

Article

Invariance as a Tool for Ontology of Information

Marcin J. Schroeder

Akita International University, Akita, Japan/193-2 Okutsubakidai, Yuwa, Akita-shi, 010-1211 Akita, Japan; mjs@aiu.ac.jp; Tel.: +81-18-886-5984; Fax: +81-18-886-5910

Academic Editors: Mark Burgin and Wolfgang Hofkirchner

Received: 31 October 2015; Accepted: 25 February 2016; Published: 2 March 2016

Abstract: Attempts to answer questions regarding the ontological status of information are frequently based on the assumption that information should be placed within an already existing framework of concepts of established ontological statuses related to science, in particular to physics. However, many concepts of physics have undetermined or questionable ontological foundations. We can look for a solution in the recognition of the fundamental role of invariance with respect to a change of reference frame and to other transformations as a criterion for objective existence. The importance of invariance (symmetry) as a criterion for a primary ontological status can be identified in the methodology of physics from its beginnings in the work of Galileo, to modern classifications of elementary particles. Thus, the study of the invariance of the theoretical description of information is proposed as the first step towards ontology of information. With the exception of only a few works among publications which set the paradigm of information studies, the issues of invariance were neglected. Orthodox analysis of information lacks conceptual framework for the study of invariance. The present paper shows how invariance can be formalized for the definition of information and, accompanying it, mathematical formalism proposed by the author in his earlier publications.

Keywords: information; ontology of information; invariance of information

1. Introduction

The main goal of this article is to examine invariance of information with respect to transformations and to use the concept of invariance as a tool for the study of its structure and its modes of existence. It is a legitimate question why we can expect any relationship between invariance and existence of information. Before the answer is given, let us consider a more general problem of the way the ontological status of concepts can be established. The questions about what actually exists and how this existence is dependent on the existence of something else are as old as philosophical inquiry and were given a multitude of answers. Several of these answers fell into oblivion due to very clear deficiencies, but some are repeated in spite of their questionable merit. Quite often the only justification is in the common sense view of reality.

The modern common sense understanding of existence is highly eclectic and frequently inconsistent. Thus, in many discussions, with a pretense to being philosophical, we can find statements mixing Aristotelian substance understood as the composition of matter formed together with the Democritean materialistic distinction of matter and void, and with the curious combination of “matter and energy” as substrata of every objectively existent entity that is “physical, not mental” (an echo of Cartesian dualism). Expressions such as “physical reality”, “physical space”, “physical entity”, or statements such as “information is physical”, are used as if they were self-explanatory.

Mixing equivocal concepts of inconsistent philosophical systems is just a matter of ignorance and does not deserve critical analysis. A more complicated issue is the use of epistemic concepts for ontological qualification, such as in the expression “physical reality”. When James Frederick Ferrier, in his 1854 *Institutes of Methaphysic the Theory of Knowing and Being*, introduced his division of philosophy

into epistemology and ontology, the distinction was much simpler and his division could be sharp. Quantum mechanics blurred the division, as the fact of performing a measurement or observation became essentially inseparable from the issue of existence and identity of the object of inquiry.

Of course, this type of philosophical problem born within theories of modern physics is not related to the most frequent errors of the confusion of epistemic and ontological concepts and criteria. Thus, qualification of something (space, reality, entity, *etc.*) as “physical”, because some or many physicists studied it, is meaningless and using this qualification as an ontological criterion is obvious nonsense. After all, physicists studied caloric or aether, not to mention notorious N-rays of René Blondlot. It makes sense to qualify an object of inquiry as physical if it has an empirically testable theory formulated according to the methodology of physics, but such qualification has limited ontological importance. The majority of concepts in physics have multiple, inconsistent theories with very different consequences for ontological interpretation.

Physics, as well as other disciplines of science, such as biology, requires continued revisions of the ontological status of its concepts. Revolutions of relativity and quantum mechanics took place a century ago, but still there is no consensus on their consequences for ontology. On the other hand, developments in modern physics can, and actually do, drive and guide development of ontology or philosophy in general. What can we learn from physics in the matters of ontology without being exposed to the danger that some near future scientific development could falsify it? More extensive discussion of this issue can be found in an earlier publication of the author [1]. For the purpose of the present article, it will be sufficient to consider only one lesson, and this lesson comes not from the content of physical theories, but from the methodology of mathematics and physics. For this reason we do not have to worry about its vulnerability to scientific progress. This lesson came out of the relatively recent developments in mathematics and physics, but its teaching applies to the entire evolution of physics starting from Galileo and, in a more general context, of knowledge from pre-Socratic philosophy.

Pre-Socratic philosophers recognized the role of that which is invariant in the changing world. Even Heraclitus, who believed that everything is changing, sought knowledge in the invariant patterns of changes. This epistemological assumption that the knowable must be invariant was sometimes appended by the ontological claims going much farther, that only that which does not change can exist, but this was just one of many possible positions. The interest in what is not changing was accompanied by the interest in the cosmos, *i.e.*, a harmonious whole and, therefore, in harmony understood as a regular structure. For many centuries these methodological principles stimulated interest in numbers and geometry, but the next essential step required a major revolution in the methods of inquiry. The earliest explicit statement by Galileo of the principle that the description of objective reality has to be invariant with respect to the change of observer (reference frame), who can be in a different place, can measure time differently, or who can move with constant speed, marks the beginning of physics as a scientific discipline.

Newton’s formalization of this rule in his principles of mechanics remained within epistemological considerations. However, both Galileo and Newton contributed to the transition in ontological foundations of physics by revitalization of the atomistic ideas of Democritus, although more in the Epicurean spirit. The Aristotelian concept of ubiquitous matter was replaced by the opposition of matter (existence) and void (nonexistence). This qualitative character of matter was soon replaced by a new quantitative concept. Newton was using the expression “bulk of matter”, but soon the concept of mass appeared. The recognition of the equivalence of the inertial and gravitational mass seemed a good confirmation of its primary ontological status. In the next century, the ancient idea of the qualitative conservation of matter was replaced by the quantitative principle of conservation of mass. Mass gained not only ontic character, but also became eternal.

It took much longer to explicitly formulate the principle of energy conservation, but by the middle of the 19th century the principles of mass, energy, and momentum conservation in isolated systems were ready. Simplicity of the division into materialistic opposition entity—matter characterized

quantitatively by mass *vs.* non-existence—the void was disturbed by the wave theory of light and later by the more general theory of electromagnetic waves. The idea of aether as an exotic form of matter was an attempt to maintain uniformity of entities, which failed when special relativity theory eliminated it. However, the same theory brought the equivalence of mass and energy and the possibility that the entities characterized by mass and energy (particles) can be transformed into entities characterized exclusively by energy (waves of fields) and *vice versa*.

Quantum mechanics and, following it, theories of elementary particles and quantum fields destroyed once again the clarity of the picture. Wave-particle duality became a universal feature of whatever exists giving all entities characteristics of both types, but while some entities are associated with waves of fields which have ontic (*i.e.*, primary) status (e.g., photons as quanta of the electromagnetic field), while some others (e.g., electrons) are associated with the waves of epistemological character (waves of probability distribution or, alternatively, wave functions). Even more disturbing consequence of these developments is the possibility that the void can have non-vanishing energy states. Thus, the void is no more equivalent to non-existence and has to be considered an entity, although of an exotic type.

It is clear that modern physics calls for a new philosophical framework in which the division into epistemology and ontology has to be reconsidered. However, together with the destruction of the traditional framework of philosophical reflection, modern physics brought some new methods of analysis of the high value for philosophy. Probably the most important is the recognition and understanding of the role of invariants of transformations.

A theorem proved by Emmy Noether [2] associates the invariance of the description of motion with respect to transformations with the conservation laws for some magnitudes. Thus, the description (law) of motion is invariant with respect to continuous translations of space (transition between observers in space) or time (change of time coordinate between observers), making momentum, respectively, energy, conserved magnitudes. Symmetry (*i.e.*, invariance) of the law of motion with respect to rotation (orientation of the observer) will result in the conservation of angular momentum. This has extraordinary importance both for physics and for philosophical interpretation. Conservation of energy turns out not to be a discovery of something that already existed independently of our inquiry, but is simply a logical consequence of our requirement that the selection of a starting point of time measuring should not influence the description of motion. Thus, we have an implication: if we want to have descriptions of reality independent from the choice of the observer (reference frame), we should consider energy, momentum, angular momentum, *etc.*, because they will be conserved in this description. This is a theoretical counterpart of the empirical rule of replicability of observations or measurements.

