Proceeding Paper

Information Is (Only) Probability †

Riccardo Manzotti

Department of Business, Law, and Consumer Behavior, IULM University, 20143 Milan, Italy;
riccardo.manzotti@gmail.com
† Presented at Philosophy and Computing Conference, IS4SI Summit 2021, online, 12–19 September 2021.

Abstract: Is information a real entity, or is it only a useful mathematical notion? Shannon's equation and a consolidated tradition have led many scientists and laymen to consider information as some kind of physical, additive quantity. The very well-known argument in favour of information as a physical quantity has not reached any definitive conclusion. Much has been derived from Shannon's formula, which was introduced to provide a convenient logarithmic measure that might be practically useful and close to most intuitive feelings. Shannon never insisted on the ontological status of information, which might be revealed to be a lot less committing than is often believed. The bottom line is that information might be a convenient mathematical notation to express the probability of independent configurations in a system.

Keywords: information; entropy; Shannon; causal overdetermination; reification; ontology

1. Introduction

Here, I will show that there is a mathematical derivation that reduces Shannon's formula to standard probability theory between independent events. I present an alternative formulation of entropy based on remapping events with different probabilities over multiple equiprobable events. This suggests that, as happens with entropy, there are not more probable configurations, but rather there are many indistinguishable configurations that are mapped onto convenient "more probable" states. Since an alternative formula that does not require any ontological commitment is available, the contention is that information does not exist (its existence does not make any difference). If H can be expressed as the product between the probabilities of independent events, there is no need to add any new entity—information is simply a convenient way to express the probability of different configurations of a physical system.

2. Shannon's Equation Revisited

Shannon's equation and a consolidated tradition have led many scientists and laymen to consider information as some kind of physical, additive quantity. The standard formula that all schoolboys learn is the familiar sum of products between probabilities and their logarithms [1]:

$$H(X) = \sum_i^N p_i Log_2 p_i \tag{1}$$

The aim of this paper is to show that the above function (1) can be rewritten in terms of standard probability theory, and that it thus does not entail any ontological commitment. Namely, that it is possible to show that:

$$H(X) = -\frac{1}{M} Log_2 \prod_j^M \widetilde{p}_j \tag{2}$$

where $\widetilde{p}_j's$ ($M \geq N$) maps the probabilities, p_i, of the random variable, X, using a method that will be described below. It can be shown that Equations (1) and (2) are mathematically

equivalent (in the sense of providing the same set of solutions). Since Equation (2) is a function of a product of probabilities, it follows that H expresses, logarithmically, the probability of independent events, i.e., P(AB) = P(A)P(B). Thus, by Ockham's principle, if Equation (2) were valid it would drain all ontological commitments of Equation (1). Information as a real quantity could be dismissed.

A few steps show how Equation (2) can be derived from Equation (1). First, Equation (1) can be rewritten as an exponent, and thus the sum can be reformulated as a product with exponents:

$$2^{-H} = 2^{\sum_i^N p_i Log_2 p_i} = \prod_i^N \left(2^{Log_2 p_i}\right)^{p_i} = \prod_i^N p_i{}^{p_i} \tag{3}$$

Equation (3) begins to take some of the spell away. The next step consists of removing exponents and suggesting suitable independent events. Let d be the greatest common factor among all p_i (assuming $p_i \in \mathbb{Q}$, $\forall i$, $d \in \mathbb{Q}$, d is the greatest rational, such that $p_i/d = q \in \mathbb{N}$, $\forall i$). By means of d, Equation (3) can then be rewritten as (Figure 1):

$$\prod_i^N p_i{}^{p_i} = p_1{}^{p_1} \cdot p_2{}^{p_2} \cdot \ldots \cdot p_N{}^{p_N} = p_1{}^d \cdot p_1{}^d \cdot \ldots \cdot p_N{}^d = \widetilde{p}_1{}^d \cdot \widetilde{p}_2{}^d \cdot \ldots \cdot \widetilde{p}_M{}^d \tag{4}$$

whereas each factor, $p_i{}^{p_i}$, is rewritten as a product, $\prod_k^{K_i} (p_i)^d$, where $K_i = p_i/d$, $K_i \geq 1$, $M = \sum_i^N \frac{p_i}{d}$. Of course, if $d = p_i$, $K_i = 1$. Thus, each p_i can be revisited by—and thus substituted with—a series of independent events, $\widetilde{x}_i$, $1 \leq i \leq M$, each having a probability, $\widetilde{p}_i$. The set of probabilities, $\widetilde{p}_i$, is built by multiplying each p_i K_i times. Equation (3) can be rewritten as:

