Foundations and Formalizations of Self-Organization

Daniel Polani

Adaptive Systems Research Group
Department of Computer Science
University of Hertfordshire
d.polani@herts.ac.uk

1 Introduction

In the study of complex systems, the relevance of the phenomenon of *self-organization* is ubiquitous. Whether it is the stripe formation in morphogenesis (Meinhardt 1972, 1982), in reaction-diffusion automata (Turing 1952) or the reorganization of a self-organizing Kohonen map, whether the seemingly effortless distributed organization of work in an ant colony or the formation of flows in pedestrian movement patterns (Helbing et al. 2005), the maintenance and creation of the complexities involved with the maintenance of life in living cells, all of these systems produce a behaviour that, in one way or another, can be called "organized" and if the source of organization is not explicitly identified outside of the system, also "*self*-organized".

Strangely enough, as much agreement as there is about whether self-organization is present or absent in a system on visual inspection, as little agreement exists concerning the precise meaning of the word. In other words, while the phenomenology of the phenomenon is pretty much agreed upon, its formal foundations are not.

Among other problems, this causes a certain amount of confusion. For instance, the difference between the notions of emergence and self-organization are being strongly emphasized (Shalizi 2001), notwithstanding the frequent co-occurrence of these notions. On the other hand, without a clean formal definition of self-organization and emergence, it is difficult to make strong points in favour (or against) a separation of these notions.

With recent interest in the exploitation of phenomena of self-organization for engineering and other applications, the importance of characterizing and understanding the phenomenon of self-organization has even increased. It is not sufficient anymore to characterize a system "ivory-tower"-like to be in one group or another according to some human classification. Rather it becomes necessary to work towards a predictable and structured theory that will also allow to make useful predictions about the performance of a system. The advent of nanotechnology and bioinspired engineering architectures increases the immediate practical relevance of understanding and characterizing self-organization.

In view of the various streams and directions of the field of self-organization, it is beyond the present introductory chapter to review all the currents of research in the field. Rather, the aim of the present section is to address some of the points judged as most relevant and to provide a discussion of suitable candidate formalisms for the treatment of self-organization. In the author's opinion, discussing formalisms is not just a vain exercise, but allows one to isolate the essence of the notion one wishes to develop. Thus even if one disagrees with the path taken (as is common in the case of not yet universally agreed upon formal notions), starting from operational formalisms helps to serve as a compass guiding one towards notions suitable for one's purposes. This is the philosophy of the present chapter.

The chapter is structured as follows: in Sec. 2, we will present several central conceptual issues relevant in the context of self-organization. Some historical remarks about related relevant work are then done in Sec. 3. To illustrate the setting, a brief overview over some classical examples for self-organizing processes is given in Sec. 4. In Secs. 5 and 6, introduces the two main information-theoretic concepts of self-organization that the present chapter aims to discuss. One concept, based on the ϵ-machine formalism by Crutchfield and Shalizi, introduces self-organization as an increase of (statistical) complexity with time. The other concept will suggest measuring self-organization as an increase of mutual correlations (measured by multi-information) between different components of a system. In Sec. 7, finally, important properties of these two measures as well as their distinctive characteristics (namely their power to identify temporal versus compositional self-organization) will be discussed, before Sec. 8 gives some conclusive remarks.

2 General Comments

In the vein of the comments made above, the present paper does not attempt to answer the question: "what is self-organization?" Rather, the question will be "where do we agree self-organization exists?" or "what are candidate characterizations of self-organization?". Thus, rather than attempting to give ultimate answers to that question, a number of suggestions will be presented that can form a starting point for the development of the reader's own notion.

The distinction of self-organization from emergence is emphasized time and time again (Shalizi 2001); this is sometimes confusing to outsiders, given that both often appear in similar contexts and that there is no universally accepted formal definition for each of them. This distinction emphasizes, though, that there is a general desire to have them carry different meanings.

Self-organization, the main focus of the present chapter[1], is a phenomenon under which a dynamical system exhibits the tendency to create organization "out of itself", without being driven by an external system, in particular, not in a "top-down" way. This requires clarification of several questions:

[1] Emergence is briefly discussed in Secs. 5.2 and 6.1.

1. What is meant by organization?
2. How and when to distinguish system and environment?
3. How can external drives be measured?
4. What does top-down mean?

Question 1 is clearly related to the organizational concept of *entropy*. However, it is surprisingly “unstraightforward” to adapt entropy to be useful for measuring self-organization, and some additional efforts must be made (Polani 2003). This is the main question the present chapter concentrates upon. All the other questions are only briefly mentioned here to provide the reader with a feel of what further issues in the context of formalization of self-organization could and should be addressed in future.

Question 2 is, basically, about how one defines the boundaries of the system; this is related with the question of autonomy (Bertschinger et al. 2006). Once they are defined, we can ask ourselves all kinds of questions about the system under investigation and its environment, its “outside”. A complication is brought into the discussion through the fact that, if organization is produced in the “inner” system out of itself (the “self” in self-organization), in a real physical system, disorder has to be produced in the outside world, due to the conservation of phase space volume (Adami 1998).

However, for many so-called self-organizing systems the physical reality is quite detached, i.e. conservation or other thermodynamic laws are not (and need not be) part of the model dynamics, so Landauer’s principles (Landauer 1961; Bennett and Landauer 1985) are irrelevant in the general scenario: for instance, a computer emulating a self-organizing map produces much more total heat in its computation than the minimum amount demanded by Landauer’s principle due to the entropy reduction of the pure computation. In other words, the systems under consideration may be arbitrarily far away from thermal equilibrium and worse, there may be a whole hierarchy of different levels of organization whose constraints and invariants have to be respected before one can even consider coming close to the Landauer limit[2].