Noether's theorem has some limiting conditions regarding the application to mechanical systems of particular types and to the way they are described (e.g., continuity of transformations), but its role in physics transcends these limits. Additionally, the importance of the study of groups of transformations was not new. The transition from classical mechanics to relativity was already recognized as a change of the group of transformations which preserves dynamical laws.

The recognition of the role of symmetries in mathematics goes back to Felix Klein and his 1872 Erlangen Program [3]. Klein proposed to study geometries through analysis of groups of transformations preserving their fundamental structure. This geometric context was the reason why invariants of transformations became called symmetric and the groups of transformations preserving some structures are called "groups of symmetries".

The program influenced not only mathematics and physics, but became, in the second half of the 20th century, the main source of inspiration for the influential direction of philosophical structuralism. The association of invariance with respect to transformations became the most important tool for the structural analysis in scientific disciplines, in the humanities, and in philosophy. Symmetries of chemical molecules (*i.e.*, groups of spatial transformations preserving their identity as given molecule) became the main tool for physical chemistry, as they turned out to be determinants of chemical

properties of compounds. Thus, we have the following correspondence: group of symmetries—internal structure of molecules—macroscopic chemical properties of substances.

In the study of artificial and natural intelligence, the invariants of the groups of transformations of the configuration of sensory mechanisms were identified with what we humans experience as objects of perception [4]. This last example is of special importance to the subject of the present paper, as the identification of objects of sensory perception is obviously related to the recognition of what actually exists in the phenomenological perspective.

The philosophy of physics is very far from the resolution of many problems in the assessment of the ontological status of the concepts used in theories of modern physics. However, there is no doubt that the most important tool for this task is in the analysis of, and the reflection on, the invariance with respect to groups of transformations which makes the description of physical reality objective.

There is another lesson that we can learn from physics, in particular from its modern developments. The invariance with respect to transformations associated with the changes of observer or, more formally, the changes of the reference frame corresponds to the preservation of the structural characteristics of the objects of inquiry. This suggests that ontological analysis corresponds to structural analysis of the objects of study. This is not a surprise, as existence of an entity is inseparable from its identity and this identity is established by structural characteristics. Thus, when we ask questions about the mode of existence, we have to focus on the structural characteristics. Both mathematics and physics give us extensive methodological knowledge of these matters.

The purpose of this article is to initiate a similar approach in the philosophy of information. The ultimate goal is to develop a tool for the study of its ontology, but also for the study of the structural analysis of information. The literature of the subject of the modes of existence of information is very broad, so the present paper will not attempt to review the large variety of earlier publications addressing the issues related to ontology of information. It would be a formidable, but pointless task, because of many different ways information is defined and understood. The diversity of the ways in which information is defined or understood leads to a large multiplicity of ramifications in its description and study. Moreover, discussions of the ontological status of information are not always carried out with clearly and correctly defined concepts. Thus, it is obvious that the large variety of definitions of information must result in differences in the views on the modes of its existence. Instead, the focus will be on the issue of invariance, with very selective literature references to most important contributions to the discussion of ontology of information, especially those which set paradigms for popular views on information existence. For analysis of the ways in which information studies considered invariance these differences between the ways information is understood or interpreted do not constitute an obstacle. Moreover, the task becomes quite easy as very little was published on information in the context of invariance or symmetry.

The constructive (as opposed to critical) part of the paper, which follows the historical remarks on invariance of information will use the concept of information introduced and elaborated by the author in his earlier publications. Although it is quite different from the variety of concepts used by other authors, its high level of generality justifies identification of virtually every other well-defined concept of information in the literature as its special case. Finally, the invariance of information will be formalized within mathematical formalism developed by the author for his concept of information. It has to be emphasized that this formalism is a result of the choice of particular mathematical foundations in general algebra. The choice is a matter of judgment which mathematical theory can be useful, but not of necessity. The author is aware that another mathematical framework, for instance, category theory, could be used as well.

2. Sources of Problems in Ontology of Information

The term “information” is one of the most recent additions to the catalog of scientific and philosophical vocabulary, but one that generates never ending discussions over its conceptualization and ontological status. Of course, the former issue is epitomized in the question “What is information?”

should be resolved before the latter. However, it is a natural course of the intellectual inquiry that every attempt to define information is first tested against the use of this term in more restricted contexts of specialized disciplines of science where the term “information” already acquired certain meaning, usually informal or intuitive. In these contexts the use of the term suggested some forms of existence of information, although typically without much care for the precise and consistent statement regarding its ontological status.

For instance, Claude Shannon, whose 1948 article on a mathematical theory of communication, republished the next year with the commentary by Warren Weaver in book format to become a paradigm for information theory, uses in the introduction the terms “information” and “message” interchangeably: “In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information. The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point” [5]. A few pages further he writes about quantities of the form of entropy that “play a central role in information theory as measures of information, choice and uncertainty” [6].

Shannon did not write explicitly that the terms “information”, “message”, “choice”, and “uncertainty” are equivalent. Actually, he did not use much the term “information” in his paper and the latter of the statements quoted above where the word “information” has its very infrequent appearance was in the section with the title “Choice, Uncertainty and Entropy” which shows that information was of secondary importance to him. It is clear that whether he considered the concept of information important or not, and whatever he understood by this term, did not have primary ontological status. The association of his entropy as a measure of “information, choice and uncertainty” with physical entropy seemed to be, for him and the early commentators of his work, rather accidental. The loose and indefinite association of information with uncertainty (whatever the choice out of the many possible ways to understand this word) suggested by Shannon was used by interpreters and popularizers of his work (Warren Weaver, Colin Cherry, and others) as a justification for their own interpretation of information as a reduction or resolution of uncertainty. However, in this case, information becomes an epistemological concept.

At first sight, the evolution of the view on the ontological status of information seems straightforward and parallel to the evolution of its epistemological status towards an increasingly fundamental, general, scientific concept. The rediscovery of the 1926 work of Leo Szilard on the solution of Maxwell’s Demon paradox by the association of the increase of physical entropy with cognitive functions of the demon [7], and the book of Erwin Schrödinger “What is Life?” in which life, with its genetic information transmission in reproduction and metabolism, was presented as generated by negative entropy of sunlight, directed attention to the more fundamental relationship between information, physics, and life [8].

The instances of information in biology, in particular, genetics, justified the necessity for disambiguation of the concepts of information and knowledge. Technological advances in computer science directed interests towards the “physics of information”. Computers as physical devices (most likely understood as technological artifacts designed with the use of physics) do not communicate information, but process it through physical operations, which stimulated interest in studying information as a subject of physical inquiry. The most prominent propagator of the idea that “information is physical” was Rolf Landauer, who wrote a series of influential articles on this subject [9–12].

The phrase, which he repeated in some variants in several titles of his papers, became an epitome of the entire direction of the study of information as a physical phenomenon. However, Landauer did not go much farther than others in his views on the ontological status of information in spite of calling it a “physical entity” in the title of his 1999 article, where he writes “Information is inevitably inscribed in a physical medium. It is not an abstract entity. It can be denoted by a hole in a punched card, by the orientation of a nuclear spin, or by the pulses transmitted by a neuron” [11]. The association with physical phenomena in his opinion is through the necessity to represent (sic!) information in

a physical medium. Representation in the physical medium does not entail existence as an object with the ontological status identical with the status of objects considered in physics. It is basically the same view as that of Donald MacKey, elaborated in his 1969 book *Information, Mechanism and Meaning*, expressed in the popular slogan “there is no information without representation” [13]. There is a close resemblance of such a representation to the concept of channels in the original paper of Shannon “The *channel* is merely the medium used to transmit the signal from transmitter to receiver” [5]. The use of the term “representation” is betraying underlying hidden assumptions of a certain “receiver” or “destination” to whom something is presented. In any case, information has, in all these views, only secondary existence dependent on the primary existence of a physical medium.

