

Proceeding

Physical Information Systems †

John Collier

Philosophy, University of KwaZulu-Natal, Durban 4041, South Africa
Tel.: +27-31-260-3248

† Presented at the IS4SI 2017 Summit DIGITALISATION FOR A SUSTAINABLE SOCIETY, Gothenburg, Sweden, 12-16 June 2017.

Published: date: 9 June 2017

Abstract: Information is usually in strings, like sentences, programs or data for a computer. Information is a much more general concept, however. Here I look at information systems that can be three dimensional, and examine how such systems can be arranged hierarchically so that each level has an associated entropy due to unconstrained information at a lower level. This allows self-organization of such systems, which may be biological, psychological, or social.

Keywords: information; dynamics; self-organization; hierarchy

Information is usually studied in terms of strings of characters, such as in the methods described by Shannon [1] which applies best to the information capacity of a channel, and Kolomogorov [2] and Chaitin [3], which applies best to the information in a system. The basic idea of all three, though they differ somewhat in method, is that the most compressed form of the information in something is a measure of its actual information content, an idea I will clarify shortly. This compressed form gives the minimum number of yes, no questions required to uniquely identify some system [4].

Collier [5] introduced the idea of a physical information system. The idea was to open up information from strings to multidimensional physical systems. Although any system can be mapped onto a (perhaps continuous) string, it struck me that it was more straightforward to go directly to multidimensional systems, as found in, say biology. Even DNA has a tertiary structure that can be functionally important, and thus contains information, though I did not address this issue in my article. A later article [6] showed the hierarchical nature of biological information, and was more specific about the cohesion concept I used in 1986 to refer to dynamical unity.

The basic idea is to use statistical mechanics to describe an informational entropy that can self-organize in a way similar to chemical systems, and apply it to speciation, development, and other biological processes. Although notions such as temperature and free energy have no obvious correlates in evolutionary biology, it is possible to use the organizational and probabilistic aspects of entropy and information to introduce a notion of biological entropy. This is not necessarily the same as thermodynamic entropy, and the exact relation is open to question. What we do need, though, is that the underlying microstates are not correlated relative to the information at higher levels except for their general properties, so they act effectively randomly with respect to those levels. Traditionally, for example, variations in DNA have been thought to be random. Biological systems depend on available free energy, but it is their organization, passed on as information in hereditary processes that is most important. So they are primarily information systems.

Biological information is subject to two sorts of variation producing new information. First, potential information (information at the lower level) can be converted into stored information (expressed at a higher level), creating new expressions in the individual or new stable structures in a

species, perhaps yielding a new species. The second sort of new information is produced by alterations to the genetic structure resulting in new information of one of the four possible types. Both types of new information add new possibilities, in the first case for development and environmental interaction, and in the second case, since it may involve the creation of potential information, for future expression as well. The increase in possibilities is generally faster than they are filled up, producing an information negentropy, or order at the higher level. This permits both order and complexity to increase together. Thus, there are two entropies important in biological systems, the entropy of information and the entropy of cohesion. The information of a system in a particular state is the difference between the maximum possible entropy of the state and its actual entropy. The entropy of information is the amount of information required, given the information of the state, to determine its microstructure. In other words, the entropy of information represents the residual uncertainty about the physical system after the ordering effect of the information contained in the biological system is subtracted.

A physical information system is a system containing stored information whose properties depend only on properties internal to the system. Stored information is like Shannon-Weaver information, except that like bound information it is physically real. It exists whenever there are relatively stable structures which can combine lawfully. These structures are the elements of the information system. The stored information of an element cannot be greater than its bound information (or else either lawfulness or the second law of thermodynamics would be violated), but the actual value is determined by its likelihood of combination with the other elements. The information content of a physical combination of elements (an "array") is the sum of the contributions of the individual elements. For example, the nucleic acids have a structure which contains a certain amount of bound information (they are not just random collections of atoms), and can interact in regular ways with other nucleic acids (as a consequence, but not the only one, of their physical structure). The stored information of a given nucleic acid sequence is determined by the a priori probability of that sequence relative to all the permitted nucleic acid sequences with the same molecules. The bound information, which will be greater, is determined by the probability of the sequence relative to all the random collections of the same molecules. (Nucleic acids, of course, have regular interactions with other structures, as well as to themselves in three dimensions, so the restriction of the information system to just nucleic acid sequences is questionable. We can justify singling out these sequences because of their special role in ontogeny and reproduction.) The lawful (regular) interactions of elements of an information system determine a set of (probabilistic) laws of combination, which we can call the constraints of the information system ([1] p. 38) for a simple example of constraints). Irregular interactions, either among elements of the information system or with external structures, represent noise to the information system.

The elements of an information system, since they are relatively stable, have fixed bound information. It is therefore possible to ignore their bound information in considering entropy variations. The elements are the "atoms" of the system, while the arrays are the states. The stored information of an array is a measure of its unlikelihood given the information system. The entropy (sensu Brillouin [7]) of this unlikelihood equals the entropy of the physical structure of the array minus the entropy of the information system constraints. This value is negative, indicating that the stored information of an array is negentropic. Its absolute value is the product of the redundancy of the information system and the Shannon entropy. Array entropy so calculated reflects more

realistically what can be done with an information system than the Shannon-Weaver entropy. In particular, random alterations to an array make it difficult to recover the array.

