

---

# The First Information Theories

---

## 1.1. Introduction

Information<sup>1</sup> has been at the heart of scientific considerations for almost 70 years now. The importance given to information in our societies keeps growing, together with the versatility of the concept and the lack of rigor in its definition. From the texts of its forerunners to the most recent publications, all agree on this point. Jean-Louis Le Moigne [LE 73, p. 10] expresses it in the following words: “Information! Is there a more familiar, a more intuitive word? Is there even a more international term? Is it not central to every management conversation? At the heart of the act of Decision-making, do we not find the immediate answer: Information? Is it not the source of this mutation that human societies are currently experiencing – not without emotion – under the name of the computer revolution? And yet, is there not a more difficult, a more multifaceted, a more ambiguous word?”

Or Jean-Paul Delahaye [DEL 94, pp. 13–14]: “The word information is used in a variety of phrases and contexts. For example, we usually say things like: “The information contained in this book, the available

---

<sup>1</sup> This chapter resumes, updates and enriches the ideas introduced in Chapter 5, pp. 69–83 of *La conceptionen communication. Méthodologie qualité*, published by Hermès Lavoisier in 1997 [LEL 97].

information we have regarding a problem, the information encoded in the genome, the poor information that his long speech provided”. [...] We sense that behind this word is hidden something complex, something changeable perhaps, something that, in any case, deserves our reflection. Thus, we are led to wonder: is it possible to make a general scientific theory of information? And if so, how to do it? It is not easy to answer seriously and it is very easy to answer badly, because some mathematical or physical theories already employ the term information, enunciate theorems and give the impression that the problem has been solved and that we can mathematically speak of information”.

The absence of scientific rigor in the characterization of the concept is absolute. We can clearly perceive this by taking a look at the definitions that appear in the dictionary:

- the action of informing, of giving information ACTION;
- the news, the information that is communicated about someone or something, STATE;
- the body of knowledge acquired regarding someone or something, A SET OF STATES;
- the actual contents of transmitted messages, CONTENT;
- a signal by which a system transmits a piece of knowledge, CONTAINER.

Unfortunately, there is no more rigor in the Sciences index of the term’s definition (reference: Petit Robert):

- an element or system capable of being transmitted by a signal or a combination of signals;
- what is transmitted, the object of knowledge or memory.

Let us observe that the word “system” frequently appears in most of the scientific definitions of the term.

The perspective of our work intends to be that of a scientific approach towards information. In this sense, we will only take into consideration the works devoted to:

- the modeling of the information processes;
- the study of the operational laws ruling the functioning of these processes;
- more or less formalized and more or less quantified proposals of abstract representations, associated with the corresponding phenomena.

We will purposefully exclude the principle of a study linked to a specific field of application or to a specific category of practices.

## **1.2. The mathematical theory of information by Shannon [SHA 48]**

Research carried out by the pioneers in the field of “information theory” had a variety of destinies. There is no doubt that the works of Claude Elwood Shannon almost immediately found their application in the field of telecommunications: this contribution arrived at its own pace, particularly after the research conducted by other signal engineers from Bell Telephone Laboratories. Nowadays, these works constitute the core of the results unanimously recognized as scientific (in the Cartesian sense of the term, and even in its strictly mathematical sense) as regards a theory of information.

### **1.2.1. *Beginnings of this theory***

The starting point can be presented in almost naïve terms. The more improbable and uncertain an event, the more information concerning its advent will be significant: the amount of information in a message depends on the improbability of the event that the message informs us about.

