already been discovered at the time when Brillouin was writing his book. After Fred Griffith's experiments in 1932, Oswald Avery showed in 1944 that DNA transmits hereditary characteristics. In his book "What is Life?" published in 1944, and which Brillouin had read, the physicist Erwin Schrödinger wrote: "Chromosomes contain, in some kind of code-script the entire pattern of the individual's future development and of its functioning". Schrödinger's idea, fifteen years before the publication of Brillouin's book, was confirmed by the discovery of the double helix by J. D. Watson and F. Crick in 1953.

> *Altogether we do hope that the scientific theory of information marks just the beginning of a new and important chapter of scientific investigation especially in physics and also in biology. There is already an appreciation that this new theory has made it possible to gather and regroup many unrelated facts, especially on the essential definitions and measurements of physics. The new theory also has the advantage of strengthening the position of statistical thermodynamics and of eliminating many paradoxes, such as Maxwell's demon.*

CONCLUSION

The development of Information Theory is one of the revolutions which brought about the all-digital era. We are still living this revolution and even if our daily life has already radically changed, we haven't yet explored all of its possibilities and it is still difficult to get an overall picture of the subject.

However, the positive effects are undeniable. Some see in the Information Age and the Digital Telecommunications Age the solution to environmental problems that industrialisation has caused our planet in view of the mass usage of fossil fuels and the development of mass transit.

With regard to Brillouin, he brought a bold scientific vision and often very spot-on perspectives, which is still the case even today with 50 years of hindsight. His work is more than a simple popularisation of Shannon's theory; he extends the thought of the latter and his predecessors. He offers a perspective on all the modern achievements which will be the result of this new mathematics.

The result is a strong, alive and passionate vision. If some of the points developed by Brillouin still remain unresolved and even if others will be discarded along the way, his vision will, nevertheless remain a stimulating source of inspiration for research for a long time.

■

(September 2009)

(translated in English by Lauren Gemmell, published September 2013)