Complex systems are always correlated but rarely information processing

Karoline Wiesner1,∗ and James Ladyman2

1 Institut für Physik und Astronomie, Universität Potsdam, Campus Golm, Haus 28, Karl-Liebknecht-Straße 24/25, 14476 Potsdam-Golm, Germany
2 Department of Philosophy, University of Bristol, Cotham House, Bristol BS6 6JL, United Kingdom
∗ Author to whom any correspondence should be addressed.
E-mail: karoline.wiesner@uni-potsdam.de and james.ladyman@bristol.ac.uk

Keywords: correlations, information theory, complex systems, information processing, self-organisation

Abstract

‘Complex systems are information processors’ is a statement that is frequently made. Here we argue for the distinction between information processing—in the sense of encoding and transmitting a symbolic representation—and the formation of correlations (pattern formation/self-organisation). The study of both uses tools from information theory, but the purpose is very different in each case: explaining the mechanisms and understanding the purpose or function in the first case, versus data analysis and correlation extraction in the latter. We give examples of both and discuss some open questions. The distinction helps focus research efforts on the relevant questions in each case.

1. Information processing and correlation formation are often taken to be the same when they are not

The literature in mathematics and the natural sciences abounds with examples of systems being interpreted as information processors or as computational devices, for instance, in dynamical systems theory [1], in neuroscience [2], and in cellular biology [3]. Complexity scientists are particularly drawn to this way of thinking. This focus issue on ‘complex systems approaches to information processing’ is a good illustration of that. However, taking the two requirements for information processing seriously—information is encoded symbolically and computation is independent of architecture—one ends up with two kinds of information-theoretic approaches in general and to complex systems in particular: one is using the language and tools of information theory to analyse a system in terms of its symbolic code and the input and output of its computation. This approach studies information processing systems. The other approach is using information theoretic tools to quantify the correlations that are generated by or present in the system. This approach studies correlation formation in complex system. Correlations are order, and the generation of correlations in a system without a central controller is called self-organisation. The purpose of this article is to clarify the difference between these two types of systems and the respective approaches to study them. It should be noted that information processing systems include those that are self-organising.

We consider examples of complex systems that have been studied in detail with each of these two approaches. We start with a complex system that we consider an information processing system. Probably the prime example of an information processing complex system is eukaryotes with their genetic information encoded in DNA and their cellular information processing machinery that reads out the code and translates it back into chemical instructions for protein synthesis. The reason DNA can indeed be thought of as storing information is that the information is encoded symbolically, and a decoding mechanism is required for read-out (hence the use of the term ‘codon’ for sequences of nucleotide triplets).3

3 The difference between protein synthesis from DNA and human-made computers is that there is no clear distinction between hardware and software in the cell, a point which we do not discuss any further. The interested reader is referred to, for example, [4] and references therein.
An example of a complex system in which correlations are forming in space and time, but which we do not consider an information-processing system, is a chemical mixture in which a Belouzov–Zhabotinsky reaction takes place. This is a class of chemical reactions involving bromine and an acid\(^4\). Through a constant influx of molecular matter, the chemical reactions lead to oscillatory patterns that are clearly visible. These patterns are correlations that form in space, in the spirals of varying colour, and in time, in the spirals moving outwards. It is an example of self-organisation since it is through the many reactions between the many molecules that the spiral patterns emerge.

In ‘what is a complex system?’ Ladyman and Wiesner formulated nine truisms of complexity [6]. They called these statements ‘truisms’ since they are generally agreed upon in the field, as opposed to many other statements involving complexity which are disputed, not least its definition. One truism is as follows: ‘complex systems are often modelled as [⋯] information processing systems’. A second truism is ‘non-living systems can generate order’. Information theory is used in practice to study both cases. But the two cases are different, though some systems can be modelled as information processing systems and also generate order. To clearly distinguish between those systems that form correlations (self-organise) and those that can be considered information processors, not only structures current research, but opens opportunities for new research strategies as we discuss at the end.

2. Mutual information is the most general measure of correlation

In Shannon’s original formulation of his mathematical theory of communication, the production of information is the reduction of uncertainty held by an observer with respect to the occurrence of an actual event from a set of possible events [7]. Its measure is the Shannon entropy \(H(X)\) of a random variable \(X\) taking on discrete values \(x\) [7]:

\[
H(X) := - \sum_x P(x) \log P(x). \tag{1}
\]

The Shannon entropy is a measure of uncertainty in the outcome of sampling the probability distribution \(P\) of random variable \(X\). Based on this definition, Shannon defined the mutual information \(I(X : Y)\) between two random variables \(X\) and \(Y\) taking on values \(x\) and \(y\), respectively, as follows:

\[
I(X : Y) := \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)}. \tag{2}
\]

