Type of Contribution: PAPER

The dichotomy of the notion of information regarding sender's data structure and recipient's background knowledge

Keywords: information, data, informativeness, sender's data structure, recipient's background knowledge

## Introduction

If it is possible to give a comprehensive overview of the notion of information – besides through an exhaustive review of the trends in the development of information theories which took place during the second half of the 20th century – then it will be achieved by observing the two theoretical approaches with opposing principles: the objectivist and subjectivist understandings of this phenomenon. As such, the aforementioned antagonism is reminiscent of that in philosophy which relates to reality itself, namely, whether there are individual 'things-in-the-world' or only 'the totality of reality' of which we are stakeholders; we will never be sure whether we directly perceive reality ('objectively') or indirectly perceive reality ('subjectively'), as if we had constructed it in our minds (Putnam 1980).

Thus, information, like reality, can be understood as "something objective, quantitative, and mainly associated with data" on the one hand, or, on the other hand, "as subjective, qualitative, and mainly associated with knowledge, meaning, and understanding" (Robinson and Bawden 2014, 131). However, it seems the objectivist and subjectivist approaches to information arises from the etymology of this term, which are based on two entirely opposite meanings. Accordingly, it is quite possible that the *ontological* meaning of information, 'to give form to', which originally referred to the formation of matter itself and only later represented 'the act of moulding the mind', may be the source of any objectivist understandings of this phenomenon; in contrast, the *epistemological* meaning of information, which prevailed in the 16th century and presented information as 'the act of communicating knowledge', or "the act of informing", in principle implies a subjective understanding of this phenomenon (Capurro 1996, Cappuro and Hjørland 2003, Buckland 1991).

It is paradoxical that the ancient ontological meaning of information reappears in C. E. Shannon's *A Mathematical Theory of Communication*, the "information theory" published in 1948 (Shannon 1948) in which 'information content', independent of its interpretation and any 'psychological factors', appears as a new cybernetic conception of the world as supplementing the totality of reality, beside the terms 'matter' and 'energy,' (Capurro, 1996). In other words, thanks to Shannon's theory and aided by a new cybernetic worldview (Wiener, 1948) as well as the biological discovery of genes (Crick & Watson, 1953), an objectivist understanding of this phenomenon seems to have returned.

At this point in the discussion, we may ask if it is possible to talk about two types of information or informativeness as opposed to simply referring to two different understandings of the same phenomenon. In accordance with Shannon's theory, by analysing the role of information as 'pure informative content' with respect to both the sender's data structure and the recipient's knowledge structure, this paper seeks to point out the inherent dichotomy of the notion of information viewed in this context.

## Shannon's theory as a "wedding" of information and data

What makes C. E. Shannon's theory exceptional is reflected in the quantification of the notion of information using a new measure: the amount of information which can be managed by probabilistic calculation. The amount of information "defined as the logarithm of the number of choices" (Shannon and Weaver 1949, 10) began to express the informativeness of the content in a message and no longer referred to the content itself.

Although the first mention of the term *data* in computer literature occurs a little earlier (1946), understood as "transmissible and storable information by which computer operations are performed" (Etymonline, 2019), the expanded use of this term definitely begins with Shannon's theory. The term *data* is nowadays easy to relate to Shannon's predefined set of messages as a kind of dataset with data transmitted via a communication channel between sender and recipient. Accordingly, this paper does not use Shannon's term *messages*, and instead uses the term *data*.

Accordingly, the informativeness of selected data from a predefined dataset is higher if it prevents the selection of other data from the predefined dataset (as alternative options). For example, selecting the letter 's' not only makes it impossible to select any of the other 25 letters of the English alphabet but also makes it impossible to select all other words which begin with those letters! This is why the informative character of the letter 's' is high. Thus, the informative nature of data arises from the effort one makes to extract it from a predefined dataset. In a strictly mathematical sense, this separation is carried out by repeatedly asking YES/NO questions. In this way, Shannon defined one bit of information as the amount of information that one is allowed to choose between two equally probable possibilities (Stone, 2014). It is easy to conclude that a longer code is needed in case of a bigger dataset to extract one datum from it. However, this legality applies only in the case of equal probabilities for all data in the dataset. When data with different probabilities are considered, the more likely data become those which are less informative because only a few YES/NO questions as bits of information are needed to extract it from the core set.

## Background knowledge of recipients

However, if the informativeness of selected data, on the one hand, arises as the result of efforts to extract it from a predefined dataset, then, on the other hand, by mere analogy, the informativeness of the same data must be coupled with an effort to incorporate it into the knowledge structure of the recipient. Successfully incorporating the selected data into the knowledge structure will result in its change in a given context. As we know, the previous idea is expressed by Bertram C. Brookes' 'fundamental equation of information science', a non-operational formula-model which states that K[S] + ΔI = K[S + ΔS] (Brookes, 1980). The knowledge structure (K[S]) of a recipient changes with each (new) inflow of information (ΔI). It follows that the same information can have different effects on different knowledge structures (Bawden, 2011). Accordingly, if the informativeness of selected data, in Shannon's sense, represents a certain potential of data to be extracted from a predefined dataset, then the informativeness of the same data in relation to the background knowledge of the recipient would represent its potential to be incorporated/embedded into it. It becomes clear that the potential of data to be embedded in the recipient's K[S] no longer depends on the size of the predefined dataset from which it originates, but instead depends upon the potential of the recipient's K[S] to receive new data.