The actual revolution came with the view expressed by John Archibald Wheeler in his famous epitome “it from bit” [14]. Wheeler not only gave information independent, primary existence, but relegated everything else to the status of secondary one: “Now I am in the grip of a new vision, that everything is Information” [15].

As it can be expected, “it from bit” is not the most popular view on the ontological status of information. Quite frequently it is presented as a mere curiosity, but there are many enthusiasts of this view among physicists. However, the objections to Landauer’s “information is physical” are rare. Unfortunately, the approval of this view is too frequently followed by the question formulated in the anachronistic language of the 19th century popularization of physics: “How is information related to matter and energy”?

This is an expression of the popular conviction that the scientific view of “physical reality” is a safe platform for the ontological analysis of the new concept of information and, in the common sense view of reality, matter and energy are two ontological categories of undeniable primary existence. After all, the entirety of social life is organized around “material goods or resources” and “energy”.

This type of naive escape from the challenges encountered in the study of information to the apparently scientifically-sanctioned common sense view of reality is a stumbling block in attempts to develop a philosophy and science of information. Of course, Wheeler’s “Everything is Information” is not a solution to the question about the ontological status of information, either. After all, “everything” is not an ontological concept and there is no clear presentation in his works of the definition of information. However, Wheeler explicitly addressed the issue of the ontological status of information without sweeping it under the carpet of “physicality”. What is missing in his view of the status of information as a more fundamental entity than traditionally recognized substances is missing also in more restrained views of Landauer and his followers. It is lack of the answer to the question of how to determine and distinguish the ontological status of concepts, such as information, fields, particles, *etc.*

In the introduction to this paper, presenting the problem in the general context, pointed at the analysis of invariance with respect to transformations as a tool for establishing the criteria for ontological status and, for the development of following it, structural analysis. Justification was in the lesson from the methodology of physics. In the following, this line of thinking will be applied to the study of information.

3. Invariance and Structure

In the popular view of physics and other “hard sciences”, the main characteristic of science is its use of quantitative methods. Information theory, as the discipline born in the work of Shannon’s famous article of 1948 [5], became so popular because it introduced a wide range of quantitative methods into the study of communication within many contexts which, earlier, were dominated by qualitative methodology. The best example is psychology. In the popular reception of information theory, its subject is a measure of information, entropy, and its use for a wide range of applications. Apparently, there is no need for qualitative methods of information, as those quantitative methods are superior, more precise, and more useful in applications. This explains why, in the course of dozens of years, so little attention was paid to structural and, therefore, qualitative characteristics of information. At least this lack of interest was common among followers of Shannon’s approach.

There was another, independent direction of the study of information, which programmatically rejected Shannon's approach and formulated its own approach in terms of structural analysis. The most explicit rejection was in the work of Rene Thom in his *Structural Stability and Morphogenesis* [16].

We will start from the question of whether qualitative and quantitative methods of inquiry are necessarily mutually exclusive. Since the category of qualitative methods is frequently defined as "all, which is not quantitative" (the typical view presented in textbooks in statistics), which of course is a gross oversimplification and overgeneralization, we have to specify what the meaning of "qualitative" characteristics of the subject of inquiry are. Otherwise we risk multiple misunderstandings.

In this paper "qualitative" is understood as equivalent to "structural". This means that an object (whatever is its ontological status—"physical", "real", mental, or other entity) has to be considered as a structure built of components in some relationship to each other, and that this structure determines its qualities, understood as modes of external manifestation of the internal structure. Thus, we eliminate from our consideration the issue of qualia and their status. Qualities are expressions of inherent characteristics of the subject of study and they are not dependent on the way of their apprehension. This does not preclude the influence of the qualities on the mutual interactions of objects, or interactions between the object and observer, the latter interaction producing perception of qualities. It is clear that this position is related to the attempts of objectification of the study, which can be identified as the main tenet of the scientific methodology.

The popular view that physics is a purely quantitative discipline comes from its identification with its pre-relativistic and pre-quantum mechanical theories (classical physics) to which typical secondary education is limited. These theories were built around the concept of physical magnitudes which have numerical values in real numbers. These magnitudes represent observables (the numerical values obtained directly from measuring devices) or the results of arithmetic operations of actions of functions on these values. The choice of physical dimension assigned to observables depends on the type of measuring device in experiments. The state of a physical system was described by a complex of the values of observables. In some cases instead of numbers (scalars), the values of observables are vectors or matrices which, in particular, coordinatization have the form of sequences or arrays of real numbers.

Modern physics destroyed this simple vision of physics as a study of physical magnitudes. The state of a physical system is not described as a collection of values of observables anymore. Structural characteristics replaced numerical values. Very different algebraic structures replaced the algebra of real numbers. The governing rule in the choice of the structures was their invariance with respect to the group of transformations; in classical cases the transformations are those already considered by Galileo (Galileo's relativity), in relativistic cases it is a different group (Lorentz group). Traditional separation of quantitative methods and structural (qualitative) methods lost its meaning.

There are many reasons, such as insufficiency of the conceptual framework of physics for biology or for the studies of other complex systems, for more radical changes in theoretical methods of science [1,17,18]. The direction of these changes is away from the traditional quantitative methodology even in higher degrees, and it is pointing at a structural analysis of a new type.

In the following sections, the methodological concept of invariance will be used to analyze the historical relationship between the quantitative and qualitative (structural) characterizations of information in the past research and to attempt setting a bridge between them for further work.

4. Historical Perspective: Hartley

Many historical accounts of information theory consider, as its original source, the paper published by Claude E. Shannon in 1948 or, alternatively, the book published one year later where this paper was followed by explanatory remarks from Warren Weaver [5]. Although the impact of Shannon's paper, and especially of the book, was so great that they were being compared to "a bomb, and something of a delayed-action bomb" [19], if we want to trace the origins of some conceptions and misconceptions regarding information, we have to go twenty years back to the paper "Transmission of Information" by Ralph V. L. Hartley [20]. Even earlier, Harry Nyquist published two papers of great importance

for telegraph transmission problems (both quoted by Shannon together with Hartley's article), but they were not addressing directly the conceptual aspects of general information theory. However, Hartley's contribution in this respect was much more influential for the further direction of the study of information, than it is usually recognized. Some ideas appearing in Hartley's paper had clear resonance in the literature of the subject in several decades. It is very unlikely that it is just a matter of coincidence.

Shannon gave credit to Hartley, but not in the full extent: "I started with information theory, inspired by Hartley's paper, which was a good paper, but it did not take account of things like noise and best encoding and probabilistic aspects" [21]. Definitely, Hartley did not address the issue of noise, explicitly, although he considered distortions. However, he was concerned about the matters of encoding and of probabilistic issues, although his decisions about how to deal with these matters, in each case accompanied with careful explanations, were different from those of Shannon. However, in some cases, he did exactly what Shannon did in his famous book.

He focused, for instance, on the "engineering problem" as can be seen in his statement "In order then for a measure of information to be of practical engineering value it should be of such a nature that the information is proportional to the number of selections" [20]. It is very likely (although it is a pure speculation) that Shannon was influenced by Hartley in writing his famous declaration of disinterest in the matters of info-semantics: "Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain 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*" [5].

Hartley formulated his view in a much more reserved and more elaborated way and was looking for a solution to more general problems (admittedly of less practical value). First, we can see in the first sentence of the abstract of his paper "A quantitative measure of 'information' is developed which is based on physical as contrasted with psychological considerations" [20]. The use of quotation marks for the word "information" seems significant. Then, in the introduction, he presents the purpose of his preoccupation with such a measure: "When we speak of the capacity of a system to transmit information we imply some sort of quantitative measure of information. So commonly used, information is a very elastic term, and it will first be necessary to set up for it a more specific meaning as applied to the present discussion. As a starting place for this let us consider what factors are involved in communication; [. . .] In the first place, there must be a group of physical symbols, such as words, dots and dashes or the like, which by general agreement convey certain meanings to the parties communicating. In any given communication the sender mentally selects a particular symbol and by some bodily motion [. . .] causes the attention of the receiver to be directed to that particular symbol. By successive selections a sequence of symbols is brought to the listener's attention. At each selection there are eliminated all of the other symbols which might have been chosen. [. . .] Inasmuch as the precision of the information depends upon what other symbol sequences might have been chosen it would seem reasonable to hope to find in the number of those sequences the desired quantitative measure of information. The number of symbols available at any one selection obviously varies widely with the type of symbols used, with the particular communicators and with the degree of previous understanding existing between them. [. . .] It is desirable therefore to eliminate the psychological factors involved and to establish a measure of information in terms of purely physical quantities" [20].