The mutual information measures the amount of uncertainty reduction in sampling one variable after sampling the other. The mutual information plays a central role in Shannon’s coding theory. His definition of the communication channel capacity is based on the mutual information. One of Shannon’s great achievements was to develop a theory for channel coding that is still at the heart of all digital communication algorithms today. It is presumably this fact that makes it tempting to call systems with non-zero mutual information ‘information processing’. From a purely mathematical point of view, however, mutual information is nothing but a measure of correlation. In fact, it is the most general measure, since it detects any form of correlation [8], not only linear (Pearson correlation) or monotonic (Spearman correlation) correlation. The reason it is not used as often in data analysis as one might expect is that the mutual information is a ‘data hungry’ function, requiring a large amount of data to be estimated with accuracy.

The mutual information has been interpreted as a measure of predictability when computed between two time sequences. Here, the two random variables \(X\) and \(Y\) are replaced by sequences of variables. If the two sequences have non-zero mutual information, then one is to some extent predictable from the other. Mutual information is symmetric in \(X\) and \(Y\), but labelling one time sequence as past and one as future gives a quantification of predictable future behaviour. The predictive information, introduced by Bialek et al [9] is exactly that: an interpretation of the mutual information as predictability of a future time sequence from the past. The same measure has been introduced under different names: the excess entropy, introduced by Crutchfield and Feldman in 2003 [10] and initially as the effective measure complexity by Grassberger in 1986 [11].

3. How information processing is different from the formation of correlation

A positive measure of mutual information is taken to be grounds for taking a system to be processing information (for example, the heart [12] or the Universe [13]). However, this is not justified, because, there are two ingredients to ‘information processing’: the ‘information’ and the ‘processing’. The main ingredient to

\(^4\) For an introduction into the mathematics of the Belouzov–Zhabotinsky reaction, see, for example [5]).
Shannon’s theory of communication is a set of messages that are encoded symbolically as sequences of symbols (usually zeros and ones). A message contains or represents information if its transmission (followed by reception and decoding) reduces the uncertainty of the receiver. Thus, in addition to non-zero mutual information, information transmission requires symbolic encoding and a receiver to whom or which that information is relevant.

Studying information processing in complex systems is interesting, because it reveals the functional nature of the system. Information processing in natural systems presupposes a symbolic (abstract) representation of something in the physical world. Thus, it also presupposes the presence of a receiver of that symbolic representation and a purpose or function with respect to which the representation is reducing the uncertainty. The field of computational neuroscience [2] works on that basis.

The brain is the canonical example of a complex system that processes information. A specific case of information processing are the neural spike sequences triggered by a visual input and, potentially, leading to some motor output. Palmer et al [14] measure the mutual information in neural spike sequences of salamanders that are exposed to visual stimuli in the form of varying videos. When viewing a video of a natural scene, the mutual information is higher than it is when the salamander is viewing a purely noisy image. This result indicates that the order in the natural scene is encoded symbolically in the spike sequence, which is a form of information processing. Furthermore, the authors suggest that the salamander brain is predicting future visual stimuli, which would be a form of computation. Whether this is true or not, what they have done is to identify patterns, using the language and tools of information theory. This can now be handed back to physiologists to understand the physical and chemical mechanisms and to find their causes.

An example of a tool set in which information theory is used explicitly to extract the information processing qualities is Friston’s ‘free energy principle’ [15]. It attempts the mathematical representation of the prediction process in the neural system using a version of the mutual information (the Kullback–Leibler divergence). Although there is a well justified critique of this approach [16], it is an example of information processing being the object of study, not merely the formation of correlations. In our view, it is here that the notion of information processing of complex systems is at its best. Nobel Laureate nurse argues for this in the opinion piece: ‘focussing on information flow will help us to understand better how cells and organisms work’ [3].

We contrast information processing with the formation of correlation, which may also be called pattern formation or self-organisation, as the process by which correlations are generated, either in time or in space or both. Self-organisation can happen in living or functional systems as well as in non-functional systems (a distinction made in [6], where functional systems include the living systems). The presence of correlations can be measured, and often this is done using the mutual information.

One way to measure ‘self-organisation’ is to measure the amount of structure that is generated or at least the amount of structure that is present with the assumption that it is the result of self-organisation. There are plenty of examples of this in the literature. For example, Shalizi et al [17] measure self-organisation in a cellular automaton by the amount of memory that would need to be stored to generate that same amount of structure computationally. Putting the question aside whether a cellular automaton is a complex system (we do not think it is since it is a humanly programed object), the authors use information and computation (finite-state automata in this case) to represent the system in question and to measure the amount of order that is generated in the self-organisation process. This does not imply, however, that the system is computing in the sense of the above notion of computation.