Therefore, unless we are aiming for understanding nanosystems where issues of Landauer’s principle could begin to play a role, we can and will ignore issues of the “compulsory” entropy production of a real physical system that exhibits self-organization. In particular, the systems we will consider in the following are general dynamical systems. We will not require them to be modeled in a thermodynamically or energetically consistent way.

As for question 3, it is not as straightforward to respond to, a difficulty that is conceded in (Shalizi et al. 2004). There, it has been suggested to study causal inference as a possible formalism to investigate the influence of an environment onto a given system. In fact, the concept of *information flow* has been recently introduced to address exactly this question (Ay and Wennekers 2003; Klyubin et al. 2004; Ay and Krakauer 2006; Ay and Polani 2006), providing an information-theoretic approach to measure causal inference. At this point these notions are still quite fresh and not much is known about their possible relevance for characterizing self-organizing systems, although it will be an interesting avenue for future work.

[2] As an example, the energy balance of real biological computation process will operate at the ATP metabolism level and respect its restrictions — but this is still far off the Landauer limit.

Question 4, again introduces an interesting and at the same time unclear notion of “top-down”. Roughly, top-down indicates a kind of downward causation (Emmeche et al. 2000), where one intervenes to influence some coarse-grained, global, degrees of freedom as to achieve a particular organizational outcome; or else, the external experimenter “micromanages” the system into a particular state. For this latter view, one could consider using a formalization of an agent manipulating its environment (Klyubin et al. 2004). This is again outside of the scope of the present paper. Nevertheless, it is hoped that this section’s brief discussion of general conceptual issues highlights some related open questions of interest that may prove amenable to treatment by a consistent theoretical framework.

3 Related Work and Historical Remarks

Shalizi et al. (2004) track back the first use of the notion of “self-organizing systems” to Ashby (1947). The bottom-up cybernetic approach of early artificial intelligence (Walter 1951; Pask 1960) devoted considerable interest and attention to the area of self-organizing systems; many of the questions and methods considered relevant today have been appropriately identified almost half a century ago (e.g. Yovits and Cameron 1960).

The notions of *self-organization* and the related notion of *emergence* form the backbone for the studies of dynamical hierarchies, and in particular those types of dynamics that lead to climbing the ladder of complexity as found in nature. Notwithstanding the importance and frequent use of these notions in the relevant literature, a both precise and useful mathematical definition remains elusive. While there is a high degree of intuitive consensus on what type of phenomena should be called “self-organizing” or “emergent”, the prevailing strategy of characterization is along the line of “I know it when I see it” (Harvey 2000).

Specialized formal literature often does not go beyond pragmatic characterizations; e.g. Jetschke (1989) defines a system as undergoing a self-organizing transition if the symmetry group of its dynamics changes to a less symmetrical one (e.g. a subgroup of the original symmetry group, Golubitsky and Stewart 2003), typically occurring at phase transitions (Reichl 1980). This latter view relates self-organization to phase transitions. However, there are several reasons to approach the definition of self-organization in a different way. The typical complex system is not in thermodynamic equilibrium (see also Prigogine and Nicolis 1977). One possible extension of the formalism is towards nonequilibrium thermodynamics, identifying phase transitions by *order parameters*. These are quantities that characterize the “deviation” of the system in a more organized state (in the sense of Jetschke) from the system in a less organized state, measured by absence or presence of symmetries. Order parameters have to be constructed by explicit inspection of the system since a generic approach is not available, although an ϵ-machine based approach such as in (Shalizi 2001) seems promising. Also, in complex systems, one can not expect the a priori existence or absence of any symmetry to act as universal indicators for self-organization; in general such a system will exhibit, at best, only approximate or imperfect symmetries, if at all.

Without a well-founded concept of characterizing such imperfect "soft" symmetries, the symmetry approach to characterize self-organization is not sufficient to characterize general complex systems.

4 Examples for Self-Organization

We now consider a number of examples of systems which are typically regarded as self-organizing. Since the field lacks a consensus on suitable formal definitions, it is helpful to consider examples of the phenomenon at hand, where there is less controversy whether or not they exhibit the desired behaviour.

4.1 Bifurcation

Consider a dynamical system with state $\mathbf{x}(t) \in \mathbb{R}^n$ at time t, whose state dynamics governed by a differential equation

$$\dot{\mathbf{x}} = F(\mathbf{x}, \mu) \tag{1}$$

where $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ is a smooth function, and $\mu \in \mathbb{R}$ is a so-called *bifurcation parameter*, and the dot denotes the usual time derivative. For a fixed μ, this defines a particular dynamical system which, amongst other, exhibits a particular *fixed point* profile, i.e. a set of points $\{\mathbf{x}^0 \in \mathbb{R}^n \mid F(\mathbf{x}^0) = 0\}$. The existence of fixed points engenders a possibly highly intricate structure of the system state space $\mathbb{R}^n$. Of particular importance are the so-called *stable* and *unstable manifolds*. The *stable manifold* of a fixed point $\mathbf{x}^0$ is the continuation (forward in time) of the local eigenspaces of the Jacobian $DF|_{\mathbf{x}^0}$ for negative eigenvalues, the *unstable manifold* is the continuation (backward in time) for positive eigenvalues (see Jetschke 1989, or any good book about dynamical systems for details). The important part of this message is that the structure of the invariant (stable and unstable) manifolds structures the state space in a characteristic way. A very simple example is shown in Fig. 1: even in this simple example, the state space is split into four regions. With a more intricate fixed points (or attractor) structure, that profile can be quite more complex.