In the following section "Elimination of Psychological Factors" Hartley observes that the sequence of symbols can be generated by conscious selection, or by an automatic mechanism as a result of chance operations. On the other hand, the receiver may be either unfamiliar with the code or its parts, or less skilled in distinguishing distorted signals (in the earlier section and in the following one he refers also to communicators using different languages). For this reason, he wants to eliminate from consideration any specific assumptions regarding the generation of symbols.

He writes: "Thus, the number of symbols available to the sending operator at certain of his selections is here limited by psychological rather than physical considerations. Other operators

using other codes might make other selections. Hence in estimating the capacity of the physical system to transmit information we should ignore the question of interpretation, make each selection perfectly arbitrary, and base our result on the possibility of the receiver's distinguishing the result of selecting any one symbol from that of selecting any other. By this means the psychological factors and their variations are eliminated and it becomes possible to set up a definite quantitative measure of information based on physical considerations alone" [20].

Hartley's strong emphasis on the physical considerations was probably motivated by the fact that he wanted to consider that communication means a very wide range of physical phenomena allowing transmission of sounds or pictures. For this purpose he involved in his study the matters of discretization of continuous magnitudes. However, these aspects of his paper will not concern us in the present paper.

It is important to observe that his derivation of the formula for the quantitative measure of information $H = n \times \log_m(s)$, where s is the number of symbols available in all selections and n is a number of selections, m is arbitrarily chosen according to the preferable choice of the unit of information, involved the assumption of invariance. In this particular case, the invariance is with respect to the grouping of the "primary symbols" (here associated with physically distinct states of the physical system) into "secondary symbols" representing psychologically determined and, therefore, subjective symbols carrying meaning. The choice of the formula makes values of H independent from grouping.

For historical reasons, we should notice that Hartley refers to the situation when "non-uniform codes" are used. In this context he observes that the choice of secondary symbols may be restricted, "Such a restriction is imposed when, in computing the average number of dots per character for a non-uniform code, we take account of the average frequency of occurrence of the various characters in telegraph messages" [20]. We can understand it in the post-Shannon perspective, and with some dose of guessing, as the text at this point is not very clear, that he postulates to use the encoding (we know now that it is optimal) by grouping primary symbols in such a way that the differences between frequencies of different characters are compensated and the direct correspondence to the equally likely primary level symbols is restored. Then the difference in the number of selections will not influence our measure of information.

Additionally, Hartley considers the issue of reduction in the number of consecutive choices related to words rather than characters in the context of speech communication: "In speech, for example, we might assume the primary selections to represent the choice of successive words. On that basis s would represent the number of available words. For the first word of a conversation this would correspond to the number of words in the language. For subsequent selections the number would ordinarily be reduced because subsequent words would have to combine in intelligible fashion with those preceding. Such limitations, however, are limitations of interpretation only [. . .]" [20].

Summarizing, we can find in Hartley's article several aspects of the study of information which can be identified with the orthodox approach:

- Information is associated with the selection from a predefined list of choices, and its measure with the number of selections, in each selection with the number of eliminated choices by actual selection (Weaver writes in his contribution to the book with Shannon "To be sure, this word information in communication theory relates not so much to what you *do* say, as to what you *could* say" [22]).
- Information is a subject of engineering, in particular with engineering of communication.
- Information is considered in the context of selections made by a sending operator and by a receiver.
- The meaning of information (necessary for Hartley, but only "frequent" for Shannon) belongs to psychological aspects of information which, because of its variability, has to be eliminated from consideration.
- The measure of information involves a logarithm of the size of a variety of symbols, from which selection (for Hartley) ensures invariance between different ways to encode information.

- Encoding (understood as grouping primary symbols to represent secondary symbols used by communicators) is arbitrary as long as we have equal probability of secondary symbols, but requires some restrictions, otherwise.
- The measure of information is invariant with respect to the permutation of symbols or words as long as we do not change the number of choices in consecutive selections.

Hartley made one important methodological assumption which did not attract much attention in the orthodox approach. It was the use of the concept of invariance, which he applied in the context of the transition to a different encoding understood as grouping of primary symbols into secondary ones. However, his preoccupation with the engineering aspects and emphasis on the elimination of psychological aspects of information prevented him from asking the fundamental question regarding a relationship between the structure of information at the primary level, its transformations, and invariants of these transformations.

5. Historical Perspective: Shannon

Popular perception, amplified by the statements expressed by people involved in the early development of information theory, even by Shannon (as we could see above in his interview), is that Hartley did not consider differences in frequencies of characters in messages. It is clearly a false view. He considered those differences but, right or wrong, as psychological aspects of communication which should be eliminated from consideration.

He wanted to have a measure of information invariant with respect to the change of language. Of course, it made more difficult the discovery of the association between his measure and physical entropy (although, for instance on Boltzmann's grave, the formula for entropy is in its simplified form, corresponding to Hartley's measure). However, Shannon did not recognize the association with physical entropy as significant, either. For him and for many others at that time it was mere curiosity.

John R. Pierce, who co-authored with Shannon and B. Oliver one of the earliest papers on information theory [23], the paper which sometimes is considered more important for the explosion of the new discipline than Shannon's book, even in the 1980 edition of his book popularizing the subject "An Introduction to Information Theory: Symbols, Signals and Noise" reiterated his view from the earlier editions "Here I will merely say that the efforts to marry communication theory and physics have been more interesting than fruitful. Certainly, such attempts have not produced important new results or understanding, as communication theory has in its own right. Communication theory has its origins in the study of electrical communication, not in statistical mechanics, and some of the ideas important to communication theory go back to the very origins of electrical communication" [24].

Someone can defend Pierce's statement, that it is about communication theory, not about information theory, but this statement is in the chapter about the origins of information theory and it follows the paragraph where he writes explicitly about information and explains its meaning "Thus, *information* is sometimes associated with the idea of *knowledge* through its popular use rather than *uncertainty* and the resolution of uncertainty, as it is in communication theory" [24].

Hartley's decision to disregard different frequencies of characters depending on the choice of the language and their order (*i.e.*, structural characteristics of messages) was purely rational, conscious, and can be understood as an expression of his concern to maintain invariance of the measure of information appropriate for the scientific study free from psychological factors. However, this concern cannot defend his position.

While the particular choice of grouping of primary symbols can be considered an arbitrary psychological factor, the fact that his measure of information is an invariant of permutations of selections makes his measure of a questionable value for analysis of information outside of the question of the rate of transmission. After all, is it important to know the value of the measure for a message, if the measure does not change when we list first all a's from the message, then all b's, and so on?

Hartley's error was in the confusion of two different matters. On one side we have the reasons for a particular choice of frequencies of characters controlled by cultural aspects of language development

and evolution, and also the reasons why characters in the strings of characters are ordered in this, or another, way. These are “psychological factors”. However, the fact that characters appear with non-uniform probability, and that the order of the characters matters, are as objective as the fall of an apple from the tree. Thus, they cannot be relegated to irrelevant characteristics of information. Hartley addressed both issues, but focused on the ways how to eliminate them from consideration.

The first error was corrected by Shannon in his revolutionary paper. We do not have to inquire why it happens or what the reasons are for differences between particular instances of languages, but we have to recognize that symbols of natural or artificial languages are subject of some probabilistic distribution which, frequently, is not uniform. For uniform distribution, Hartley’s formula is sufficient. If the distribution is not uniform (and finite) we can generalize it to Shannon’s entropy, which describes the contribution to the measure of information carried by each of the symbols:

$$H(n,p) = -\sum p_i \log_2(p_i), \sum p_i = 1, \text{ with } i = 1, \dots, n \quad (1)$$

However, it does not mean that this is the ultimate solution of the problem. We can see that the measure of information proposed by Shannon is reducing invariance. It is not invariant with respect to replacements of characters by other characters of different probability. Thus, the measure of information becomes dependent on the probability distribution. We have to clarify the role of probability distribution in the formula. Does it describe the information, message, or the language?

