Atti del XLIV Congresso Nazionale Sisfa - Firenze 2024 (2025), **43**_10, pp. 89–94
Doi:10.6093/978-88-6887-352-3-10

From Shannon to von Neumann: A Partial Understanding of Information

Niccolò Covoni[1]

[1]Università degli Studi di Urbino Carlo Bo, Urbino & Università della Svizzera Italiana,
n.covoni@campus.uniurb.it.

Abstract: The talk aims to analyse the definition of information by comparing the crucial definitions formulated by Shannon and von Neumann and highlighting some of their criticalities. Statistical mathematics poses the basis for understanding information in the form of Shannon's proposal in A Mathematical Theory of Communication. In this work, information is explained by the concept of information entropy. However, it creates confusion in understanding what information is, first because of its comparison to the thermodynamic counterpart and second because Shannon's definition is very precise, as it analyses entropy as the quantity of uncertainty in a transferred message. Instead, the advent of quantum mechanics also posed a turning point in the definition of information and, more precisely, in the exchange of information. Von Neumann proposed a new definition that, at first glance, seems pretty close to the definition of Shannon; in fact, the shape of both formulas is very similar, and the von Neumann entropy seems to be the same equation of the information entropy but for the microscopic world. However, the von Neumann definition captures something different: the entanglement property. The two equations coincide in some situations, but the same applies to Shannon entropy and thermodynamics. We will investigate if the definitions are specular. The seminar will be structured as follows: in the first part we will analyse Shannon's definition; in the second part we will discuss von Neumann's definition; in the last section we will compare them using recent works on this topic.

Keywords: Information, Physics, Entropy, Philosophy of Science

1. Entropy

The Boltzmann Demon is the easiest way to think and understand the concept of entropy. It reflects the number of possible microscopic arrangements (or states) the demon can create with the particles in a system while still achieving the same macroscopic state. All attempts to define information use the concept of entropy, so a brief analysis was necessary before looking at the various proposals for defining information. For this reason, the paper has the following structure: the first paragraph is an overview of the concept of entropy, and paragraphs §2 and §3 analyze the proposals given by Nyquist (1924) and Shannon (1948) to define information, instead the paragraph §4 shows the von Neumann concept of information entropy in quantum mechanics. At the end of paragraph §5, the paper aims to show the conclusion that the attempts to define information cover different aspects and we lack a unified definition of information.

Entropy appeared in the middle of the 19th century in the context of thermodynamics, proposed by Rudolf Clausius (Hanlon, 2020). At first, it was considered a measure of disorder and the grade of irreversibility of a physical process. The language of thermodynamics measures the quantity of energy in a system that is not available to make a work in this way:

$$dS = \frac{\delta \overline{Q}}{T} \tag{1.1}$$

$$\frac{dS}{dt} \geq 0 \tag{1.2}$$

The first equation is the Clausius definition of entropy. It describes the amount of heat $\delta\overline{Q}$ transferred from a system at a given temperature T. The equation states that entropy changes proportionately to the amount of heat added to the system and inversely to the temperature. Instead, equation (1.2) expresses that entropy always tends to increase, which means that entropy can never decrease over time.

In 1877, Ludwig Boltzmann developed the statistical explanation of Clausius's second law of thermodynamics. He formalized the concept of entropy S, defined as the ratio of heat flow to temperature. The law states that the entropy for a closed system (with constant energy, volume, and number of particles) can never decrease (Boltzmann, 1877).

$$S = k_B ln\Omega \tag{1.3}$$

K_B is a physical constant that relates temperature to energy at the particle level. The number of distinct microscopic configurations, or "microstates", that a system can be in, given its macroscopic constraints, is represented by Ω. As Ω increases, there are a greater number of possible ways of arranging the particles while preserving the system's overall macroscopic characteristics.

However, "the theory of thermodynamics, taken by itself, does not connect entropy with information. This only comes when the results are interpreted in terms of a microscopic theory, in which case temperature can be interpreted as being related to uncertainty and incoherence in the position of particles" (Bais & Farmer, 2008, p. 7).