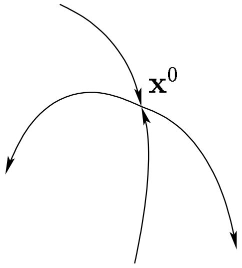

Fig. 1. Stable and unstable manifold of a fixed point $\mathbf{x}^0$

The importance of above observation stems from a number of facts. In the example above (as shown in Fig. 1), there are only positive or negative eigenvalues of the Jacobian $DF|_{\mathbf{x}^0}$. However, if we consider μ a parameter that is scanned through, the eigenvalue spectrum of the Jacobian changes smoothly, and eigenvalues may change sign, i.e. may crossing the 0 level. Generically, an eigenvalue changing sign on changing μ will travel with nonzero speed through 0, i.e. for which $DF_\mu|_{\mathbf{x}^0} \neq 0$, where DF_μ is the partial derivative of the Jacobian with respect to μ. If this is the case, the number or character of the fixed points may change, sometimes dramatically, and with it the whole split of the space into attractor regions of different character. This process is known as *bifurcation*. In systems which have a fast dynamics F parametrized by a slow-varying (and perhaps externally controlled) parameter μ, the appearance of new fixed points is often interpreted as a process of self-organization.

4.2 Synergetics

The concept of a slow varying parameter has been made part of the above analysis by a number of approaches, most notably the synergetics approach (Haken 1983), but is also known under the name of *slow manifold* and *fast foliation* (Mees 1981). If we consider a dynamical system where the Jacobian of F has a few eigenvalues very close to 0, and a large number of strongly negative eigenvalues, those components of x which fall into the degrees of freedom of the strongly negative eigenvalues will vanish quickly, reducing the essential dynamics of the system to the low-dimensional submanifold of the whole system which corresponds to the "slow" degrees of freedom of the system. In the more colourful language of synergetics, these "slow" degrees of freedom are the "master" modes that "enslave" the fast, quickly decaying modes belonging to the strongly negative eigenvalues.

Synergetics provided an historically early formal and quantitative approach for the treatment of self-organization phenomena by decomposing a possibly large system into hierarchically separate layers of dynamics. It has been successfully applied to a number of physical systems and models, including laser pumping, supraconductivity, the Ginzburg-Landau equations and the pattern formation and the Bénard instabilities in fluids (Haken 1983). However, it only works properly under certain conditions (namely the particular structure of the eigenvalue spectrum) and there are self-organization phenomena it fails to capture fully (Spitzner and Polani 1998).

4.3 Pattern Formation in Spatial Media

To define the dynamical system concept from Secs. 4.1 and 4.2 one requires the concept of for smooth manifolds, i.e. a space with consistent differentiable structures. If one adds the requirement that F obeys given symmetries, i.e. that there is a symmetry group $\mathit{\Gamma}$ such that $\gamma \in \mathit{\Gamma}$ operates on $\mathbb{R}^n$ and (1) obeys

$$\dot{(\gamma\mathbf{x})} = F(\gamma\mathbf{x}, \mu)$$

for all $\gamma \in \mathit{\Gamma}$, then it can be shown this imposes restrictions on the solution space, including the bifurcation structure (Golubitsky and Stewart 2003).

A special, but important case is a spatial symmetry governing the state space of the dynamics. In the most generic case, the state space is not the finite dimensional space $\mathbb{R}^n$, but rather the space $C^m(\mathbb{R}^k, \mathbb{R})$ of sufficiently smooth functions on $\mathbb{R}^k$, and the symmetries are the Euclidean symmetries (translations and orthogonal rotations) on $\mathbb{R}^k$, and (1) actually becomes a partial differential equation[3]. In a discrete approximation one can replace the space $\mathbb{R}^k$ by a *lattice* $L = \{\sum_{i=1}^k l_i \mathbf{v}_i \mid l_i \in \mathbb{Z}\}$ for linear independent $\mathbf{v}_i \in \mathbb{R}$ (Hoyle 2006), and thus reobtain a finite-dimensional version of (1), this time the n components of space not anymore forming an unstructured collection, but rather being organized as a lattice and subject to its symmetries. Here, the system dynamics, together with the symmetries governing the system give rise to particular stable states and attractors where, due to their easily visualisable spatial structure it is easy to detect phenomena of self-organization.

A classical example for such a model is Turing's reaction-diffusion model (Turing 1952). He was among the first to study the dynamics of reaction-diffusion systems as possible model for computation; his particular interest was to study pattern formation in biological scenarios. In a time where there was still a debate which would be the most adequate model for computation, Turing's suggestion of a spatially organized computational medium with an activator and an inhibitor substance with differing diffusion coefficients, provides a wide range of spatial organization dynamics. Depending on the chemical dynamics, there are different instances of reaction-diffusion machines. Figure 2 shows some Turing patterns emerging from having an activator and an inhibitor with concentrations a, b, respectively, computed from the definition of their rates of change

$$\frac{\partial a}{\partial t} = \delta_1 \Delta a + k_1 a^2/b - k_2 a \tag{2}$$

$$\frac{\partial b}{\partial t} = \delta_2 \Delta b + k_3 a^2 - k_4 b \tag{3}$$

in the discrete approximation of a square lattice (Bar-Yam 1997). The starting configuration were randomly initialized concentration fields for a and b. The simulations show the dynamics is constrained to produce structured patterns from virtually arbitrary initial states.