## Model of informativeness regarding data and knowledge structure

When we talk about the background knowledge of the recipient in the most general sense, aside from its structural aspects and any kind of concretisation, we must still first imagine it as a K[S] connected by certain relations. Regardless of what is considered the 'constituents of knowledge' in such an imagined structure, we hold that relations represent the primary phenomenon which comprises the K[S] itself. For a K[S], in Brookes’ sense, can be said to have the potential to receive new data or information only if it is in the anomalous state of knowledge (ASK) in a given context. As such, it should be noted that 'ASK' is not used here in its original sense (Belkin, 1980). We only want to emphasise that ASK is characterised by a smaller number of established relations in the recipient's K[S] in the given context. In principle, we expect that a more pronounced ASK means the transmittance of more informative data via the communication channel.

However, even in the case of a K[S] with a pronounced ASK, there is a possibility that the received data will be unable to establish any relations in the K[S], first, because it already exists in it (for example, the case of receiving a message that is not new to us), and, second, because there is no default context in which the relations could be established (for example, if we were given a book in a language we do not know). However, in both cases, it is important to note that the informativeness of data can still be regarded as the number of relations with the K[S] of the recipient. Although ASK cannot directly help us to assess the informativeness of data regarding a K[S], it ultimately gives us insight into how we can achieve it. The calculation seems simple, although it took years to express it: in a recipient's K[S], in any given context, more informative data are those which establish a greater number of relations. At the same time, this is a main hypothesis of this paper. Furthermore, established relations can be relations not only between existing constituents of knowledge but also newly formed constituents of knowledge (regardless of what they might be), and from which, after all, the meaning of received data or information derives. In this sense, we refer to *data informativeness* in relation to the sender's data structure as Shannon's informativeness (ΔIs) and *data informativeness* in relation to the structure of the recipient's knowledge as Brookes' informativeness (ΔIb) (Figure 1).



Figure 1 – Shannon's and Brookes' informativeness of transmitted data via communication channel.

This paper will demonstrate the presented hypotheses using concrete examples.

## Conclusion

We conclude that if Shannon's informativeness is inversely proportional to the probability of extracting selected data from a data structure (D[S]), which is relative to the total amount of data in a sender's D[S], then Brookes' informativeness is proportional to the probability that selected data could establish non-existing relations with a recipient's K[S], which is relative to the total number of relations in a given context. The communication process, therefore, does not encapsulate the knowledge acquisition process. If information is understood as a 'measure of surprise' or the 'unexpectedness of content' in a predetermined environment in which 'everything is already known' in the context of the communication process, then, by analogy, it is also a 'measure of novelty' which can be expressed through the number of relations that selected data can establish through the existing background knowledge of the recipient, in which not everything is known in the context of the knowledge acquisition process.

# REFERENCES

# Bawden, David. 2011. "Brookes Equation: The Basis for a Qualitative Characterization of Information Behaviours." *Journal of Information Science* 37, 1 (February): 101–8.

Belkin, Nicholas J. 1980. "Anomalous states of knowledge as a basis for information retrieval." *Canadian journal of information science* 5, 1: 133-143.

# Brookes, Bertram C. 1980. "The foundations of information science. Part I: philosophical aspects." *Journal of Information Science* 2, 3-4: 125-133.

Buckland, Michael K. 1991. "Information as Thing". *Journal of the American Society for Information Science* 42, 5: 351–60.

Capurro, Rafael. "On the genealogy of information." In *Information. New questions to a multidisciplinary concept*, edited byKornwachs, Klaus, and Konstantin Jacoby, 259-270. *Berlin: Akademie Verlag*, 1996.

Capurro, Rafael and Birger Hjørland. 2003. "The concept of information.” *Annual Review of Information Science and Technology* 37, 1: 343–411.

Etymonline. 2019. "Data." Last modified September 11. 2019. https://www.etymonline.com/word/data

Putnam, Hilary. 1981. *Reason, truth and history*. Vol. 3. Cambridge: University Press.

Robinson, Lyn and David Bawden. 2014. "Mind the Gap: Transitions Between Concepts of Information in Varied Domains." In *Theories of Information, Communication and Knowledge: a Multidisciplinary Approach, Studies in History and Philosophy of Science*, edited by Ibekwe-SanJuan, Fidelia and Thomas Dousa, 121-41. Dordrecht, Heidelberg, New York, London: Springer.

Shannon, Claude E. 1948. "A mathematical theory of communication", *The Bell System Technical Journal* 27: 379–423, 623–656.

Watson, James D., and Francis H. C. Crick. 1953. "Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid." *Nature* 171: 737-738.

Weaver, Warren. 1949. "Recent Contributions to the Mathematical Theory of Communication", In Shannon, Claude E. and Warren Weaver. 1949. "The mathematical theory of communication", 3-28. Urbana: The University of Illinois Press.

Wiener, Norbert. 1948. *Cybernetics or Control and Communication in the Animal and the Machine*. Cambridge, MA:Technology Press.