It is possible to quantify the ‘information flow’ between the scales in this system. For an introduction, see, for example, Lindgren’s lecture notes [18] and references therein. However, Lindgren does not adopt the interpretation that chemical systems are processing information, but instead advocates information theory as a powerful tool for describing, quantifying, and analysing the phenomenon of self-organisation, and rightly so.

4. Information theory provides powerful tools for data analysis

One of the strengths of complexity science is its technique of abstracting away the physical or chemical particularities of a system and studying an abstracted version with suitable mathematical and computational tools such as network analysis, dynamical systems theory, or statistical data analysis [6]. It is likely that it is for this reason that complex systems are often called ‘information processing systems’, since, for something to be an information processor, the architecture (hardware) in which the information is stored and its processing is taking place must be irrelevant for the outcome of the computation [19].

Focussing on information theory as a tool for statistical analysis is also often very fruitful. Mutual information is the most general measure of correlations, and it is applicable to discrete as well as continuous variables. Thus, it is a tool immediately available for correlation analysis and is much used, not least in bioinformatics [20]. Another example of a data-analytic tool that makes use of information theory is the infomap principle,
introduced by Rosvall and Bergstrom [21]. It is a clustering technique for complex networks that is based on coding theory. It has become one of the most popular tools in clustering analysis. Other examples are the maximal information coefficient [22] or the use of mutual information in medical image processing [23]. None of these tools assume that the system in question is processing information. In this context, information theory merely provides the tools for extracting the correlations in the systems.

In summary, information processing is distinct from self-organisation. As a consequence there are two uses of information theory in the study of complex systems. One is modelling the system as a computational device with input, output, and symbolic encoding. It can serve as intermediary between observation and functional explanation. The other use of information theory is as a data-analytic tool set. Both uses are important, but they are different.

Data availability statement

No new data were created or analysed in this study.

ORCID iDs

Karoline Wiesner https://orcid.org/0000-0003-2944-1988

References

[1] Crutchfield J P, Ditto W L and Sinha S 2010 Introduction to focus issue: intrinsic and designed computation: information processing in dynamical systems—beyond the digital hegemony Chaos 20 037101
[2] Kriegeskorte N and Douglas P K 2018 Nat. Neurosci. 21 1148–60
[3] Nurse P 2008 Nature 454 424–6
[4] Keller E F 2011 Stud. Hist. Phil. Sci. C 42 174–9
[5] Strogatz S H 2014 Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering (Boca Raton, FL: Westview Press)
[6] Ladyman J and Wiesner K 2020 What is a Complex System? (New Haven, CT: Yale University Press)
[7] Shannon C E 1948 A mathematical theory of communication Technical Report Bell Labs
[8] Li W 1990 J. Stat. Phys. 60 823–37
[9] Bialek W, Nemenman I and Tishby N 2001 Neural Comput. 13 2409–63
[10] Crutchfield J P and Feldman D P 2003 Chaos 13 25–54
[11] Grassberger P 1986 Int. J. Theor. Phys. 25 907–38
[12] Ashikaga H, Aguilar-Rodríguez J, Gorsky S, Lusczek E, Marquitti F M D, Thompson B, Wu D and Garland J 2015 J. R. Soc. Interface. 12 20141201
[13] Lloyd S 2006 Programming the Universe (Vintage, New York: Knopf)
[14] Palmer S E, Marre O, Berry M J and Bialek W 2015 Proc. Natl Acad. Sci. USA 112 6908–13
[15] Friston K 2010 Nat. Rev. Neurosci. 11 127
[16] Biehl M, Pollock F A and Kanai R 2020 arXiv:2001.06408
[17] Shalizi C R, Shalizi K L and Haslinger R 2004 Phys. Rev. Lett. 93 118701
[18] Lindgren K 2003 Information Theory for Complex Systems Lecture Notes (Department of Physical Resource Theory, Chalmers and Göteborg University) (updated in 2014)
[19] Simon H A and Newell A 1964 Am. Sci. 52 281–300
[20] Steuer R, Kurths J, Daub C O, Weise J and Selbig J 2002 Bioinformatics 18 S231–40
[21] Rosvall M and Bergström C T 2007 Proc. Natl Acad. Sci. 104 7327–31
[22] Reshef D N, Reshef Y A, Finucane H K, Grossman S R, McVean G, Turnbaugh P J, Lande R E S, Mitzenmacher M and Sabeti P C 2011 Science 334 1518–24
[23] Pluim J P W, Maintz J B A and Viergever M A 2003 IEEE Trans. Med. Imaging 22 986–1004