Other reaction-diffusion systems exhibiting spatial self-organization phenomena are, for instance the Beluzov-Zhabotinski reaction which can be implemented in vitro. Reaction-diffusion processes are believed to influence the morphogenesis processes in living beings (Meinhardt 1982) and, as such, of central importance to understand how morphological complexity can be obtained by "unpacking" the relatively compact genotype.

5 Information-Theoretic Approaches to Self-Organization

The models from Secs. 4 have in common that they require the dynamical system to "live" on a differentiable manifold to describe self-organization. In addition, the

[3] Here we ignore technical details necessary to properly define the dynamics.

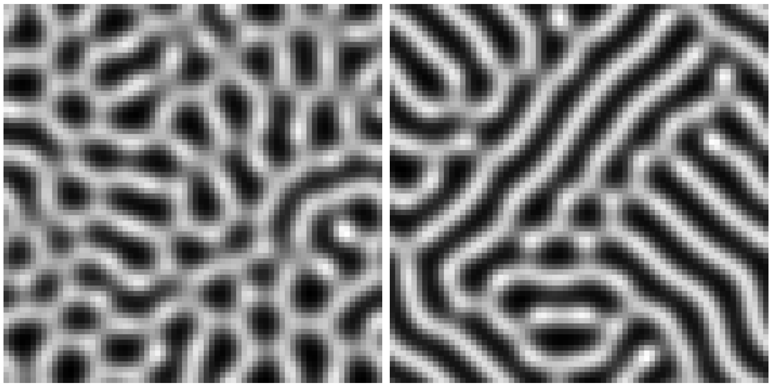

Fig. 2. Turing patterns of the concentration a of the activator, as emerging from (2) for different parameters k. See (Bar-Yam 1997) for discussion how to obtain different patterns.

synergetics model of self-organization requires a particular grouping of the eigenvalues of the Jacobian and the pattern formation models require the presence of a spatially structured state space. These are relatively specific requirements. In a unified treatment of self-organization, it would be very limiting to exclude self-organization in discrete, not differentiable, worlds. Similarly, it would be inappropriate to assume that systems must have a Euclidian spatial organization similar to reaction-diffusion systems. It is easy to envisage scenarios where a system may possess other topological/structural properties, such as social or food web networks.

In addition, the example systems from Sec. 4 were implicitly assumed to be deterministic, which is usually far too strong an assumption. Neither this assumption nor the assumption from the last paragraph needs to hold: one can imagine self-organization in an agent system (such as relevant for engineering problems) which is neither deterministic nor organized on a regular spatial structure, and certainly nothing can be assumed in terms of distributions of eigenvalues close to fixed points (if these at all exist).

Can anything at all be analysed in such structurally impoverished scenarios? Indeed, it turns out that still a lot can be said with much less structure, and the toolbox of choice is information theory. In the following, we will outline two approaches to model self-organization using information theory.

Information theory operates on probability distributions. These require only minimal structure (a probability measure) on the space of interest, and make no assumption about differentiability or spatial structure. Information theory has crystallized as a promising common language for the study of general systems, to tackle issues of complex phenomena exhibiting a wide variety of seemingly incompatible properties.

5.1 Notation

Due to space limitations, the formalization of the exposition is restricted to a minimum. Consider a random variable X assuming values $x \in \mathcal{X}$, $\mathcal{X}$ the set of possible values for X. For simplicity, assume that $\mathcal{X}$ is finite. Define the *entropy* of X by

$$H(X) := -\sum_{x \in \mathcal{X}} p(x) \log p(x) \ ,$$

the *conditional entropy* of Y as

$$H(Y|X) := \sum_{x \in \mathcal{X}} p(x) H(Y|X = x)$$

with

$$H(Y|X = x) := -\sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x)$$

for $x \in \mathcal{X}$. The *joint entropy* of X and Y is the entropy $H(X, Y)$ of the random variable (X, Y). The *mutual information* of random variables X and Y is $I(X; Y) := H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y)$. A generalization is the *intrinsic information*: for random variables $X_1, \dots, X_k$, the intrinsic or multi-information is

$$I(X_1; \dots; X_k) := \left[\sum_{i=1}^{k} H(X_i)\right] - H(X_1, \dots, X_k) \ .$$

This notion is also known e.g. as *integration* in (Tononi et al. 1994). Similar to the mutual information, it is a measure for the degree of dependence between the different X_i.

5.2 Self-Organization as Increasing Statistical Complexity

One influential approach to study complex systems and the notions of self-organization and emergence is based on the ϵ-machine formalism which provides a model to describe complex temporal processes (Crutchfield and Young 1989). Using this formalism, Shalizi (2001) develops a quantifiable notion of self-organization. In the following, we briefly describe the ϵ-machine formalism and the ensuing model for self-organization.

Consider a stochastic process (with, say, infinite past and future):

$$\mathbf{X} = \dots X^{(t-3)}, X^{(t-2)}, X^{(t-1)}, X^{(t)}, X^{(t+1)}, X^{(t+2)}, X^{(t+3)}, \dots \ .$$

Denote the (ordered) sequence of variables up to $X^{(t)}$ by $\overleftarrow{X}$ (*past*) and the sequence of variables from $X^{(t+1)}$ upwards by $\overrightarrow{X}$ (*future*). Consider the equivalence relation that identifies all pasts $\overleftarrow{x}$ for which the probability distribution $P(\overrightarrow{X}|\overleftarrow{x})$ of the possible futures is exactly the same. This equivalence relation partitions the pasts into disjoint sets, which, for the sake of simplicity, we name $\tilde{x}$. Any past $\overleftarrow{x}$ is member of exactly one equivalence class $\tilde{x}$.