The question is less trivial than it may seem at first. Pierce reports the information received from William F. Friedman about curious facts related to the frequency of the use of letters: “Gottlob Burmann, a German poet who lived from 1737 to 1805, wrote 130 poems, including a total of 20,000 words, without once using the letter R. Further, during the last seventeen years of his life, Burmann even omitted the letter from his daily conversation. In each of five stories published by Alonso Alcalá y Herrera in Lisbon in 1641 a different vowel was suppressed. Francisco Navarrete y Ribera (1659), Fernando Jacinto de Zurita y Haro (1654), and Manuel Lorenzo de Lizarazu y Berbuizana (1654) provided other examples. In 1939, Ernest Vincent Wright published a 267-page novel, *Gadsby*, in which no use is made of the letter E” [25].

How much information is being carried by each letter of Wright’s novel? Should we use the probability distribution with no character of probability 0, or that transformed by the exclusion of the letter “e”? Hartley’s concerns seem vindicated.

The process of text generation is governed by the probability distribution which reflects the structural aspects of the text and which reflects much more than just the probability distribution of characters used in statistical research for the English language. Even taking into account idiosyncratic features of someone’s way of expression, such as suppression of some letters, or simple preference, may not be enough.

Shannon was aware of the influence of the structural aspects of language. He recognized the problem of the lack of invariance with respect to the order of generation of symbols (letters or words), but his attempt to solve the problem through the use of conditional probabilities and frequencies of groups of characters of increasing size, or even frequencies of word sequences was inconclusive. His comments on the randomly generated examples of “approximations to English” are disarming in their naiveté: “The particular sequence of ten words ‘attack on an English writer that the character of this’ is not at all unreasonable. It appears then that a sufficiently complex stochastic process will give a satisfactory representation of a discrete source” [26]. It is not satisfactory at all. At least, Shannon did not provide any criterion for his satisfaction and the sequence is just gibberish. That we have parts of it that follow grammatical rules is an obvious result of the fact that probability distribution gives precedence to typical combinations of words, and their typicality comes from being in agreement with the rules of grammar. Non-grammatical sequences have probability close to zero in actual texts. However, this does not change the fact that the sequence in the example does not make any sense.

It is not clear why later on the next page he claimed that we have any reason to believe that the generation of texts should be described as a discrete Markoff process. Grammatical rules of inflection

allow, in some languages, almost unlimited structural permutations. The errors in writing made by dyslexics and everyone's way of reading show that words in a natural language do not function as sequences created letter after letter, where the latter is selected based on the choice of the former or even several preceding ones.

To avoid misunderstanding, there is possibility that, in the human language acquisition, the frequencies of the patterns of words play an important role, as supporters of the Distributional Hypothesis claim. This idea goes back a long time to the studies of Zelig Harris [27] and to more influential, but slightly more general views of John Rupert Firth [28]. However, it is completely unlikely that the actual process of language production has the form which can be associated with a stochastic process, either directly or as a heuristic method of study.

In the general study of information as a concept independent from the specific, "engineering" issues of the theory of communication, Shannon must be given credit for going beyond Hartley's initial insight in developing a powerful tool of probabilistic methods of inquiry. Hartley tried, intentionally, to eliminate the need for taking into account various probability distributions and postulated the use of the uniform distribution. He was aware of the difficulties in the choice of particular distributions, due to their dependence on structural characteristics, which for him belonged to psychological factors. Due to this, he accepted the invariance of the measure of information to allow unlimited structural transformations.

Shannon, twenty years later and with the knowledge of more recent developments in logic and computation theory was aware that the authentic theory of information must take into account its structural characteristics. He disregarded the importance of semantics of information, but in his time this attitude was not unusual. Of course, the assumption that the meaning of information is irrelevant for its study is against our intuition, especially because in the common use of the word information, "information without any meaning", *i.e.*, one which cannot generate knowledge, seems to be an oxymoron.

We have to remember that, in many other disciplines, semantics was relegated to the study of the matters on the mind side of the mind-body problem. Meaning was still associated with intention or "aboutness" which, by Brentano's Thesis, was the main characteristic of the mental, as opposed to the material, and only the latter could be a subject of scientific inquiry in the popular view of the time. Furthermore, the focus of theoretical studies in the related disciplines (for instance logic or linguistic) in the period between the two World Wars was on the syntactic studies.

To avoid philosophical problems with the concept of meaning, various substitute concepts were explored (sense or model). There was no commonly accepted methodology of semantics. Therefore, Shannon's declaration of his disinterest in the meaning, "These semantic aspects of communication are irrelevant to the engineering problem" [5], may seem little bit arrogant, but was in the spirit of the times. Finally, although the criticism of this declaration from the side of those who tried in the next few years to develop semantics for information, such as Yehoshua Bar-Hillel and Rudolf Carnap [29,30] and who, for the reason of the negligence of the semantic issues disqualified Shannon's approach from the status of a theory of information was, in the opinion of the present author, well justified, their attempts had similar weakness. The proposed semantics of information was formulated in purely syntactic terms, as it was pointed out by the present author elsewhere [31].

Shannon's approach in this respect was in the full agreement with logic and, with the recently born new discipline, studying computation. These disciplines were also mainly interested in the structural issues of information. The problem was that the methods he proposed were not very effective, at least outside of the quite specific "ergodic sources" of information and even, in this particular case, he did not propose anything regarding the nature of information beyond the calculation of entropy. This, of course, does not depreciate his tremendous contributions to the study of communication, where these type of issues are irrelevant.

Here, someone can question the phrase "anything regarding the nature of information beyond the calculation of entropy." Why is it not sufficient? Entropy can be calculated for every finite probability distribution, and whenever it is finite for the continuous distributions. Thus, there is nothing in entropy

which was not already in the probability distribution. Moreover, different probability distributions may produce the same entropy. It can be easily verified that very different instances of information, characterized by different probability distributions, have the same quantitative expressions in entropy. Thus, the invariance of entropy here is going beyond significant distinctions of information. We can avoid this problem by making the assumption that information is actually characterized by a probability distribution, entropy is just one of the possible quantitative characteristics of probability distributions and, therefore, also of information. However, in this case, we reduce unwanted invariance only slightly and definitely not sufficiently, and at the same we eliminate information theory. Information theory becomes indistinguishable from the probability theory.

It does not help much if we reformulate probability theory in terms of random variables and refer to entropy as a characteristic of random variables. We can say that a random variable is carrying information and its measure is entropy, but it is just assigning the name “information” to a nebulous concept without any specific meaning. Moreover, in this approach shortcomings of entropy start to be even more visible. In an earlier article, the author advocated the use of a better, alternative, but closely-related measure of information [32]:

$$Inf(n,p) = \sum p_i \log_2(np_i), \sum p_i = 1, \text{ with } i = 1, \dots, n. \quad (2)$$

It can be easily recognized that this measure can be associated with an instance of the Kullback-Leibler measure (but here there is present an important issue why this particular instance), or that it comes from the difference between Hartley’s measure (maximum information irrespective to actual probability distribution) and Shannon’s entropy: $Inf(n,p) = H_{Hartley} - H_{Shannon}$.

Equivalently, Shannon’s entropy is the difference between the maximum of the alternative measure $Inf(n,max)$, which is the case when the probability measure of one particular choice is 1, and other choices have probability of all other choices is 0 (which happens to be equal to $H_{Hartley}$) and the alternative measure for a given probability distribution $Inf(n,p)$.

Shannon’s entropy tells us how much of unknown and potential information we can have in a system, if we already know that the information has some specific form, for instance, enforced by the use of a particular language or particular encoding. Thus, *if* information has some structure going beyond the use of particular letters with the frequency corresponding to that of the English language, this structural information will be accounted for in entropy. *If not*, then entropy is not telling us anything. The problem is that we know only the probability distribution for letters in the particular language of the message. Thus, entropy tells us only about the “space” for information in the system, not how much information is actually there.