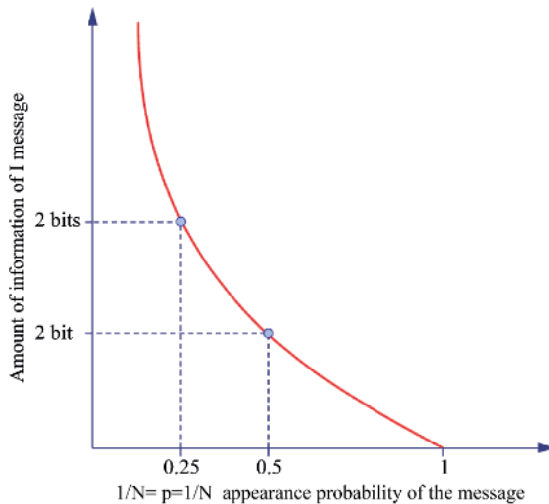
To support this hypothesis, we will illustrate it with a simple intuitive example. Let us imagine that, for decades, all your family members have been without any news from uncle Sam, who left a long time ago to lead an adventurous life in unexplored territories. A message announcing uncle Sam's arrival the next day contains a lot of information because, given his long absence and prolonged silence, the probability of a message announcing this specific event is extremely low; in fact, such an occurrence is close to zero. Besides, it is precisely due to the fact that the event itself is perceived as improbable that the probability of the message announcing it is feeble. Therefore, we can perceive that in this case there is confusion between the probability of the event taking place and the announcement of the event, a matter that we will discuss at further length. On the other hand, let us imagine that uncle Sam sent a letter that the post was unable to deliver promptly – which does sometimes happen. If this same message, by the same means, arrives one day after the reappearance of uncle Sam, it contains no more information, because the event has already taken place; so it is no longer improbable, it has become certain. As a matter of fact, there was a time when this kind of inconvenience was commonplace; for example, in *Tristes Tropiques*, Claude Lévi-Strauss observed: “Since the ‘official and urgent telegram’, sent from Lahore on the previous day in order to announce my arrival, reached the director only five days later, due to the floods that raged the Punjab, I might as well have come impromptu” [LÉV 55, p. 473]. Thus, we reckon that the informational estimate of the same message, in an identical form, can vary from one extreme to the other in a short time, depending on the probability of occurrence of the event to which the message makes reference.

Starting from this observation, Shannon's theory establishes a biunivocal relation between the amount of  $I$  information and the probability of occurrence of a message or, more precisely, the number of  $N$  states that the expected message can eventually adopt.

This relation takes the following form:  $I = K \cdot \log N$ ,  $K$  being constant.

$p = \frac{1}{N}$  being the appearance probability of one of the  $N$  possible states, these  $N$  states being equally likely, the above relation can also be expressed as:

$$I = K \cdot \log \frac{1}{p} = -K \cdot \log p$$



**Figure 1.1.** Relation between the amount of information of a message and the appearance probability of such a message

This curve represents the relationship between the amount of information and the probability of occurrence of such a message.

### 1.2.2. Shannon's generalization

The preceding definition of the measurement of the amount of information is filtered by the restrictive hypothesis of the equiprobability of possible states. One of Shannon's key contributions lies precisely in the fact that he suggested such a generalization. In fact, all possible states no longer have the same probability of occurring, and then, when approaching the limit, the distribution of

probabilities assumes a continuous form. By doing so, Shannon associates an amount of information with an amount of entropy.

In this way, we get a measure of the information of a message considered from the viewpoint of the appearance probabilities of the message, which assumes the very general form:

$$I = -\sum_{i=1}^n p_i \cdot \log p_i \quad \text{with} \quad \sum_{i=1}^n p_i = 1$$

$p_i$  being the appearance probability of one of the  $n$  states:  $N_1, N_2, \dots, N_i, \dots, N_n$ .

Nevertheless, in its generality, this measure still has some limits, mainly associated with the notion of probability. In fact, who are these states likely or unlikely for? For an objective, statistical receiver or for a manager who often estimates a probability via a legitimately subjective judgment? The manager using this notion will often reach the limit. If he does so consciously, he will be right.

Therefore, we should always bear in mind that the distribution of probabilities assumes a continuous form when it approaches the limit.

### **1.2.3. Information and entropy**

The theory of systems has given great importance to the notion of entropy, a measure for uncertainty, disorder, for diversity. We can briefly summarize the evidence of the negentropic equivalence of information as follows:

Given an initial situation, about which we know nothing ( $I_0 = 0$ ) and *a priori* characterized by  $N$  equally probable situations,  $I_1$  information ( $I_1 > 0$ ) makes it possible to reduce the number of equally probable alternatives from  $N_0$  to  $N_1$  ( $N_1 < N_0$ ).

The evolution of the physical entropy of this system is measured by:

$$S_0 - S_1 = k \cdot \log N_0 - k \cdot \log N_1 \text{ with } S_0 > S_1$$

Hence, we have seen that the information about a situation is measured by:

$$I = k \cdot \log N$$

with an adequate choice of units. Then:

$$S_0 - S_1 = I_0 - I_1$$

And because  $I_0 = 0$ :

$$I_1 = -(S_0 - S_1)$$

If we define negentropy as the negative of physical entropy, we obtain the amount of information supplied to the system, which equals the corresponding increase in system negentropy.