To construct an ϵ-machine from a given process $\mathbf{X}$, define an automaton such that its states are identified one-to-one by the equivalence classes $\tilde{x}$ arising from above procedure. When a transition from t to $t+1$ is made, it means replacing a past $\dots X^{(t-3)}, X^{(t-2)}, X^{(t-1)}, X^{(t)}$ by a past $\dots X^{(t-2)}, X^{(t-1)}, X^{(t)}, X^{(t+1)}$, and thus is acts as a transition from an equivalence class $\tilde{x}$ to an equivalence class $\tilde{x}'$, corresponding to the new past. Together with labeling the transition by the realization $x^{(t+1)}$ of $X^{(t+1)}$, this defines the automaton.

The ϵ-machine, when it exists, acts as the unique minimal maximally predictive model of the original process (Shalizi and Crutchfield 2002), including highly non-Markovian processes which may contain a significant amount of memory[4]. It allows to define the concept of *statistical complexity* as the entropy $H(\tilde{X}) = -\sum_{\tilde{x}} p(\tilde{x}) \log p(\tilde{x})$ of the states of the ϵ-machine. This is a measure of the memory required to perform the process $\mathbf{X}$.

Note that the statistical complexity is, in general, different from another important quantity, known, among other, as *excess entropy* and which is given by $\eta(\mathbf{X}) := I(\overleftarrow{X}; \overrightarrow{X})$ (see e.g. Grassberger 1986). One always has $\eta(\mathbf{X}) \leq H(\tilde{X})$ (Shalizi 2001). The interpretational distinction between statistical complexity and excess entropy is subtle. Of the selection of interpretations available, the author prefers the view inspired by the "information bottleneck" perspective (Tishby et al. 1999; Shalizi and Crutchfield 2002): The excess entropy is the actual information contained in the complete past (for a given time) about the complete future *as it could be reconstructed if the complete past were available as a whole*. As opposed to that, to obtain the statistical complexity one has to force this information through the "bottleneck" given by the ϵ-machine state at the *present* time slice which has to provide a sufficient statistics about the past concerning the future constrained to the current time. Because of the constraint of this bottleneck to the present time slice it is, in general, less parsimonious in description than the "idealized" excess entropy that, in principle, assumes availability of the whole process (past and future) to produce its prediction. In the bottleneck picture, we have the sequence

$$\tilde{X} \longleftarrow \overleftarrow{X} \longleftrightarrow \overrightarrow{X}$$

where the left arrow indicates the projection of $\overleftarrow{X}$ to the ϵ-machine state and the arrows between the past and future variables indicate their informational relation. The process of "squeezing" their mutual information into the bottleneck variable $\tilde{X}$ produces in general a variable with a larger entropy than the actual mutual information content between $\overleftarrow{X}$ and $\overrightarrow{X}$ (this is a general property of the information bottleneck, of which the relation between statistical complexity and excess entropy is just a special case).

[4] Note that, in general, the construction of an ϵ-machine from the visible process variables $\mathbf{X}$ is not necessarily possible, and the reader should be aware that the Shalizi/Crutchfield model is required to fulfil suitable properties for the reconstruction to work. I am indebted to Nihat Ay and Wolfgang Löhr for pointing this out to me.

In computing the ϵ-machine, the picture of an infinite time series is idealized. In the empirical case one will consider finite time series, giving rise to a “localized” ϵ-machine, operating on a certain window size. Here, one will expect to encounter a slow drift superimposed on the fast dynamics of the process which will slowly change the localized ϵ-machine with the passing of time. Shalizi (2001) calls a system self-organizing if this statistical complexity grows with time. We will call this flavour of self-organization *SC-self-organization* (“SC” for statistical complexity).

This approach basically considers organization to be essentially the same as complexity. In this model, self-organization is an intrinsic property of the system and unambiguously measurable. In particular, this approach makes the relation to emergence unambiguously clear, as Shalizi gives a definition of emergence based on the ϵ-machine notion in the same work. The essence is that in the ϵ-machine perspective emergence is present if there is a coarse-grained description of the system that is more predictively efficient than the original description of the process, i.e. if it has a higher ratio $\eta(\mathbf{X})/H(\tilde{X})$ of excess entropy vs. statistical complexity, a better ratio between the total amount of process information that ideally needs to be processed and the process memory that is actually necessary to achieve that.

This approach to emergence is descriptive, as it characterizes a property of the particular description (i.e. perspective or “coordinate system”) through which one looks into a system. As opposed to that, self-organization in the ϵ-machine model is a purely intrinsic property of the system. Through this split into description and intrinsic properties, Shalizi (2001) argues that while emergence may allow to simplify descriptions of a system, there may be cases of self-organization which humans do not recognize as such because there is no appropriate simplified (emergent) coordinate system through which the self-organization would become apparent. It is only visible through the ϵ-machine construction. This provides a transparent picture how self-organization and emergence turn out to be two mathematically distinct concepts that represent different aspects of a system. While this is a motif that one finds repeatedly emphasized in the scientific discourse, it is rarely formulated in an as compelling and crisp language.