$$2^{-H} = \prod_i^N p_i{}^{p_i} = \prod_j^M \widetilde{p}_i{}^d = \left[\prod_j^M \widetilde{p}_i\right]^d \tag{5}$$
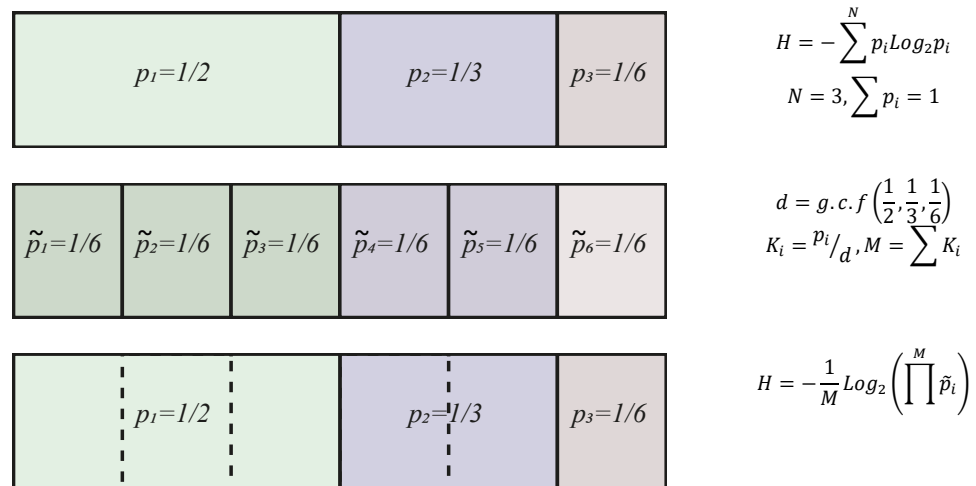

Figure 1. Each discrete set of probabilities p_i can be split into a larger set of equally probable events $\widetilde{p}_i$.

Additionally, then, since $M = \sum^N (p_i/d) = \sum^N p_i/d = 1/M$, it is trivial to obtain Equation (2). The derivation from Equation (1) to Equation (2) might appear complicated, but it is not. For instance, if the elements are equiprobable symbols ($p_i = \frac{1}{N}$), the familiar Hartley's expression, $H = Log_2 N$, will follow [2]. An important caveat: $\prod^N p_i$ does not correspond to a probability ($p(i \cap j) = 0$, $\forall i \neq j$), $\prod^M \widetilde{p}_i$ does. In fact, $\prod^M \widetilde{p}_i$ is the probability of any sequence of M events whose frequency matches exactly that of the outcomes of X (Figure 2). In the future, it would be interesting to check whether this proposal might be extended to the continuous case, thereby providing an even more compelling case.
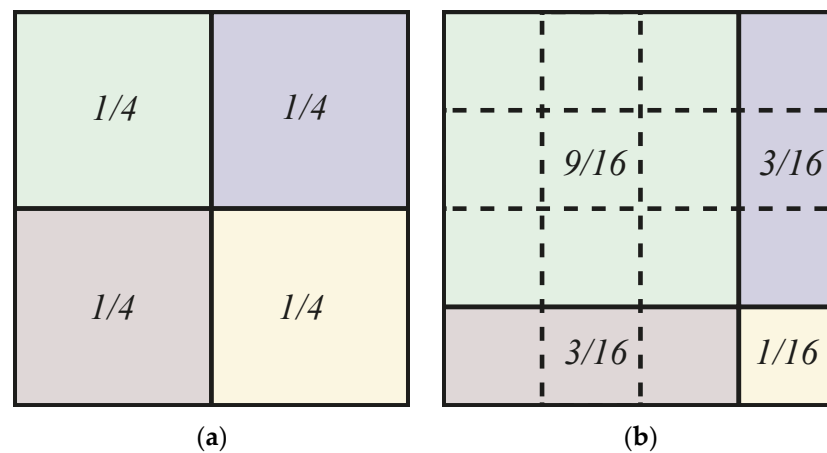
Figure 2. (a) X, such that $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{2}$, since $\prod^M \widetilde{p}_i = \widetilde{p}_1\widetilde{p}_2 = \frac{1}{4}$; (b) X, such that $p_1 = \frac{3}{4}$, $p_2 = \frac{1}{4}$, since $\prod\limits^M \widetilde{p}_i = \widetilde{p}_1\widetilde{p}_2 = \frac{3}{16}$.

Consider a handful of numerical examples.

First, suppose having two outcomes with probabilities $\left\{\frac{1}{3}, \frac{1}{2}\right\}$, then $d = \frac{1}{6}$, since $\frac{1}{6}$ is their greatest common factor, i.e., d is a rational number by which $\frac{1}{3}/\frac{1}{6} = 2$, $\frac{1}{2}/\frac{1}{6} = 3$. In fact, $\left(\frac{1}{3}\right)^{\frac{1}{3}} \cdot \left(\frac{1}{3}\right)^{\frac{1}{3}} = \left(\frac{1}{3}\right)^{\frac{1}{6}}\left(\frac{1}{3}\right)^{\frac{1}{6}} \cdot \left(\frac{1}{2}\right)^{\frac{1}{6}}\left(\frac{1}{2}\right)^{\frac{1}{6}}\left(\frac{1}{2}\right)^{\frac{1}{6}}$.