It is crucial to acknowledge that although entropy was initially conceptualized within the domain of thermodynamics as a measure of disorder, its mathematical formulation established the foundation for subsequent advancements in information theory. The interconnection between entropy and information became evident only when researchers began to interpret entropy in terms of uncertainty and probabilistic processes.

2. Nyquist: a first attempt

One of the 'prototype' attempts to define information can be found in the work of Harry Nyquist, *Certain Factors Affecting Telegraph Speed*, published in 1924. The scope of this paper is to work on some troubles that affect telecommunication systems. Here, Nyquist defines information as the token that carries itself (Binary or Signal Elements).

Aside from the technical results, one of the most exciting concepts of the paper is the 'speed of transmission of intelligence.' It can be defined as "the number of characters representing different letters and figures that can be transmitted in a given length of time, assuming that the circuit transmits a given number of signal elements per unit of time" (Nyquist, 1924, p. 333) Nyquist also proposes a mathematical formulation of the speed of transmission of intelligence:

$$W = K \log m \tag{2.1}$$

K is a constant, and m is the number of current values employed. Suppose we assume that we are working with a code whose characters are all the same duration (this is already an indirect assumption of the ontology of information). In that case, we can consider n as the number of signal elements per character, and then the total number of them that can be constructed is m^n or $n \log m$. The speed of transmission of intelligence is directly proportional to the line speed; we can call it s and inversely proportional to the number of signal elements per character n and obtain the equation W proposed by Nyquist.

His formulation is one of the earliest attempts to link entropy, as proposed by Boltzmann (1.3), with information. He tries to connect statistics with the world of information, but only when every case has the same transmission probability. A big problem in his approach is that information ('Intelligence,' using his words) is sometimes imperfect in the real world. So, Nyquist's formula captures the transmission of intelligence (information) at maximum in the perfect case. Still, it remains very unclear what happens when a message passes through noises or turbulence.

3. Shannon's definition of Information

It was Claude Elwood Shannon to study and give a more solid shape to information theory at the Bell Labs. He took a big step in the definition of information, and his revolutionary paper can be considered the year zero for the theory of information. He formulated a definition of information in a 1948 paper called *A Mathematical Theory of Communication.* The author states: "the fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point" (Shannon, 1948, p. 379). In this paper, the author proposes his theory of information, in which he extends Nyquist's theory. For Shannon, how information is transmitted through messages is not linear: information is, from his point of view, a search in a set of possible good messages. In his view, the core structure of a communication system has a specific structure with some crucial elements:

- The informational source that produces messages for the receiver.
- The transmitter: this is something that modifies the message in the right way for transmission.
- The channel: it is the object that transmits the message.
- The receiver: it has the role of encoding the message received from the transmitter.
- The destination: it is the recipient of the message.

With this taxonomy of the different elements that make up a system, we can build different models of communication systems, not only the most common ones created by computer devices or telephones, but it is also possible to model communication systems formed by people, objects, and very different kinds of actors. Also, in Shannon's approach, like in the Nyquist theory, the measure of information is represented mathematically by a logarithmic function but with different elements. This approach is very convenient for engineering since standard variables, which are helpful in that field, scale linearly with the logarithmic number of possibilities; the logarithmic approach is also very similar to our view of measurement, where we tend to weigh things by comparison. However, information looks pretty different from a measure like length, so what type of information can we describe with this theory?

To describe information in this way we use a mathematical model that describes sequences of symbols with a set of probabilities, and that is the stochastic process (Hoffman, 2013), which in statistical mathematics is a family of random variables. The randomness that we can find in statistical mathematics is called Shannon's entropy:

$$H = -k \sum_{i=1}^{n} p_i \log p_i \tag{3.1}$$

If we analyze the formula, we find that:

- The minus sign is because the result of the rest is every time negative because $\log p_i$ is negative or at maximum 0.
- k is a constant that derives from the choice of a unit of measurement. This value helps to control the basis of the logarithm. The value of k is equal to one if the basis of the log is 2, and it is different otherwise.