One interesting aspect of the ϵ-machine view is how it reflects the bifurcation concept from Sec. 4.1. Consider as process an iterator map for $t \to \infty$. As long as there is only a single fixed point attractor, the ϵ-machine will (asymptotically) have only one state. As a bifurcation into two fixed point attractors occurs, these two attractors will be reflected by the ϵ-machine. With the bifurcation behaviour becoming more intricate (as would happen, say, in the logistic map example with an adiabatically slowly growing bifurcation parameter), the ϵ-machine also grows in complexity. In this way, the ϵ-machine can grow significantly in size and with it the statistical complexity.

5.3 Observer-Induced Self-Organization

Conceptually, pattern formation which we gave in Sec. 4.3 as a prominent example of self-organization, does not fit smoothly into the picture of growing statistical complexity. One reason for that is that statistical complexity by its very foundations is a concept that operates on an process that has no a priori structure on the $X^{(t)}$, except for the orderedness (and, implied, directedness) of time.

Quite different from that, spatial pattern formation inextricably requires a spatial structure on its state space $\mathcal{X}$. The patterns that develop during the experiments form in space, and space has strong structural constraints. For this purpose, in (Shalizi 2001), a spatial ϵ-machine is developed to deal specifically with this problem. Thus, the spatial structure is brought in explicitly as a part of the model.

An alternative approach to model self-organization using information theory is suggested in (Polani 2003). This approach no longer considers an unstructured dynamical system on its own, but adds the concept of an *observer* which acts as a particular “coordinate system” through which the given system is represented at a given time step. For this model of self-organization, an observer or coordinate system needs to be specified *in addition* to the dynamics of the system. The suspicion that observers may be of importance to characterize complex systems has been voiced quite a few times in the past (Crutchfield 1994; Harvey 2000; Baas and Emmeche 1997; Rasmussen et al. 2001). In formalizing this idea here, we follow the particular flavour from (Polani 2003).

6 Organization via Observers

A *(perfect) observer* of a (random) variable X is a collection $X_1, X_2, \dots, X_k$ of random variables allowing full reconstruction of X, i.e. for which $H(X|X_1, X_2, \dots, X_k)$ vanishes. We define the the *organization information* with respect to the observer as the multi-information $I(X_1; \dots; X_k)$. We call a system *self-organizing* (with respect to the given observer) if the organization information increase with respect to the observer variables is positive as the system dynamics progresses with time. $I(X_1; \dots; X_k)$ quantifies to which extent the observer variables $X_1, X_2, \dots, X_k$ depend on each other. We call this flavour of self-organization *O-self-organization* (“O” for observer-based).

The set of observer variables can often be specified in a natural way. For instance, systems that are composed by many, often identical, individual subsystems, have a canonical observer, defined via the partition of the system into subsystems. For instance, the observer variables could denote the states of agents that collectively make up a system. An increase in the multi-information of the system with respect to the agent states then indicates an increasing degree of coordination between the agents: this is consistent with our intuitive understanding of self-organization. Reaction-diffusion systems are also naturally described in this framework. Each point in space becomes an observer variable; in the attractor states with spot-and-wave patterns, these observer variables are intrinsically correlated.

Note, however, that for the multi-information not to vanish, it is still necessary that the whole system has some degree of freedom and that there is not just a single fixed pattern the system converges to. This makes sense, since otherwise one would just be talking about a single-attractor, and thus trivial, system.

Using the Self-Organizing Map as model system, and the individual neuron weights as observer variables, (Polani 2003) discusses in detail the advantage of multi-information as a measure for self-organization as compared to other information-theoretic candidates for such a measure (except for the comparison with SC-self-organization, which is discussed in the present chapter for the first time). Many of the arguments carry over to other typical scenarios.

Note that compared to SC-self-organization (Sec. 5.2), O-self-organization is different in several respects. SC-self-organization does not require observers, and arises from the intrinsic dynamics of the system. This is orthogonal to the view of O-self-organization. In principle, the need for fewer assumptions by SC-self-organization is a conceptual advantage. On the other hand, to model the appearance of regular patterns (e.g. of a reaction-diffusion system) as a self-organization process one must anyway specify the extra spatial structure in which the patterns appear. In O-self-organization, this can be directly made a part of the specification. Thus, O-self-organization would be a natural candidate for these types of scenarios.

6.1 Observer Dependence

For the observer-based measure, a central question is how the measure changes as one moves from one observer to another, i.e. what happens to the measure on change of the "coordinate system". It turns out that it is possible to formulate an interesting relation between fine-grained observers and a coarse-graining of the very same observers. We will discuss this relation in the following.

Let $X_i, i = 1 \dots k$ be a collection of jointly distributed random variables; this collection forms the *fine-grained observer*. Obtain the *coarse-grained observer* by grouping the X_i according to

$$\underbrace{X_1, \dots, X_{k_1}}_{\tilde{X}_1}, \underbrace{X_{k_1+1}, \dots, X_{k_2}}_{\tilde{X}_2}, \dots, \underbrace{X_{k_{\tilde{k}-1}+1}, \dots, X_k}_{\tilde{X}_{\tilde{k}}} \,, \tag{4}$$

i.e. each of the coarse-grained variables $\tilde{X}_j, j = 1 \dots \tilde{k}$ is obtained by grouping several of the fine-grained variables X_i together.