### **1.3. Kolmogorov's algorithmic theory of information [KOL 65]**

A little less than twenty years later, appeared another theory of information, known as the "algorithmic theory of information" or "Kolmogorov's theory of information" [KOL 65, LI 93]. It is sometimes introduced as a substitute for Shannon's theory.

#### **1.3.1. Succinct presentation**

The initial idea is to define the complexity of an object according to the size of the smallest program capable of engendering it [CHA 77].

When Turing [TUR 36] established that certain mechanisms had a maximum calculating power, that is to say, as soon as the notion of a universal machine became available, the idea of using it as a universal measure of complexity naturally followed. This was almost simultaneously discovered by Minsky [MIN 62], Chaitin [CHA 66] and Kolmogorov [KOL 65], whereas it is to Solomonoff [SOL 64] that we owe the first technical formulation mentioning the notion of a universal Turing machine.

### **1.3.2. First algorithmic information theory**

Kolmogorov's notion of complexity is defined by the concept of information content of finite objects (for us, sequences of 0 and 1), itself formulated with the help of the notion of a universal Turing machine.

A universal Turing machine is a Turing machine capable of simulating all Turing machines. By definition, the information content of the  $s$  finite sequence, or Kolmogorov's algorithmic complexity of  $s$ , written as  $K(s)$ , corresponds to the size of the smallest program (for a universal Turing machine) capable of producing  $s$ .

We show that, by difference of an additive constant, this notion is independent of the universal Turing machine used in the definition. In other words, if  $K_U(s)$  refers to the algorithmic complexity of  $s$  obtained by using the universal Turing machine  $U$  and  $K_V(s)$  to the one obtained with  $V$ , there is a  $C_{UV}$  constant, which solely depends on  $U$  and  $V$ , in such a way that for any  $s$  finite sequence:

$$K_U(s) - K_V(s) < C_{UV}$$

This result receives the name of invariance theorem.

We immediately show that there exists a  $C'$  constant, so that for every  $s$  finite sequence:

$$K(s) < (s) \text{ length} + C'$$

The idea this result proves is very simple: for every  $s$  sequence, there is a program producing  $s$  whose meaning is simply “to print  $s$ ” and whose length is equal to the length of  $s$  plus what is needed so as to express the copy algorithm, which has a  $C'$  length. This relation defines the size of the smallest program capable of producing a finite sequence.

### 1.3.3. *Second algorithmic information theory*

The initial expression of Kolmogorov’s information theory needed to be corrected in function of the expression of the end of the program.

In fact, the above definition is only possible in case the end of a program does not need to be indicated on the program itself. The end of the program is reached when all its digits have been read. The second theory of algorithmic complexity requires that the end of programs are specified on the programs themselves: we will then speak of self-delimiting programs. Nowadays, this is the case for virtually all computer programming languages.

The consequences of this requirement are:

- that no program can be extended to another one, and therefore, each program can be specifically weighed;
- that the preceding inequality becomes:

$$H(s) < (s) \text{ length} + 2 \log ((s) \text{ length}) + C'$$

Indeed, the copy program must contain the information regarding the length of  $s$ , besides the description of  $s$  digit by digit.

In any case, the information content of a finite sequence is always expressed through the size of the smallest program capable of producing it.

## 1.4. Delahaye's further developments [DEL 94]

Nevertheless, even if these information theories (respectively mathematical and algorithmic) are universally acknowledged and continue to be professed all over the world, they do not cover all the problems related to information.

### 1.4.1. Gaps

In his book *Information, complexité et hasard* – an essential work that we have already quoted and which constitutes the main source of the presentation above – Jean-Paul Delahaye is rather reticent concerning the merits of these multiple theories. “Neither Shannon’s theory, nor Kolmogorov’s theory, nor any other theory, states everything about what information is. What is more, while claiming to be *the* information theory, each of these prevents progress. Information must be conceived in general and diverse ways which are unlikely to be perused in haste and for which, until now, no mathematical theory provides the ultimate secret. Along the way, we will ponder the real interest of the theories mentioned” [DEL 94, p. 14].

Thus, the author suggests further extensions that address and go a step beyond the limitations he observed.