This is in complete agreement with Weaver’s statement “To be sure, this word information in communication theory relates not so much to what you *do* say, as to what you *could* say” [22]. It is significant that Shannon referred to redundancy, not entropy, in his rediscovery of the characterization of the languages through their frequencies of characters used by al-Khindi already more than a thousand years earlier for cryptographic purposes.

The author’s earlier articles provided an extensive argumentation for the advantages of the use of the alternative measure, which will be not repeated here [31,32]. It is enough to say that it eliminates problems in the use of the concept of information in physics (where the curious concept of “negentropy” had to be introduced, as a positive magnitude which has the opposite value to the positive entropy, in order to save consistency with observed reality), and eliminates several deficiencies of entropy when applied to random variables, such as its divergence in the limit for transition into continuous distributions, its lack of invariance with respect to linear transformations of coordinates, *etc.*

Unfortunately, but not unexpectedly due to the close relationship with entropy, the alternative measure, itself, does not resolve the problems related to the invariance of either of the measures, which, in both cases, goes way beyond the invariance with respect to structural transformations of information. After all, both measures are for one choice of the character, which is only a small component of the whole, and usually singular characters do not carry meaning, nor can reflect the structure of entire

information. This fact has a natural consequence that the class of transformations determined by the properties of these small components is too large and does not reflect the invariant properties of entire structure. However, when focusing on the information which actually is in the system, not on the available “space” for this information within the variety of possible choices, we can try to find a better description of the relationship between the quantitative and structural characteristics of information.

6. Historical Perspective: From Turing to Kolmogorov and Chaitin

The 1936 paper of Alan Turing opens the new era of computation [33]. This epoch-making paper, together with the paper of Alonzo Church [34] refining his own earlier work, and the slightly earlier published paper of Emil Post [35], opened new ways of thinking about information, although in this time nobody used the term “information” in the context of computation. The expression “information processing” became popular much later.

Turing and Post directly referred to the processes performed by a machine involving manipulation of symbols and, therefore, put their work in the context similar to some extent to that of the work of Hartley. However, the subject of their work was understood in the key terms of logic, numbers, or calculation.

The leading theme of these, and many other works, was Kurt Gödel’s Theorem [36]. While Gödel’s result by ruining Hilbert’s hopes, in hindsight, may seem a “black swan” of Nassim Taleb [37]. It was surprising in its apparent predictability (after all, Peano’s arithmetic involves the Axiom of Choice, and the furious resistance against this axiom at the turn of the century intended to prevent involvement of any concepts which are not results of the finitary, well defined constructions). For the subject of the present paper, Gödel’s Theorem is not so important, although it motivated Turing and Church in their work (the formal objective of their papers was a reproduction of Gödel’s Theorem in different terms), but the method of Gödel numbers which he used in his proof.

Here we can see what was missing in the work of Hartley and what will be soon missing in the work of Shannon. Gödel managed to harness the apparently variable, elusive, and psychologically-determined aspects of the language into arithmetical description. Every expression, no matter how long, if finite, and no matter how complicated of any linearly structured language can be encoded in a unique way by the natural (Gödel) numbers [38]. Moreover, the concepts belonging to the analysis of these expressions (such as being well defined, being in logical relations of inference) can be encoded exactly the same way. Gödel used this encoding to show the existence of some sentences for which neither verification, nor refutation, is possible if, in the theory, we can reconstruct the arithmetic of natural numbers.

Of course, the numbers here do not measure or count anything. However, the structural relations within the text are expressed in the form which can be analyzed the same way as any other arithmetical formulas. This stimulated Church, Turing, and Post to describe processes of arithmetic in as simple a way as possible, and then the latter two authors described these fundamental processes in terms of the work of a simple machine which gets some input number and is producing an output number. Of course numbers can represent arbitrary text encoded, for example, in the form of Gödel’s number, or in some other way. Turing showed that it is possible to design not only an a-machine (automatic machine) for every particular process, but also a unique simple machine, such that its work can be controlled by the input to produce output of any other a-machine applied to the part of the input, a Universal Turing Machine.

This is not only an earlier, but also a much deeper, revolution in the analysis of information, but nobody used in 1936 the word “information”, and the theory of information has its popularly recognized date of birth 12 years later, if we forget or marginalize, as most people today do, the little known contribution of Hartley.

Turing (as well as Church, Post, and others) showed that the structural aspects of information can be examined not necessarily through statistical analysis, and that they are accessible to direct logical and mathematical studies. We can see here not only a methodology for dealing with the structural aspects of information, but also an approach where invariance plays the main role. Turing’s universal

machine is an excellent tool for this purpose. For instance, we can replace the question about the existence of an a-machine achieving some task, by the question whether a universal Turing machine can achieve it. Additionally, to compare two instances of information (if in the format acceptable to the machine) we can compare the inputs, which produce these instances as the machine's output.

Andrey Kolmogorov [39] and Gregory Chaitin [40], published independently and only after earlier publications of Ray Solomonoff [41,42] regarding related ideas, the description of an approach to the measuring of information in a string of symbols (its algorithmic complexity) by the length of the shortest program (input) for a universal Turing machine that produces this string. The standard terminology—algorithmic complexity—is referring to the fact that this measure depends on the structure of information, while Shannon's entropy does not.

We have here a measure of information which is invariant with respect to transformations performed by a universal Turing machine, but these transformations form a structure significantly different from those usually considered in the context of invariance. Instead of the group of transformations (each transformation has its inverse) we have, in this case, only a semigroup. There are many issues regarding the assessment of the meaning of computability (the necessary condition for measurability and application of the method) for some instances of information, which cannot be presented in the form of a finite string of symbols (e.g., irrational numbers). Turing, himself, considered a number computable if its arbitrary finite substring can be produced by a universal machine which gets in its input the number of required digits to be printed (more precisely, in his original 1936 paper he writes about the production of the number's n-th digit).

With this assumption, or without, we have some fundamental problems which are crucial for the domain of artificial intelligence, for the questions about implementation of the work of the brain, *etc.* Since the resistance to the claims of adequate interpretation of the work of brain is well known from works of John R. Searle [43,44], e.g., his Chinese Room Argument, and the arguments given by him and others do not address exactly the issues considered in this paper, I will omit the discussion of these matters. However, it is interesting that Searle uses as an argument against the possibility of consciousness in a device of the type of Turing machine its "multiple realizability", *i.e.*, the fact that a Turing machine can be implemented in many different physical systems. Thus, the invariance of the outcomes of the work of the machine (information) with respect to the changes of physical implementation is an argument against any authentic artificial intelligence.

In earlier papers of the author [45,46] these problems were addressed in the context of the autonomy of a Turing machine, which is relevant here. Can we say that a Turing machine without any involvement of human beings actually performs calculations of the values of functions defined on the set of natural numbers and with their values in the set of natural numbers?

My negative answer to this question was based on the claim that there is nothing in the machine which can put together the sequence of symbols (for instance of 0 s and 1 s) into a whole which is interpreted as a natural number. Moreover, someone who can see only the outcome of the work of the Turing machine cannot say what exactly is the number supposedly produced by the machine (100 can be a binary representation of the number 4 as represented in the decimal system, or can be itself a decimal representation, or can be a Gödel's number standing for some expression from an unknown language). Of course, we are coming back to the issue of meaning which was already a significant concern of Hartley. Yes, we can say that it is not an "engineering problem" or that it is "psychological factor", but it does not solve it, whether it is or is not. This is not the only problem of this type.

The issue can be addressed without any reference to the involvement of human consciousness. In such an approach, we can ask about the integration of information. Physics, more exactly quantum mechanics, provides examples of integrated information in the instances of superposition of states necessary to describe particles which cannot be considered as having definite properties or of states describing entangled particles. However, we do not need quantum physics and we can find more intuitive examples. We have in our human experience many other instances of objects that lose their identity when we separate them into parts.

This calls for the methodology which considers an additional characteristic of information: its level of integration.