Then the multi-information of the fine-grained observer can be expressed as[5]

$$I(X_1; X_2; \dots; X_k) = I(\tilde{X}_1; \tilde{X}_2; \dots; \tilde{X}_{\tilde{k}}) + \sum_{j=1}^{\tilde{k}} I(X_{k_{j-1}+1}; \dots; X_{k_j}) \,, \tag{5}$$

(where we adopt the convention $k_0 := 0$ and $k_{\tilde{k}} := k$). Relation (5) states that the multi-information as measure of self-organization in the fine-grained case can be expressed as th multi-information for the set of coarse-grained variables, corrected by the *intrinsic* multi-information of all these coarse-grained variables, or to put it snappily, "the fine-grained system is more than the sum of its coarse-grained version". The proof is sketched in Appendix A[6].

[5] This is a generalization of Eq. (3) from (Tononi et al. 1994) for the bipartite case to the multipartite case.

[6] This property is related to a property that can be proven for graphical models, see e.g. Proposition 2.1 in (Slonim et al. 2001).

Equation (5) states how the intrinsic information of a system changes under a "change of coordinates" by regrouping of the random variables that represent the system. This "bookkeeping" of multi-information while changing the basis for the system description in general only applies to regrouping of variables, but not to recoding. Under recoding of variables (i.e. re-representing the variables X_i by random variables $Y_i = f_i(X_1, \dots, X_i, \dots, X_k)$ where f_i is some deterministic function), there is not a canonical way of transforming the multi-information in a simple and transparent manner.

To see that, note that recoding may entirely remove dependencies between the recoded X_i (e.g. in Independent Component Analysis Comon 1991). In fact, further requirements can be added to the component independence; indeed, this has been proposed as a way of discovering degrees of freedom representing *emergent levels of description* (Polani 2004, 2006). In this sense, O-self-organization is distinct from and effectively conjugate to the "emergent descriptions" concept. This dichotomy mirrors the analogous dichotomy exhibited by the formalization of self-organization and emergence using the ϵ-machine formalism (Sec. 5.2).

7 Discussion

7.1 SC- and O-Self-Organization

We have emphasized that self-organization is a phenomenon often discussed in conjunction with complex systems. While there is a manifold selection of processes that are associated with this phenomenon, most notions used to characterize self-organization are either too vague to be useful, or too specific to be transferable from a system to another. The information-theoretic notions of (statistical complexity) SC-self-organization (Shalizi 2001; Shalizi et al. 2004) and that of (observer-based) O-self-organization (Polani 2003) strive to fill this gap. While similar to each other in the general philosophy, they are essentially "orthogonal" as quantities.

SC-self-organization takes in a single time series and measures the growth in statistical complexity during time. O-self-organization requires an observer, i.e. a set of variables through which the system state is observed. Such a set of variables is often naturally available, for instance, in multiagent systems. Similar to SC-self-organization, O-self-organization seems to capture essential aspects of self-organization — for instance, the freezing of seemingly unrelated degrees of freedom (the observer variables) into highly coordinated global behaviour.

While SC-self-organization concentrates on the complexity of the temporal dynamics, O-self-organization concentrates on the compositional aspects of the system (this compositionality can, but need not be spatial). This distinction is also what indicates the use of each of the measures. If one focuses on the temporal dynamics, SC-self-organization may be more relevant, if on the spatial or compositional dynamics, O-self-organization may be the measure of choice. As the precise mathematical conceptualizations of self-organization are relatively recent, it will take some time until enough experience is accumulated to make an informed decision which measure is appropriate to use in a given constellation, or whether still other, more suitable measures will need to be discovered.

A word of caution at this point: the calculation of multi-information is difficult if the number of variables is large (Slonim et al. 2005). Particularly in unorganized and random states a Monte-Carlo estimate of the entropies and multi-information is likely to be undersampled and to overestimate the organization of the system in that state. Research is underway to develop methods that are able to estimate the organization of unstructured states properly (and to distinguish it from the organized states) in general settings.

7.2 Introducing Observers

For O-self-organization, we have assumed the existence of natural observers. What if none exist? Which ones to introduce and to use? The multi-information term consists of the entropies of the individual variables as well as the entropy of the joint variables. The latter depends only on the total system, not the observer. It is therefore the entropies of the *individual* variables that will change if we choose different observers. In general, it is quite possible to choose them in such a way as to make them independent — while this choice of observer is interesting (it essentially corresponds to an Independent Component Analysis, Sec. 6.1), it makes the system maximally un-self-organized. This clearly shows that O-self-organization is not intrinsic. O-self-organization is “in the eye of the beholder” (Harvey 2000), but in a formally precise way.

Now, for O-self-organization to be present at all, the whole system must have some degrees of uncertainty, otherwise the individual variable entropies will also collapse and the multi-information will vanish. This is a property of the whole system. Thus, one could consider a natural observer one that *maximizes* the multi-information (as opposed to minimizing it), and thus making the system as self-organized as possible. If this is the case, O-self-organization could be viewed as the opposite to independent component decomposition.

But there is yet another way of constructing a natural observer: if one considers units (agents) that operate in the system and which possess sensors and actuators, the former which attain information about the system, and the latter which act upon and modify the system, then the perception-action loop of these agents forms a structured information channel. It can be shown (Klyubin et al. 2004) that maximizing the information flow through this channel allows the units to extract features from the system that are pertinent to the structure of the system.

This view is particularly interesting since it does not look at a system with a preordained temporal dynamics, but rather the units (agents) have the option to choose their own actions. Nevertheless, once they perform the information flow maximization, they attain perceptual properties specially appropriate for the system at hand. The thus attained filters or feature detectors could act as another form of natural observer variables for the given system. Similarly, principles of informational organization can lead to a joint coordination of a sensorimotor device (Klyubin et al. 2005; Prokopenko et al. 2006) and direct systems to an equipment with embodiment-appropriate pattern detector loops.

