

# The Difference that Makes a Difference for the Conceptualization of Information <sup>†</sup>

**Marcin J. Schroeder**

Akita International University, 010-1211 Akita, Japan; mjs@aiu.ac.jp; Tel.: +81-18-886-5984

<sup>†</sup> Presented at the IS4SI 2017 Summit DIGITALISATION FOR A SUSTAINABLE SOCIETY, Gothenburg, Sweden, 12–16 June 2017.

Published: 9 June 2017

**Abstract:** Information is a subject of multiple efforts of conceptualization leading to controversies. Not frequently sufficient effort is made to formulate the concept of information in a way leading to its formal mathematical theory. Discussions of conceptualizations of information usually are focusing on the articulation of definitions, but not on their consequences for theoretical studies. This paper compares two conceptualizations of information exploring their mathematical theories. One of these concepts and its mathematical theory were introduced in earlier publications of the author. Information was defined in terms of the opposition of one and many and its theory was formulated in terms of closure spaces. The other concept of information was formulated in a rather open-ended way by Bateson as “any difference that makes a difference”. There are some similarities between Bateson’s concept of information and that of MacKay. In this paper a mathematical theory is formulated for this alternative approach to information founded on the concept of a difference in terms of generalized orthogonality relation. Finally, the mathematical formalisms for both approaches are compared and related. In conclusion of that comparison the approach to information founded on the concept of difference is a special case for the approach based on one-and-many opposition.

**Keywords:** information concept; information definition; information theory; generalized orthogonality; closure spaces

---

## 1. Introduction

The concept of information is a subject of never ending discussions. The fact that these discussions do not lead to consensus generates a lot of anxiety among those who are engaged in the study of information, while this should be considered best evidence for the non-trivial character of this concept and as such be a source of joy. The actual problem is not in the variety of different definitions, but in the fact that many of them are deficient in logical rigor and that their mutual comparisons rarely go beyond the surface of verbal articulation. It seems that more attention is payed to the normative question what “should” be called information than to the issue of the explanatory power of the concept in the contexts of its use. There is nothing necessitating the choice of the particular definition of any concept and of course this applies to the concept of information too. Therefore, criteria for evaluation and comparisons of definitions can be found only in their consequences for the development of the theory of information understood as a complex of assertions regarding its characteristics, structure, properties and relations to other concepts.

This is exactly why so called information theory developed by Shannon is not a theory of information at all, but a theory of communication. Shannon never defined the concept of information in his great study of communication, which does not tell us anything about the structural characteristics or properties of information and even its quantitative characteristic in the form of entropy is problematic [1]. Actually, the word “information” in his famous article appears only few

times and its only important occurrence (and probably last in entire text) is in the context of quantities that have form of entropy known from “statistical mechanics” and that “play a central role in information theory as measures of information, choice and uncertainty” [2]. Probably Shannon’s unfortunate reference to “information theory” as if such theory existed already contributed to persisting confusion regarding what information theory is in spite of the continuing strong objections to its identification with Shannon’s theory of communication [3].

Shannon’s goal was to develop a mathematical theory of communication and therefore he cannot be blamed for not paying enough attention to the concept of information and its characterization. It is more problematic that frequently contributions to the discussion of information are equally vague regarding what exactly information is, how its concept can be described in a formal way and what we can assert about it. Competing voices about information are usually so incompatible (information as representation, information conceived through conduit metaphor, information in linguistic context, information as data in computation, etc.) that no comparison of the concepts involved is possible. Even more controversial are very strong claims, for which their authors do not provide any justification (e.g., “no information without representation” used as a slogan by followers of MacKay’s approach to information as “that which adds to a representation” [4]).

Not always, or even not frequently sufficient effort is made to formulate the concept of information in a way leading to its formal mathematical theory. Mathematical formulation is important, because mathematical theories of concepts can be easily compared through analysis of their theorems. This paper is exploring such comparison between mathematical theories of information for two conceptualizations of information. One of these concepts and the theory derived from it were introduced in earlier publications of the author. Information was defined by him in terms of the categorial opposition of one and many, as that which makes one out of many either by the selection or by structuralization [1,5]. Thus, this many can be made one by a selection of an element of the variety constituting the many, or by a structure which unify the many into one. Mathematical theory of such concept was presented and analyzed in many earlier publications of the author [6,7].

