

Federico Flueckiger

Contributions Towards a Unified Concept of Information

Zdroj: <http://mypage.bluewin.ch/federico.flueckiger/Uci/Download/Ctrb2uci.pdf>

[upraveno]

2.1.2 Claude E. Shannon's statistical model

Wiener's idea of a statistical theory about the amount of information was realised by Claude E. Shannon in [Shannon 1969] {Footnote⁽⁴⁾: [Shannon 1969] is the fourth edition of the work first published in 1948.}. The basic problem that needed to be solved according to Shannon was the reproduction at one point of a message produced at another point. He deliberately excluded from his investigation the question of the meaning of a message {Footnote⁽⁵⁾: In chapter 2 'meaning' designates the extension of a message, i.e. the reference of the message to the things of the real world. An exception to this is chapter 2.3.3, where the deviation from this usage is explicitly noted. A detailed analysis of the concept of meaning in particular and the theory of semantics in general with reference to our concept of information will follow in chapter 3.}, arguing that:

"Frequently the messages have meaning; that is they refer to or are correlated according to some system with physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design." [Shannon 1969, p. 31]

According to Shannon, in order for messages to be transmitted, we need a communication system. By this he means an arrangement such as is represented in Figure 1.

Shannon describes the components of the communication system as follows:

1. The information source (or message source) produces messages or parts of messages intended for a particular destination.
2. On the basis of the message, the transmitter produces a sequence of signals such that they can be transmitted over a channel.
3. The channel is merely the medium used to transmit the signal from transmitter to receiver. During transmission the signals may be perturbed and distorted by a so-called noise source.
4. The receiver usually performs the reverse operation to the transmitter, reconstructing if possible the original message from the signals.
5. The destination is the addressee of the message and can be either a person or a thing. It requires a priori knowledge about the information source which enables it to

understand the message transmitted. In any case, the destination must know the set of signs available to the information source.

On the basis of his communication system, Shannon defines statistical quantities such as channel capacity and amount of information for discrete and continuous, noisy and noiseless systems. For the purpose of the present thesis, only the derivation of the amount of information for discrete noiseless information sources is relevant:

- Channel capacity is irrelevant for a definition of information. It is not of interest how much information per time unit can be sent over a channel, but only the fact that information has been sent and what it consisted of.
- Except for degenerated cases, continuous behaviour can be simulated to any desired degree of precision by means of discrete models.
- The noise source can be represented as an additional source of information which informs (or misinforms) the channel and thereby the information transmitted.

Shannon assumes that the set of messages which a discrete information source can produce consists of a finite number of elements N_1, N_2, \dots, N_n (the elementary messages or signs), with a probability of occurrence of p_1, p_2, \dots, p_n respectively. Since under very general conditions an information source combines the messages from sign sequences produced in a non-deterministic way, the information source can be interpreted as a stochastic process {Footnote⁽⁶⁾: More detailed descriptions of stochastic processes and their special cases will be found in [Heller et al. 1978], [Ehrenstrasser 1974] or [Beyer et al. 1978].}. In many cases, such as the transmission of a text in a natural language, the probabilities moreover depend on the preceding states. Thus in English the letter 'd' is more likely to be followed by the letter 'e' than the letter 'z'. In such cases the information source operates as a discrete Markov process (a special case of a stochastic process) or Markov chain respectively. Where transition probabilities converge on a particular probability distribution independently of the starting probability, such as is typically the case with information sources that produce texts in a given language, it is usual to talk of an ergodic Markov chain.

The ergodicity of transition probabilities can be illustrated by imagining that an information source can be brought to produce texts automatically and without human input [Shannon 1969, p. 43 ff.]. If digrams {Footnote⁽⁷⁾: By a digram one understands the fact that the occurrence of a sign depends not only on its own probability of occurrence but also on the sign preceding it.} are used, the resulting sequence will already differ greatly from a sequence produced using only normal probabilities of occurrence. If trigrams or indeed tetragrams {Footnote⁽⁸⁾: For trigrams the two preceding signs and for tetragrams the three preceding signs determine the occurrence of a sign.} are used for generating text, it is possible to produce almost meaningful text.

Apart from the requirement that abstraction be made from form and content of the information, the measure for the amount of information - called $H(p_1, p_2, \dots, p_n)$ by Shannon - should have the following properties:

1. H is continuous in p_i .
2. If all p_i are equal ($p_i = 1/n$), then H must increase monotonically with n .
3. If a choice is broken down into two successive choices, the original H should be the weighted sum of the two new values of H .

In the form of a theorem, Shannon postulates in [Shannon 1969, p. 49 f.] that the only function satisfying the three criteria above is of the following form:

Formula 1

where K is a normalising positive constant. Shannon shows in [Shannon 1969, p. 87 ff.] that Formula 1 is also valid for the more general case of real p_i .

With this formula Shannon created a direct equivalent to entropy as defined in statistical mechanics and thus contradicted Wiener's assertion that the amount of information was the same as negative entropy. The debate about the minus sign continued for quite a while until Shannon's formula was generally adopted, probably because it was sufficiently formalised so that it could be applied to practical cases immediately.

The amount of information H from Formula 1 is a quantity which allows us to calculate the average number of signs (from a set of signs x_1, x_2, \dots, x_n) needed for coding a message. Since the resulting value is not always an integer, Shannon talks about the measure of information H^* of the coding operation where: $H^* \geq H$. The resulting superfluous information, redundancy r , is calculated as follows:

Formula 2

Shannon's theory of the amount of information was dismissed by many later authors as merely a theory of syntactical information, because it excluded the semantic and pragmatic levels. In particular, Yehoshua Bar-Hillel and authors drawing on his work, such as Doede Nauta jr., vehemently attacked Shannon's theory. They said that, except for the theory about the amount of information of an information source, considered irrelevant by this school of thought, information as a notion was not defined by Shannon at all (cf. [Bar-Hillel 1964, p. 301]). Moreover, they claimed that physical entropy as a purely empirical concept was unsuitable for a definition of semantic information, which for them was a logical concept (cf. [Bar-Hillel 1964, p. 309]). Any attempt to develop Shannon's theory into a universal theory of information would necessarily reach an impasse. To this day, though, there is no quantity in information theory which is as well-supported and as generally accepted as Shannon's amount of information. On the other hand, Shannon's work is rightly seen as lacking indications for a conceptual clarification of information.