7. Duality of Structural and Selective Manifestations of Information

Integration of information was considered the author's earlier publications in the context of two aspects or manifestations of information [47,48]. These two manifestations are always associated with each other in the coexisting dual characteristics of the same concept of information, but in two different, although related, information carriers. Information carrier is a variety or multiplicity standing in the opposition to the one (selected or constructed) for which some mode of transformation into unity (selection or construction) is considered. The original formulation of this definition was that information is an identification of a variety, where this identification can be made by the selection of "one" out of "many" (selective manifestation), or making "one" out of "many" by binding the "many" into a whole "one". The selective manifestation can be characterized in a quantitative manner, for instance through the probability distribution function describing the choice (transition from many to one) and, consequently, by functional magnitude; for instance, of the type of entropy, quantifying ("measuring") this distribution in some respects, or as explained in an earlier section of this article, by the preferred alternative measure [32]. Structural manifestation can be characterized by the level of information integration understood as a degree in which a binding structure can be decomposed into its components (in the mathematical formalism developed earlier by the author it was factorization into the direct product of component structures [48,49]).

The duality can be understood as a consequence of the definition of information introduced above. The only possibility for making the specific selection of an element from the information carrier is that each of the elements of the carrier has some structure consisting of some lower level variety bound into a whole. This structure gives each of the elements an identity (typically described in terms of "properties") allowing directed selection. On the other hand, when we construct a structure from some set of elements, there is a variety of ways how this structure as a whole can be built. The variety of potential structures from which one particular structure is selected forms the upper level information carrier.

In the special, but limited, context of information systems described in the form of a tape for Turing machine, the dual character of these two aspects, structural and dynamic, was utilized in algorithmic complexity measures. It is worth noticing the shift of attention from the length of the computation understood as a number of selections of the values for the current cell or square of the tape (traditional focus of algorithmic complexity) to the length of the input necessary to produce the measured information item expressing the minimal size of the structured input.

Finally, we can formulate what we mean by an information system. It is an information carrier in which the mode of transformation into unity is defined.

This conceptual framework for information is very general. The concept which is used in the definition is a categorical opposition of one and many. This means that this opposition cannot be defined and has to be considered a primitive concept for the theory of information. The high level of abstraction may generate doubts whether it is possible to develop a sufficiently-rich theory of information. The following section will disprove such concerns. The positive aspect of the approach presented in this section is the fact that virtually all clearly-defined concepts of information in the literature can be considered special cases of the concept defined above.

This is quite obvious when we compare the selective manifestation of information described above with all approaches inspired by the work of Shannon. The selection of the one out of many can be described by a probability distribution, or by an instruction within the head of a Turing machine. Similarly, Rene Thon's approach to study information through the structures defined on manifolds can be associated with the structural manifestation of information.

Thus, if we find methods to analyze information in terms of invariance in the conceptual framework presented here, we can extend the study to a wide range of particular instances of information in the literature.

8. Formalism for the Theory of Information

The concept of information defined as identification of a variety can be formalized with the use of concepts of general algebra. More extensive presentation of the formalization of information theory can be found in several of my earlier publications [49,50]. The point of departure in formalization of the duality of information manifestation can be found in the way we are associating information, understood in the linguistic way, with the relation between sets and their elements formally expressed by " $x \in A$ ". The informational aspect of the set theory can be identified in the separation axiom schema, which allows interpretation of $x \in A$ as a statement of some formula $\varphi(x)$ formulated in the predicate logic which is true whenever $x \in A$. The set A consists then of all elements which possess the property expressed by $\varphi(x)$.

If we are interested in a more general concept of information, not necessarily based in any formal language, we can consider a more general relationship than $x \in A$ described by a binary relation R built between the set S and its power set 2^S by the membership of elements of S in the closures $f(A)$ of subsets A of S for some closure operator f . If this closure operator is trivial (for every subset A its closure $f(A) = A$) we get the usual set-theoretical relation of belonging to a set. In a more general case, only closed subsets correspond to properties.

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 \mathfrak{S} satisfying conditions: entire S is in \mathfrak{S} and, together with every subfamily of \mathfrak{S} , its intersection belongs to \mathfrak{S} ; *i.e.*, \mathfrak{S} 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 [51]):

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

The set S with a closure operator f defined on it is usually called a closure space and is represented by the symbol $\langle S, f \rangle$.

The Moore family \mathfrak{S} of subsets is simply the family $f\text{-Cl}$ of all closed subsets, *i.e.*, subsets A of S such that $A = f(A)$. The family of closed subsets $\mathfrak{S} = f\text{-Cl}$ 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 information systems that are not necessarily linguistic, although it does not have to be a Boolean algebra. In many cases it maintains all fundamental characteristics of a logical system [31].

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

The Moore family \mathfrak{S} 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. Filter \mathfrak{S}_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 \mathfrak{S} , 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 , \mathfrak{S}_0 consists of elements in S which have probability measure 1, while \mathfrak{S} is simply the set of all subsets of S .

Now, when we have mathematical formalism for information, we can proceed to formalization of the theory of invariants. For this purpose we can use the well-developed theory of functions between closure spaces preserving their structures, *i.e.*, homomorphisms of closure spaces.

If we have two closure spaces $\langle S, f \rangle$ and $\langle T, g \rangle$, then a function ϕ from closure space $\langle S, f \rangle$ to $\langle T, g \rangle$ is called a homomorphism of closure spaces if it satisfies the condition:

$$\forall A \subseteq S: \phi(f(A)) \subseteq g(\phi(A)).$$

This condition defines continuous functions in the case of topological spaces and, as in topology, for general transitive closure spaces it is equivalent to the requirement that the inverse image of every g -closed subset is f -closed.

Now, when we add a condition that the function ϕ is bijective, we get an isomorphism of closure spaces. Finally, isomorphisms from $\langle S, f \rangle$ on itself (*i.e.*, when $S = T$) are called automorphisms or transformations of closure spaces. It can be easily shown that the class of all automorphisms on a closure space $\langle S, f \rangle$ forms a group with respect to composition of functions.

Now, with the accumulated knowledge of mathematical theory developed for the study of closure spaces [51] we have a complete conceptual toolkit for the study of the invariants of the transformations of information systems.

It is not a surprise that, in addition to the extensive study of topological invariants, the entire Erlangen Program of Felix Klein can be formulated in this mathematical formalism.

9. Conclusions

A closure operator defining an information system is a very general concept, which can be used to define geometric, topological, logical, and algebraic structures. This gives us an opportunity to formalize a very broad class of different types of information associated with geometry, topology, logic, *etc.* The invariance of the description of information is here identical with the invariance of the closure space with respect to transformations preserving its structure. For instance, for topological information, such transformations are continuous functions.

Now, when the toolkit for the study of the transformations of information systems and their invariants is ready, the next task is to apply it to the analysis of more specific instances of information. This task is going beyond the scope of the present paper and will be attempted in future publications.

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

References

- Schroeder, M.J. Crisis in Science: In Search for New Theoretical Foundations. *Prog. Biophys. Mol. Biol.* **2013**, *113*, 25–32. [[CrossRef](#)] [[PubMed](#)]
- Noether, E. Invariante Variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* **1918**, *1918*, 235–257. (In German)
- Klein, F.C. A comparative review of recent researches in geometry. **2008**, arXiv: 0807.3161v1. [[CrossRef](#)]
- Pitts, W.; McCulloch, W.S. How we know universals; the perception of auditory and visual forms. *Bull. Math. Biophys.* **1947**, *9*, 127–147. [[CrossRef](#)] [[PubMed](#)]
- Shannon, E.C. A mathematical theory of communication. In *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, USA, 1949; p. 3.
- Shannon, E.C. A mathematical theory of communication. In *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, USA, 1949; p. 20.
- On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings. In *Quantum Theory and Measurement*; Wheeler, J.A., Zurek, W.H., Eds.; Princeton University Press: Princeton, NJ, USA, 1983; pp. 539–548 (*Engl. Transl.*); Szilard, L. Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Zeitschrift für Physik* **1929**, *53*, 840–856.
- Schrödinger, E. *What is Life?*; Cambridge University Press: Cambridge, UK, 1945.
- Landauer, R. Information is Physical. *Phys. Today* **1991**, *44*, 23–29. [[CrossRef](#)]
- Landauer, R. The Physical Nature of Information. *Phys. Lett. A* **1996**, *217*, 188–193. [[CrossRef](#)]
- Landauer, R. Information is a Physical Entity. *Phys. A* **1999**, *263*, 63–67. [[CrossRef](#)]
- Landauer, R. Information is Inevitably Physical. In *Feynman and Computation: Exploring the Limits of Computers*; Hey, A.J.G., Ed.; Perseus: Reading, MA, USA, 1999; pp. 77–92.