The other concept of information considered here is probably the most popular of all attempts in conceptualization of information was formulated in a rather open-ended way by Gregory Bateson in several of his publications from the 1970’s [8]. But it was the glossary appended to his last book that made it a famous, commonly invoked slogan “information is any difference that makes a difference” [9]. This description of information is not a precise definition, but not just a game of words either. Of course, its popularity owes a lot to its polysemic, proverbial form and vernacular language. The lack of precision may increase its attractiveness, as everyone can find it consistent with own views. In particular, the use of the idiomatic expression “makes a difference” opens it to a variety of interpretations. It can indicate effectiveness, for instance in the sense of causation, or it can have a normative interpretation as an indication of importance. Actually Bateson apparently appreciated this ambiguity, as he dropped the ending “in some later event” suggesting the former interpretation from his “definition” as formulated in earlier papers (“information is any difference that makes a difference in some later event” [8]).

To be fair, we can find similar idiomatic expression in MacKay’s study of information in the context of what he considered “operational” definition of information: “We shall find it profitable to ask: ‘To what does information make a difference? What are its effects?’ This will lead us to an ‘operational’ definition covering all senses of the term, which we can then examine in detail for measurable properties” [10]. He tries to answer the question about the effects of information, but not how information makes a difference. So his use of the idiomatic expression has the same intention as that of Bateson to avoid being bound by any commitment to a specific interpretation.

Bateson’s way to information as “any difference that makes a difference” began already in 1951 in the spirit much closer to MacKay’s representational view of information: “Every piece of information has the characteristic that it makes a positive assertion and at the same time makes a denial of the opposite of that assertion” [11]. But already at that time he recognized the role of differences: “In this sense, our initial sensory data are always ‘first derivatives’, statements about

*differences* which exist among external objects or statements about changes which occur either in them or in our relationship to them. [...] What we perceive easily is difference and change—and difference is a relationship” [12]. In the following years we can see that his view of information became increasingly general, but instead of lifting the level of abstraction and looking for more abstract conceptual framework, Bateson remained at the level of common sense concepts, but tried to formulate his description increasingly open-ended.

Why are Bateson’s and MacKay’s studies of information distinct among so many other attempts? They both are motivated by the interest in structural aspects of information, but try not to sever the connection to Shannonian theory of communication. Neither includes actual structural analysis of information or goes beyond purely declarative interest in structures, but both recognize the importance of structural characteristics of information. MacKay explicitly refers to the concept of a structure, for instance when he writes: “By representation is meant any structure (pattern, picture, model) whether abstract or concrete, of which the features purport to symbolize or correspond in some sense with those of some other structure” [13]. Also, he writes about Structural Information-Content as “The number of *distinguishable groups or clusters* in a representation [...] Thus structural information is not concerned with the *number* of elements in a pattern, but with the possibility of *distinguishing between them*” [14]. There is nothing here about what actually structure is, except some scattered common sense examples of “pattern, picture, model” and a vague statement that structure’s presence is manifested by some grouping or clustering of elements and that this introduces possibility of making distinctions, i.e., to recognize differences.

Since neither Bateson, nor MacKay clarified the qualifying expression of “making difference” and the former intentionally leaves this qualification open-ended, in this paper the second, alternative to that of the present author approach to information is understood as founded on the concept of a difference without its qualification. It will be shown in the next section of the paper that this concept has a surprisingly rich philosophical consequences and interesting mathematical theory. Finally, in the third section the mathematical formalisms for both approaches are compared and related. The surprising conclusion of that comparison is that the approach to information founded on the concept of difference is a special case for the approach based on one-and-many opposition and its formalism in closure spaces.

## 2. Difference and Structure

The concept of difference (Latin *differentia*, Greek *diaphora*) assumed very early prominent position in philosophy along with those of a genus and species due to its role in Aristotelian logic (*Prior Analytics* 24<sup>a</sup>16-25<sup>a</sup>13) [15]. Differentia between species became a fundamental tool in defining universals. Aristotle gave it also an important role in the study of substance (*Metaphysics* 1037<sup>b</sup>8-1039<sup>a</sup>8) [15]. However after the decline of the interest in Scholastic philosophy in the advent of the Scientific Revolution of the 17th century it was relegated to the secondary role of the negation of the equality or equivalence relations. There was more interest in what makes things similar than different.