- p_i is the probability of the success of the event i.
- $\log p_i$ comes from the fact that we have disjoint events, and all of them are distinguished by the number of characters with the base. Also, the uncertainty has to respect the additive property, so the only possible operation that can produce the maximum probability equal to one is the logarithm (for our purposes, 2 is the norm because of the number of bits in computing).
- $\sum_{i=1}^{n}$ represents that if we have a discrete condition, H is the average sum of each possible event.

Shannon's entropy is not the same as its dynamical equivalent, Boltzmann's entropy (1.3), but there is a correlation at equilibrium. The Shannon entropy represents the average level of 'uncertainty' about possible outcomes. Rather, thermodynamic entropy represents all possible combinations of a system, and it is not possible to calculate the entropy of an intermediate situation because this property only has value at equilibrium. Instead, Shannon's entropy has different results in different cases, with several possible outcomes, while the physical entropy has only one pointed result. An essential feature of the Shannon entropy is that $H = 0$ if and only if a probability has a value of one (the remaining p_i could be zero). That result is another big difference with thermodynamic entropy, where you can't have a value of zero as imposed by the third principle of thermodynamics (absolute zero is unattainable, the only situation where entropy could be zero).

Using the definition of informational entropy and the common structure (at least in Shannon's view) of each communication system, his approach considers information such as a message from a transmitter to a receiver, with a correlation of information, acquired and missed, in the form of entropy. This view is widespread today, and Shannon's entropy is now widely used in computer communications. "This operational motivation for defining entropy in terms of data compression expresses a key concept in the philosophy of information theory: fundamental measures of information arise as answers to fundamental questions about the physical resources required to solve some information processing problem" (Nielsen & Chaung, 2010, p. 501).

However, Shannon's theory has several implications and problems that need to be considered to use it as a unified definition of information. First, information is regarded merely as a selection of symbols from a given set, so information has only a technical meaning in this definition. One possible consequence is that two messages from two different sets of symbols could contain the same amount of information, without considering their meanings. Thus, "according to *A Mathematical Theory of Communication*, the classic monkey randomly pressing the typewriter keys is indeed producing a lot of information" (Floridi, 2009, p. 33), the same that we can find, for example, in a copy of an Italian vocabulary. This problem seems to be a considerable lack in Shannon's proposal to research a unified concept of information. Another critical point to note is that Shannon's proposal resembles a probability theory more than an information theory. It can be an advantage if we want to apply the theory to communication technologies (this is Shannon's aim when he developed the Mathematical Theory of Communication). However, when we try to use this method with our thoughts, where memory plays an important role, we can easily say that "the mathematical theory of communication deals with the transmission of information, not the information itself" (Weaver, 1949, p. 12).

4. Von Neumann's entropy

Shannon entropy is also used in quantum mechanics, where von Neumann entropy can be conceived as the quantum generalization of Shannon entropy. This concept of entropy was proposed by von Neumann in 1955. While Shannon's entropy applies to classical probability distributions, von Neumann's entropy applies to quantum states described by the density matrix von Neumann entropy S is defined as:

$$S(\rho) = -Tr(\rho \log \rho) \tag{4.1}$$

Where Tr denotes the trace operation (sum of the diagonal elements of the matrix) and log is the logarithm of the density matrix. The entropy of a quantum state provides a quantitative measure of how 'mixed' a system is. The von Neumann Entropy of a pure state is equal to zero. Otherwise, a value greater than zero represents a mixed state. An interpretation of values like this reflects the idea that when we have a pure state, we do not have any uncertainty about the system. In contrast, for a mixed state, we are, in a sense, lacking information about the system's state. Thus, Shannon's Entropy (3.1) is a special case of von Neumann's Entropy in quantum systems that can be described by classical statistics (when the density matrix ρ is diagonal).