13. MacKay, D.M. *Information, Mechanism and Meaning*; MIT Press: Cambridge, MA, USA, 1969.
14. Wheeler, J.A. Information, Physics, Quantum: The Search for Links. In *Complexity, Entropy, and the Physics of Information*; Zurek, W.H., Ed.; Addison-Wesley: Redwood City, CA, USA, 1990; pp. 3–28.
15. Wheeler, J.A.; Ford, K. *Geons, Black Holes, and Quantum Foam: A Life in Physics*; W. W. Norton: New York, NY, USA; 1998.
16. Thom, R. *Structural Stability and Morphogenesis*; Benjamin: Reading, MA, USA, 1975.
17. Schroeder, M.J. The Role of Information Integration in Demystification of Holistic Methodology. In *Integral Biomathics: Tracing the Road to Reality*; Simeonov, P.L., Smith, L.S., Ehresmann, A.C., Eds.; Springer: Berlin, Germany, 2012; pp. 283–296.
18. Simeonov, P.L.; Brezina, E.H.; Cottam, R.; Ehresmann, A.C.; Gare, A.; Goranson, T.; Gomez-Ramirez, J.; Josephson, B.D.; Marchal, B.; Matsuno, K. Stepping Beyond the Newtonian Paradigm in Biology: Towards an Integrable Model of Life—Accelerating Discovery in the Biological Foundations of Science INBIOSA White Paper. In *Integral Biomathics: Tracing the Road to Reality*; Simeonov, P.L., Smith, L.S., Ehresmann, A.C., Eds.; Springer: Berlin, Germany, 2012; pp. 319–418.
19. Pierce, J.R. The early days of information theory. *IEEE Trans. Inf. Theory* **1973**, *19*, 3–8. [[CrossRef](#)]
20. Hartley, R.V.L. Transmission of Information. *Bell Syst. Tech. J.* **1928**, *7*, 535–563. [[CrossRef](#)]
21. Eilersick, F.W. A conversation with Claude Shannon. *IEEE Commun. Mag.* **1984**, *22*, 123–126. [[CrossRef](#)]
22. Shannon, E.C.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, USA, 1949; p. 100.
23. Oliver, B.; Pierce, J.; Shannon, C.E. The Philosophy of PCM. *Proc. Inst. Radio Eng.* **1948**, *36*, 1324–1331. [[CrossRef](#)]
24. Pierce, J.R. *An Introduction to Information Theory: Symbols, Signals and Noise*; Dover: New York, NY, USA, 1980; p. 24.
25. Pierce, J.R. *An Introduction to Information Theory: Symbols, Signals and Noise*; Dover: New York, NY, USA, 1980; p. 48.
26. Shannon, E.C.; Weaver, W. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, IL, USA, 1949; p. 14.
27. Harris, Z. Distributional structure. *Word* **1954**, *10*, 146–162. [[CrossRef](#)]
28. Firth, J.R. A synopsis of linguistic theory 1930–1955. In *Studies in Linguistic Analysis*; Philological Society: Oxford, UK, 1957; pp. 1–32. Reprinted: *Selected Papers of J. R. Firth 1952–1959*; Palmer, F.R., Ed.; Longman: London, UK, 1968.
29. 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. Reprinted: In *Language and Information: Selected Essays on Their Theory and Application*; Addison-Wesley: Reading, MA, USA, 1964; pp. 221–274.
30. Bar-Hillel, Y. Semantic Information and Its Measures. In *Transactions of the Tenth Conference on Cybernetics*; Foundation: New York, NY, USA, 1952; pp. 33–48. Reprinted: In *Language and Information: Selected Essays on Their Theory and Application*; Addison-Wesley: Reading, MA, USA, 1964; pp. 298–310.
31. Schroeder, M.J. Search for Syllogistic Structure of Semantic Information. *J. Appl. Non-Class. Logic* **2012**, *22*, 83–103. [[CrossRef](#)]
32. Schroeder, M.J. An Alternative to Entropy in the Measurement of Information. *Entropy* **2004**, *6*, 388–412. [[CrossRef](#)]
33. Turing, A.M. On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc. Ser. 2* **1936**, *42*, 230–265, cor. 43, 544–546.
34. Church, A. An unsolvable problem of elementary number theory. *Am. J. Math.* **1936**, *58*, 345–363. [[CrossRef](#)]
35. Post, E.L. Finite Combinatory Processes—Formulation 1. *J. Symb. Logic* **1936**, *1*, 103–105. [[CrossRef](#)]
36. Gödel, K. *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*; Dover: New York, NY, USA, 1992; (Engl. Transl.); Gödel, K. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik* **1931**, *38*, 173–198.
37. Taleb, N.N. *The Black Swan: The Impact of the Highly Improbable*; Random House: New York, NY, USA, 2007.
38. Gödel, K. *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*; Dover: New York, NY, USA, 1992; p. 40 (Engl. Transl.); Gödel, K. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik* **1931**, *38*, 173–198.
39. Kolmogorov, A.N. Three approaches to the definition of the “quantity of information”. *Problemy Peredachi Informatsii* **1965**, *1*, 3–11.
40. Chaitin, G.J. On the Length of Programs for Computing Finite Binary Sequences. *JACM* **1966**, *13*, 547–569. [[CrossRef](#)]

41. Solomonoff, R. *A Preliminary Report on a General Theory of Inductive Inference*; Technical Report ZTB-138; Zator Co.: Cambridge, MA, USA, 1960.
42. Solomonoff, R. A Formal Theory of Inductive Inference. *Inf. Control* **1964**, *7*, 1–22. [[CrossRef](#)]
43. Searle, J.R. *The Rediscovery of the Mind*; MIT Press: Cambridge, MA, USA, 1992.
44. Searle, J.R. *The Mystery of Consciousness*; NYREV: New York, NY, USA, 1997.
45. Schroeder, M.J. From Proactive to Interactive Theory of Computation. In Proceedings of the 6th AISB Symposium on Computing and Philosophy: The Scandal of Computation—What is Computation, Exeter, UK, 2–5 April 2013; pp. 47–51.
46. Schroeder, M.J. Autonomy of Computation and Observer Dependence. In Proceedings of the 7th AISB Symposium on Computing and Philosophy: Is Computation Observer Dependent Bishop, Goldsmiths, UK, 1–4 April 2014. Available online: <http://doc.gold.ac.uk/aisb50/AISB50-S03/AISB50-S3-Schroeder-paper.pdf> (accessed on 26 February 2016).
47. 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. Available online: <http://mdpi.org/fis2005/F.58.paper> (accessed on 26 February 2016).
48. Schroeder, M.J. Quantum Coherence without Quantum Mechanics in Modeling the Unity of Consciousness. In *Quantum Interaction*, Proceedings of the Third International Symposium, QI 2009, Saarbrücken, Germany, 25–27 March 2009; Bruza, P., Sofge, D., Lawless, W., van Rijsbergen, K., Klusch, M., Eds.; Springer: Heidelberg, Germany, 2009; Volume 5494, pp. 97–112.
49. Schroeder, M.J. From Philosophy to Theory of Information. *Int. J. Inf. Theor. Appl.* **2011**, *18*, 56–68.
50. Schroeder, M.J. Algebraic Model for the Dualism of Selective and Structural Manifestations of Information. In *RIMS Kokyuroku*; No. 1915; Kondo, M., Ed.; Research Institute for Mathematical Sciences, Kyoto University: Kyoto, Japan, 2014; pp. 44–52.
51. Birkhoff, G. *Lattice Theory*, 3rd ed.; American Mathematical Society Colloquium Publications: Providence, RI, USA, 1967; Volume XXV.



© 2016 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).