One notable exception was the recognition by John Wilkins of the importance of difference in cognition and especially in matters related to cryptography in his 1642 book on the subject of cryptography *Mercury or the Secret and Swift Messenger*: “For in the general we must note, that whatever is capable of a competent Difference, perceptible to any Sense, may be a Sufficient Means whereby to express the Cogitations. It is more convenient, indeed, that these Differences should be of as great Variety as the Letters of the Alphabet; but it is sufficient if they be but twofold, because Two alone may, with somewhat more Labour and Time, be well enough contrived to express all the rest” [16].

Bateson’s description of information as “a difference that makes a difference” and MacKay’s references to structural content of information clearly associated with differences are always considered as independent, original and unprecedented contributions to the study of information. Sometimes there are voices that at least chronological priority should be given to MacKay in the setting foundations for information in the concept of difference, which is disputable. However, they

both must have been influenced by the dominating at the time philosophical and methodological structuralism. It is extremely unlikely that they both were unaware of the works of Herman Weyl [17], Jean Piaget [18], Claude Levi-Strauss [19] and stayed insulated from the philosophical discourse on the fundamental role of structures across all domains of human inquiry.

Furthermore, it is very unlikely that they were not familiar with the original source of the structuralistic methodology in the works of Ferdinand de Saussure, specifically in his 1916 book *Course in General Linguistics*. His general study of the language (after all the primary example of information system) was based on the idea of the transition from the traditional diachronic approach focusing on the derivations of linguistic forms from historically earlier ones to the synchronic methodology analyzing structural characteristics. But the structure of the language according to de Saussure is manifested in differences: “Everything that has been said up to this point boils down to this: in language there are only differences. [...] Language has neither ideas nor sounds that existed before the linguistic system, but only conceptual or phonic differences that have issued from the system. [...] Any nascent difference will tend invariably to become significant but without always succeeding or being successful on the first trial. Conversely, any conceptual difference perceived by the mind seeks to find expression through a distinct signifier, and two ideas that are no longer distinct in the mind tend to merge into the same signifier [20].

### 3. Mathematical Formalisms

We can proceed to mathematical formalisms of the two approaches to information. Thus, the author of this paper defined information as a resolution of the one-many opposition, or in other words as that, which makes one out of many. There are two ways in which many can be made one, either by the selection of one out of many, or by binding the many into a whole by some structure. The former is a selective manifestation of information and the latter is a structural manifestation. They are different manifestations of the same concept of information, not different types, as one is always accompanied by the other, although the multiplicity (many) can be different in each case.

Now we can interpret this definition within mathematical theory of closure spaces [21]. The concept of information requires a variety (many), which can be understood as an arbitrary set  $S$  (called a *carrier of information*). *Information system* is this set  $S$  equipped with the family of subsets  $\mathcal{F}$  satisfying conditions: entire  $S$  is in  $\mathcal{F}$ , and together with every subfamily of  $\mathcal{F}$ , its intersection belongs to  $\mathcal{F}$ , i.e.,  $\mathcal{F}$  is a Moore family. Of course, this means that we have a closure operator defined on  $S$  (i.e., a function  $f$  on the power set  $2^S$  of a set  $S$  such that:

- (1) For every subset  $A$  of  $S$ ,  $A \subseteq f(A)$ ;
- (2) For all subsets  $A, B$  of  $S$ ,  $A \subseteq B \Rightarrow f(A) \subseteq f(B)$ ;
- (3) For every subset  $A$  of  $S$ ,  $f(f(A)) = f(A)$ .

The Moore family  $\mathcal{F}$  of subsets is simply the family  $f$ -Cl of all closed subsets, i.e., subsets  $A$  of  $S$  such that  $A = f(A)$ . The family of closed subsets  $\mathcal{F} = f$ -Cl is equipped with the structure of a complete lattice  $L_f$  by the set theoretical inclusion.  $L_f$  can play a role of the generalization of logic for not necessarily linguistic information systems, although it does not have to be a Boolean algebra. In many cases it maintains all fundamental characteristics of a logical system [22].