This definition of array entropy is inadequate, since it is defined in terms of properties not in the system, namely the entropies of the constraints and the structure constituting the array. The entropy of a system is usually defined in terms of the likelihood of a given macrostate. Two microstates are equivalent macrostates if they have same effect at the macro level (ignoring statistical irregularities). If we assume that all states must be defined internally to the system, the above analysis of arrays does not allow any non-trivial macrostates; each macrostate has just one microstate. This forces a definition of entropy in terms of elements not in the system, or else a "cooked" definition, like Shannon-Weaver entropy. A satisfactory definition of array entropy must be given entirely in terms of the defining physical properties of the information system elements. Such a definition can be given by distinguishing between actual and possible array states.

By assumption, the elements of the system are relatively stable and combine lawfully to form arrays. Possible maximal arrays of elements are the microstates. The macrostates are the actual array states. The microstates of an array are the possible maximal arrays of which it is a part. The information and entropy of a macrostate are defined in the usual way in terms of probabilities of microstates. In abstract information systems this definition degenerates, since arrays can be arbitrarily large. In realistic information systems, though, there is an upper limit on possible array size (though it might be somewhat vague). In organisms the maximum array size is restricted largely by the lengths of the chromosomes. In species it is restricted to the maximum number of characteristics of a member. (There must be such a maximum, since the amount of genetic information is finite.) The array information is a form of bound information, but also has an entropy defined only in terms of the information system characteristics. The external entropy of the null array is the entropy of the constraints on the information system. The external entropy of a maximal array is the base line from which the internal entropy can be measured. It can be called the entropy of the information system. The size of the information system is the difference between these two entropies.

The external entropy of an array is the internal entropy plus the entropy of the information system, equal to the entropy of the constraints minus the array information. The internal entropy of information systems is an extension of the classical statistical entropy of thermodynamic systems. It treats information systems as closed with respect to information but open to matter and energy, whereas mechanical systems are closed if they allow energy to flow in and out of the system, but not matter. The internal entropy of an array is determined by the physically possible ways it could be realized, just as the entropy of a thermodynamic state is determined by its possible microstates. The internal entropy is no less physical than the thermodynamic entropy, unlike the sequence or configurational entropy of Shannon-Weaver information. Array information is a special case of message information, just as bound information is a special case of free information. In this sense it is not anthropomorphic to speak of a biological code or a chemical message.

Codes can be hierarchical. Units concatenated out of elements of a lower level can form natural elements of a higher level. An example is the hierarchy of characters, words and sentences. Sequences of characters terminating with a special character, like a space, comma or period, form possible words. Sequences of words terminated by a period or other sentence terminator form possible sentences. Not all possible words are words, nor are all possible sentences sentences. Otherwise the hierarchy would be trivial. Words are distinguished from non-words by having a meaning or

grammatical function, and sentences are distinguished from non-sentences by being grammatical. Because these properties of words and sentences are useful, words and sentences tend to outnumber other character strings. Some non-words and non-sentences are present in the language, however, which are potentially words or sentences, since they would be so if they fell into common use.

Brillouin [7] p. 55 points out that a more efficient code for English would exploit the fact that not all potential words are words by encoding words so as to permit fewer non-words. The information required per character could be reduced by a factor of more than two, yet the same amount of information could be conveyed by the same number of characters. An even larger reduction could be achieved by eliminating potential sentences, and even more, no doubt, by eliminating unverifiable sentences. This would not only make language learning difficult, but would also reduce the likelihood of change in the language.

Using Brooks-Wiley (1986) terminology, the distinguished set of higher level messages contain the stored information of the information system, while the variants contain the potential information. The stored information is what distinguishes a system from other systems. In physical information systems the basis of the individuation must be some physical property.

Acknowledgments: This work was partially supported by an HCID grant from the National Research Foundation of South Africa from 2005-2020.

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

References

1. Shannon, C.E. and Weaver, W., *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL, USA 1949.
2. Kolmogorov, A.N., Three Approaches to the Quantitative Definition of Information. *Problems of Information Transmission* 1, pp. 1-7, 1965.
3. Chaitin, G. Chaitin, Gregory J., *Algorithmic Information Theory*. Cambridge University Press, Cambridge, UK 1987.
4. Collier, John. Information, causation and computation. In Gordana Dodig Crnkovic and Mark Burgin (eds) *Information and Computation: Essays on Scientific and Philosophical Understanding of Foundations of Information and Computation*, World Scientific, Singapore 2012.
5. Collier, John.. Entropy in Evolution. *Biology and Philosophy*, 1, pp. 5-24, 1986.
6. Collier, John. Hierarchical dynamical information systems with a focus on biology. *Entropy* 5, pp. 100-124, 2003.
7. Brillouin, L. *Science and Information Theory*, Academic Press, New York, 1962.



© 2017 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>)