Second, suppose having three outcomes, x_1, x_2, x_3, with probabilities $\left\{\frac{1}{6}, \frac{1}{3}, \frac{1}{2}\right\}$, $N = 3$. Their greatest common factor common is $d = 1/6$. By multiplying all elements according to the ratio $K_i = p_i/d$, a set, $\{\widetilde{x}_j\}$, $M = 6$, where each element has a probability of $d = \frac{1}{6} = \frac{1}{M}$, we obtain $\{p_1, p_2, p_3\} \Rightarrow \{\widetilde{p}_1, \widetilde{p}_2, \widetilde{p}_3, \widetilde{p}_4, \widetilde{p}_5, \widetilde{p}_6\}$. Equation (1) obtains $H = -\left(\frac{1}{6}Log_2\frac{1}{6} + \frac{1}{3}Log_2\frac{1}{3} + \frac{1}{2}Log_2\frac{1}{2}\right) \cong 1.459147917\ldots$ Equation (2) obtains $H = -\frac{1}{6}Log_2\left(\frac{1}{6} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}\right) \cong 1.459147917\ldots$

Finally, suppose having two outcomes with probabilities $\left\{\frac{3}{4}, \frac{1}{4}\right\}$, $N = 2$. Then, $d = \frac{1}{4}$, $M = 4$. $\{p_1, p_2\} \Rightarrow \{\widetilde{p}_1, \widetilde{p}_2, \widetilde{p}_3, \widetilde{p}_4\}$. The probability $\prod^M \widetilde{p}_i$ represents the probability of any sequence of M outcomes of the kind 0001, 0100, and 1000 (or whatever matches X's profile).

3. Discussion

Is information something real or not? In many fields, the notion that information is a physical quantity has led scientists to debatable conclusions about both its nature and role [3,4]. Many arguments have been advanced both in favour of and against the reality of information. It is fair to maintain that information cannot be directly measured, as can be done with other, less problematic physical quantities (say, mass and charge). It is impossible to tell how much information is contained in a physical system unless more knowledge about the relation between that physical system and another system is available. It is also disputable whether information has any real causal efficacy, or if it is an epiphenomenal notion, as is centres of mass. The very well-known argument in favour of information as a physical quantity has not reached any definitive conclusion [5]. In this context, much have been derived from Equation (1), which, as Shannon himself stated, was introduced mostly because "the logarithmic measure is more convenient [...] it is practically more useful [...] it is mathematically more suitable [...] it is nearer to our intuitive feelings" [1]. Shannon hardly provides final arguments for the physical existence of information.

An interesting outcome of Equation (2) is that it is achieved by remapping events with different probabilities over multiple equiprobable events. This suggests that, as happens with entropy, there are not more probable configurations, but rather there are

many indistinguishable configurations that are mapped onto convenient "more probable" states. Therefore, Equation (2) (and the steps from three to six) is a way to unpack one's limited knowledge or perspective about physical states and represent them in a neutral way where each state is just as probable.

This is not to say that the notion of information is not a very successful and useful one. Yet, the existence of an alternative formula that does not require any ontological commitment to information makes a strong cause against its existence—the existence of information does not seem to make any difference. Equations (1) and (2) causally overdetermine what a system does, and, since Equation (1) is ontologically less parsimonious, for the Ockham's razor Equation (2) is to be preferred.

In this contribution, Equation (2) expresses H in terms of the product between the probabilities of independent events. Therefore, if Equation (2) was used in place of Equation (1) (despite being more cumbersome), there would be no need to add any new entity—information is simply a convenient way to express the probability of different configurations of a physical system. Since Equation (2) is mathematically equivalent to Equation (1), and it does not require anything but probabilities, it follows that Equation (2) provides a mathematical proof that information does not exist apart from as a useful conceptual tool.

References

1. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656. [CrossRef]
2. Hartley, R.V.L. Transmission of Information. *Bell Syst. Tech. J.* **1928**, *7*, 535–563. [CrossRef]
3. Wheeler, J.A. Information, Physics, Quantum: The Search for Links. In *Third International Symposium on the Foundations of Quantum Mechanics*; Physical Society of Japan: Tokyo, Japan, 1989.
4. Tononi, G.; Boly, M.; Massimini, M.; Koch, C. Integrated information theory: From consciousness to its physical substrate. *Nat. Rev. Neurosci.* **2016**, *17*, 450–461. [CrossRef] [PubMed]
5. Levy, A. Information in Biology: A Fictionalist Account. *Noûs* **2011**, *45*, 640–657. [CrossRef]