*Information* itself is a distinction of a subset  $\mathcal{F}_0$  of  $\mathcal{F}$ , such that it is closed with respect to (pair-wise) intersection and is dually-hereditary, i.e., with each subset belonging to  $\mathcal{F}_0$ , all subsets of  $S$  including it belong to  $\mathcal{F}_0$  (i.e.,  $\mathcal{F}_0$  is a filter in  $L_f$ ).

The Moore family  $\mathcal{F}$  can represent a variety of structures of a particular type (e.g., geometric, topological, algebraic, logical, etc.) defined on the subsets of  $S$ . This corresponds to the structural manifestation of information and gives the expression “structural” explicit meaning. Filter  $\mathcal{F}_0$  in turn, in many mathematical theories associated with localization, can be used as a tool for identification, i.e., selection of an element within the family  $\mathcal{F}$ , and under some conditions in the set  $S$ . For instance, in the context of Shannon’s selective information based on a probability distribution of the choice of an element in  $S$ ,  $\mathcal{F}_0$  consists of elements in  $S$  which have probability measure 1, while  $\mathcal{F}$  is simply the

set of all (measurable) subsets of  $S$ . Thus, this approach combines both manifestations of information, the selective and the structural.

Now we can consider the formalism for the general concept of difference. In mathematics this concept is usually called *generalized orthogonality* (with possible qualifications indicating its variations as “strong”, “weak”, etc.). The reason is that orthogonality in vector spaces equipped with scalar product is a good model of the relationship in a very general case.

The abstract orthogonality relation is defined on a set  $S$  by the conditions [23,24]:

1.  $\forall x, y \in S: x \perp y \Rightarrow y \perp x$ , i.e. relation  $\perp$  is symmetric,
2.  $\forall x \in S: x \perp x \Rightarrow x \perp y$  for all  $y$  in  $S$ ,

Of course, the second condition may seem strange. How anything can be different from, or orthogonal to itself. However zero vector in vector spaces with a scalar product is orthogonal to itself. Also, if we assume that the relation is irreflexive (no element is orthogonal to itself) the second condition is satisfied. Therefore there is no reason to object such generalization when it merges several different mathematical concepts analogous to the common sense word “difference”.

If the set  $S$  has an additional structure of a partial order, then we can enrich the theory of orthogonality in the following way.

We can consider more general structure of a poset  $[P, \leq]$  with the so called *strong orthogonality* relation  $\perp$  defined as  $\langle P, \leq, \perp \rangle$  by the conditions:

- (i)  $\forall x, y \in P: x \perp y \Rightarrow y \perp x$ , i.e. relation  $\perp$  is symmetric,
- (ii)  $\forall x \in P: x \perp x \Rightarrow \forall y \in P: x \perp y$ ,
- (iii)  $\forall x, y \in P, x \leq y$  iff  $\perp(y) \subseteq \perp(x)$ , where  $\perp(x) = \{z \in P: z \perp x\}$ .

For instance Aristotelian syllogistics can be considered an example of such structure [22].

Thus, we have two mathematical concepts within general algebra representing two ways of understanding information. The uniform mathematical theory underlying these concepts opens them to comparative study.

#### 4. Conclusions

In the extended version of this paper a theorem is provided that shows the way every orthogonality space is associated with a unique closure space. On the other hand we have specific properties for closure spaces to be derived from a generalized orthogonality relation. It turns out that only relatively narrow class of closure spaces can be associated with orthogonality relations. This shows that information formalized with the concept of difference understood as a very general orthogonality relation is a special case of information described in terms of closure spaces.