In the quantum realm, the concept of the von Neumann entropy is correlated with the property of entanglement. Indeed, entropy becomes a natural measure of the quantum correlation between subsystems when two or more quantum systems are entangled (A and B, for example). The system's overall state cannot be decomposed into the individual states of the subsystems because they do not have well-defined properties, even though we know the total state of the combined system perfectly. We can describe them by density matrix ρ_{AB}, but we cannot write the overall density matrix as the product of the density matrices of the individual subsystems $\rho_A \otimes \rho_B$. Instead, the von Neumann entropy of subsystem A, obtained by taking the partial trace of subsystem B, is a measure of the entanglement between A and B:

$$S(\rho_A) = -Tr(\rho_A \log \rho_A) \tag{4.2}$$

If $\rho_A > 0$, the subsets are entangled; therefore, we cannot completely know the state of A without taking B into account. The concept of von Neumann entropy seems very far from that of Shannon. In the case of the similitude between the latter and the thermodynamic formulation of entropy (Hemmo & Shenker, 2006), the only closeness is related to the shape of the equations. Here, the case seems very similar. It is important to notice that the von Neumann entropy, from a theoretical point of view, is used to measure a property in the quantum world: the entanglement, that is completely different from what the Shannon entropy value tries to achieve and, once again, seems not clear from this definition what is the concept of information.

5. Conclusions

Any of the attempts that I tried to cover in this article lack in some points to achieve a general definition of information; one of the most general definitions of information that we can reach is something similar to this: "The information is what is produced by an information source that is required to be reproducible at the destination if the transmission is to be counted a success" (Timpson, 2013, p. 22). But of course, this definition is not enough and also is a cyclical definition where the term information comes inside the definition itself. This analysis of some historical attempts seems to give that information as a quantity is not associated with individual messages but rather characterizes the source of messages. Being a piece of Shannon's information has nothing to do with being a piece of information in the everyday sense. The common ground of the interpretations we briefly considered is that information is always related to an object called a bit (also, von Neumann's entropy can be correlated with something similar: the qubit). What seems more suitable to say about information is that pieces of information are abstract items, while transmitted information as a quantity is a property, as compressibility or channel capacity, so by no means a concrete thing. To have a token of a piece of transmitted information, we need some physical systems, but it does not make what is encoded, stored, or written down physical; the fact that tokens are physical does not mean that the types of which they are instances are.

Acknowledgements

The author would like to thank Prof. Vincenzo Fano and Prof. Claudio Calosi for their invaluable contributions through insightful discussions. Furthermore, I would like to express my gratitude to Prof. Flavia Marcacci for providing me with the invaluable opportunity to present at the LXIV National Congress of SISFA. It was a remarkable chance to showcase my work and to receive constructive feedback from an esteemed audience. I would like to extend my gratitude to the entire Synergia Research Group at the University of Urbino.

Bibliography

Bais, F.A. & Farmer, J.D. (2008). "The Physics of Information", *Philosophy of Information*. Amsterdam: Elsevir, pp. 609-684.

Boltzmann, L. (1877). "Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung resp. den Sätzen über das Wärmegleichgewicht", *Wiener Berichte*, **76**, pp. 373–435.

Floridi, L. (2009). "Philosophical Conceptions of Information", in Sommarufa, G. (eds.), *Formal Theories of Information. Lecture Notes in Computer Science*, Berlin: Springer, pp. 13-53.

Hanlon, R.T. (2020). "Clausius and the road to entropy", *Block by Block: The Historical and Theoretical Foundations of Thermodynamics*. Oxford: Oxford University Press, pp. 438-477.

Hemmo, M. & Shenker, O. (2006). "Von Neumann's Entropy Does Not Correspond to Thermodynamic Entropy", *Philosophy of Science*, **73**(2), pp. 153–174.

Hoffmann, L.D. (2013). *Calculus for business, economics, and the social and life sciences*. New York: McGraw-Hill.

Nielsen, M.A. & Chaung, I.L. (2010). *Quantum Computation and Quantum Information*. New York: Cambridge University Press.

Nyquist, H. (1924). "Certain Factors Affecting Telegraph Speed", *The Bell System Technical Journal*, **3**(2), pp. 324-346

Shannon, C.E. (1948). "A Mathematical Theory of Communication", *The Bell System Technical Journal*, **27**(3), pp. 379-423

Timpson, C.G. (2013). *Quantum Information Theory and the Foundations of Quantum Mechanics*. Oxford: Oxford University Press.

Weaver, W. (1949). "The Mathematics of Communication", *Scientific American*, **181**(1), pp. 11-15.