### 1.4.2. Information value

First, it is necessary to go beyond mere information contents and substitute this with the notion of information value.

Let us consider a sequence of  $S$  symbols, for example, a string of characters. The information content of the various objects of this nature has or had a certain variable value, at some point in the past. They were bought and sold accordingly; it was possible to invest in order to produce them or to continue spending large sums in order to keep them. Their assigned value at the exchange market was a testimony of their inherent value, in this case, expressed in a monetary equivalent. “A general theory of information which does not praise the informational value of one of these objects can only be incomplete or

inadequate. An information theory which does not provide the means for comparing the informational content value of these objects, cannot claim to be recognized as *the* theory of information” [DEL 94, p. 16].

For this reason, Delahaye suggests considering information value as a complement to its sole “raw” contents.

### **1.4.3. Raw information content**

The raw information content of each object can be understood as a weight estimate in a certain coding system. The corresponding measurement unit in a digital environment is the bit and its multiples (bytes, kilobytes, megabytes, gigabytes, terabytes, etc.). We can define this measure of raw content as the space that the stored object digitally occupies in the memory of a computer, when it is not subjected to any particular treatment other than the formatting process compatible with the computer’s operating system. Moreover, the identifier of any file recorded in a digital database is nowadays systematically accompanied by a weight measure – that is to say, its amount of information or raw content – expressed in bits.

“It is clear that the raw information content does not determine the value of the information. This is evident: the value of information is something more complicated and relative. And it is because the information value of a string is relative to a certain purpose and to a certain situation that there are several information theories and that not a single theory can really cover all the problems that the notion of information elicits” [DEL 94, p. 16].

Given a certain goal, that we will call goal B, this is written as  $Val(S,B)$ , that is to say, the value of information contained in  $S$  regarding goal B.

In the particular case where we set the very specific goal of compressing the  $S$  character string as much as possible, assuming that we have an  $M$  machine for this, then the information value of  $S$  is the length of the smallest program available (expressed in binary terms), which enables it to reconstitute the  $S$  string when given to  $M$ . In this case, information value is the incompressible content of  $S$  (as regards

M). Consequently, the notion of information content according to Kolmogorov corresponds to the general definition of information value  $\text{Val}(S,B)$  when goal B is to:

B = (compress for universal machine M)

Thus, only the assignment of a goal makes it possible to measure the value relative to this goal. Evidently, there exists an infinity of purposes other than compressing for a universal machine. Bennett's notion of logical depth [BEN 88] corresponds to the general definition of information value  $\text{Val}(S,B)$  when goal B is to:

B = (limit the number of calculation steps necessary to retrieve S as much as possible).

This last value can be seen as a computational value. However, beyond the technical goals of compressing or calculating the S string itself, in everyday life a great number of goals are defined in a pragmatic way.

#### ***1.4.4. Pragmatic aspects related to information value***

We still have to mention certain new limitations to the formulations that have been brought forward as information theories. "If we consider that the targeted aim is of a practical nature, for example, to survive in a certain environment or to make as much money as possible on a certain day at the Paris stock exchange, then the value of the information contained in a string of characters S will be measured according to the intention pursued. For example, precious information will be which place to attend to obtain certain types of food or the name of the stock market that must be bought because its price will increase. There is no general information theory that takes into account all the pragmatic aspects which determine the value of a string of characters" [DEL 94, p. 20].

Jean-Paul Delahaye concludes that: "The mathematical theories available or still in development which concern information claim a degree of universality and of applicability which we have to be wary of" [DEL 94, p. 27].

In the present state of scientific formalization, the assessment of the value of information – although more essential than measuring quantities or the raw contents of information – is still indefinite.

## 1.5. Final remarks

This section outlined the very first theories recognized as scientific approaches to information. The first theory, that of Shannon [SHA 48], is mathematical. The second one, by Kolmogorov [KOL 65], is algorithmic.

These theories share the feature that they disregard the semantic dimension of information in favor of a measurement of raw content or amounts of information. They do not contribute to estimating the value of this type of information. Regardless of their real interest, the two theories have often been misinterpreted, maybe due to misunderstanding, caricature, generalization or simplification beyond what is reasonable. In order to establish its fundamental notions and supporting paradigms, the following chapter will revisit Shannon's theory, by revealing it in an intuitive and playful way so as to fully grasp its subtleties and nuances.