**Conflicts of Interest:** The author declares no conflict of interest.

#### References

1. Schroeder, M.J. An Alternative to Entropy in the Measurement of Information. *Entropy* **2004**, *6*, 388–412.
2. Shannon, E.C. A mathematical theory of communication. In *The Mathematical Theory of Communication*; Shannon, E.C., Weaver, W., Eds.; University of Illinois Press: Urbana, IL, USA, 1949; p. 20.
3. Bar-Hillel, Y.; Carnap, R. *An Outline of a Theory of Semantic Information*; Technical Report No. 247; Research Laboratory of Electronics, MIT: Cambridge, MA, USA, 1952. In *Language and Information: Selected Essays on Their Theory and Application*; Bar-Hillel, Y., Ed.; Addison-Wesley: Reading, MA, USA, 1964; pp. 221–274.
4. MacKay, D.M. *Information, Mechanism and Meaning*; MIT Press: Cambridge, MA, USA, 1969; p. 163.
5. Schroeder, M.J. Philosophical Foundations for the Concept of Information: Selective and Structural Information. In *Proceedings of the Third International Conference on the Foundations of Information Science, Paris, France, 4–7 July 2005*; MDPI: Basel, Switzerland, 2005. Available online: <http://www.mdpi.org/fis2005/F.58.paper.pdf> (accessed on 15 May 2017).
6. Schroeder, M.J. From Philosophy to Theory of Information. *Int. J. Inf. Theor. Appl.* **2011**, *18*, 56–68.
7. Schroeder, M.J. Towards Autonomous Computation: Geometric Methods of Computing. *Philos. Comput. (Newsl. Am. Philos. Assoc.)* **2015**, *15*, 9–27. Available online: <http://c.ymcdn.com/sites/www.apaonline>.

- [org/resource/collection/EADE8D52-8D02-4136-9A2A-729368501E43/ComputersV15n1.pdf](http://org/resource/collection/EADE8D52-8D02-4136-9A2A-729368501E43/ComputersV15n1.pdf) (accessed on 15 May, 2017).
8. Bateson, G. A Re-examination of 'Bateson's Rule'. *J. Genet.* **1971**, *60*, 230–240.
  9. Bateson, G. *Mind and Nature: A Necessary Unity*; E.P. Dutton: New York, NY, USA, 1979.
  10. MacKay, D.M. *Information, Mechanism and Meaning*; MIT Press: Cambridge, MA, USA, 1969; p. 157.
  11. Bateson, G. Information and Codification: A Philosophical Approach. In *Communication: The Social Matrix of Psychiatry*; Ruesch, J., Bateson, G., Eds.; Norton: New York, NY, USA, 1951; p. 175.
  12. *ibid.* p. 173.
  13. MacKay, D.M. *Information, Mechanism and Meaning*; MIT Press: Cambridge, MA, USA, 1969; p. 161.
  14. *ibid.* p. 165.
  15. *Aristotle: Selections*; Ross, W.D., Ed.; Charles Scribner's Sons: New York, NY, USA, 1955.
  16. Gleick, J. *The Information: A History, A Theory, A Flood*; Pantheon Books: New York, NY, USA, 2011; p. 161.
  17. Weyl, H. *Symmetry*; Princeton University Press: Princeton, NJ, USA, 1952.
  18. Piaget, J. *Structuralism (Le Structuralisme)*; Harper & Row: New York, NY, USA, 1971.
  19. Lévi-Strauss, C.; Jacobson, C.; Schoepf, B.G. *Structural Anthropology*; Translation; Doubleday Anchor Books: New York, NY, USA, 1967.
  20. de Saussure, F. *Course in General Linguistics*; Baskin, W., Transl.; Columbia University Press: New York, NY, USA, 2011; pp. 120–121.
  21. Birkhoff, G. *Lattice Theory*, 3rd ed.; American Mathematical Society Colloquium Publications: Providence, RI, USA, 1967.
  22. Schroeder, M.J. Search for Syllogistic Structure of Semantic Information. *J. Appl. Non-Class. Log.* **2012**, *22*, 83–103.
  23. Schroeder, M.J.; Wright, M.H. Tolerance and weak tolerance relations. *J. Comb. Math. Comb. Comput.* **1992**, *11*, 123–160.
  24. Schroeder, M.J. *Logico-Algebraic Structures for Information Integration in the Brain (Algebras, Languages, Computations and their Applications)*; Research Institute for Mathematical Sciences, RIMS Kokyuroku, Kyoto University: Kyoto, Japan, 2007; Volume 1562, pp. 61–72.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).