8 Conclusion

The present chapter discussed the general problem of defining self-organization and presented two operational approaches, both based on information-theoretic principles. One approach, based on the ϵ-machine formalism, defines self-organization as an intrinsic property of a system, as a growth of the memory required to process a time series of random variable. The other approach defines self-organization via an observer, in typical cases realized as a family of variables of more-or-less similar type; a growing coordination between these variables with time is then identified as self-organization. Depending on one's aims, one will choose one or the other model to identify self-organization in a system. In particular, SC-self-organization will be the notion of choice if one is interested in characterizing the increase in complexity of the temporal dynamics, while O-self-organization emphasizes the self-organization process in a system composed of many individual subsystems.

The advantage of using information-theoretic notions for quantifying self-organization is that they provide a precise language for identifying the conditions of self-organization and the underlying assumptions, as opposed to vague or uncomputable qualitative characterizations. The quantitative character of information measures also allows one to actively search for "more self-organized" systems in a given context, rather than just state whether a system possesses or does not possess this property (as e.g. an algebraic characterization would do). In addition, the information-theoretic language forces one to specify exactly the assumptions and requirements underlying the notions one is using.

In short, information theory proves to be a powerful language to express self-organization and other central concepts relevant to complex systems. Even if one ultimately should prefer a different route to characterize self-organization in a complex system, it is probably a good first bet to strive towards a formulation that profits from the clarity and transparence of the information-theoretic language.

A Proof of Relation between Fine and Coarse-Grained Multi-Information

Proof. First, note that a different way to write the composite random variables $\tilde{X}_j$ is $\tilde{X}_j = (X_{k_{j-1}+1}, \dots, X_{k_j})$ for $j = 1 \dots \tilde{k}$, giving

$$H(\tilde{X}_j) = H(X_{k_{j-1}+1}, \dots, X_{k_j}) \ . \tag{6}$$

Similarly, the joint random variable $(\tilde{X}_1, \dots, \tilde{X}_{\tilde{k}})$ consisting of the composite random variables $\tilde{X}_j$ can be seen as a regrouping of the elementary random variables

$X_1, \ldots, X_k$.Therefore the joint random variable constructed from the $\tilde{X}_j$ and that constructed from the X_i have both the same entropy:

$$H(\tilde{X}_1, \ldots, \tilde{X}_{\tilde{k}}) = H(X_1, \ldots, X_k)\ . \tag{7}$$

For consistency of notation, write $k_0 = 0$ and $k_{\tilde{k}} = k$. One then obtains

$$\begin{aligned}
&I(\tilde{X}_1; \tilde{X}_2; \ldots; \tilde{X}_{\tilde{k}}) + \sum_{j=1}^{\tilde{k}} I(X_{k_{j-1}+1}; \ldots; X_{k_j}) \\
&= \sum_{j=1}^{\tilde{k}} H(\tilde{X}_j) - H(\tilde{X}_1, \ldots, \tilde{X}_{\tilde{k}}) + \sum_{j=1}^{\tilde{k}} \Bigg(\sum_{j'=k_{j-1}+1}^{k_j} H(X_{j'}) - H(X_{k_{j-1}+1}, \ldots, X_{k_j}) \Bigg) \\
&= \underbrace{\sum_{j=1}^{\tilde{k}} \sum_{j'=k_{j-1}+1}^{k_j} H(X_{j'})}_{=\sum_{i=1}^{k} H(X_i)} + \sum_{j=1}^{\tilde{k}} \underbrace{\Big(H(\tilde{X}_j) - H(X_{k_{j-1}+1}, \ldots, X_{k_j}) \Big)}_{=0} - \underbrace{H(\tilde{X}_1, \ldots, \tilde{X}_{\tilde{k}})}_{=H(X_1,\ldots,X_k)}
\end{aligned}$$

where the first term results from a regrouping of summands, the second term results from Eq. (6) and the third from rewriting the whole set of random variables from the coarse-grained to the fine-grained notation, thus giving

$$\begin{aligned}
&= \sum_{i=1}^{k} H(X_i) - H(X_1, \ldots, X_k) \\
&= I(X_1; \ldots; X_k)
\end{aligned}$$

which proves the equation. ■

References

Adami, C. (1998). *Introduction to Artificial Life*. Springer.

Ashby, W. R. (1947). Principles of the self-organizing dynamic system. *J. Gen. Psychol.*, 37:125–128.

Ay, N. and Krakauer, D. C. (2006). Information geometric theories for robust biological networks. *J. Theor. Biology*. In Press.

Ay, N. and Polani, D. (2006). Information flows in causal networks. Proc. NIPS Workshop on Causality and Feature Selection.

Ay, N. and Wennekers, T. (2003). Dynamical properties of strongly interacting markov chains. *Neural Networks*, 16(10):1483–1497.

Baas, N. A. and Emmeche, C. (1997). On emergence and explanation. *Intellectica*, 2(25):67–83.

Bar-Yam, Y. (1997). *Dynamics of Complex Systems*. Studies in Nonlinearity. Westview Press, Boulder, Colorado.

Bennett, C. H. and Landauer, R. (1985). The fundamental limits of computation. *Scientific American*, pages 48–56.

Bertschinger, N., Olbrich, E., Ay, N., and Jost, J. (2006). Autonomy: an information theoretic perspective. In *Proc. Workshop on Artificial Autonomy at Alife X, Bloomington, Indiana*, pages 7–12.

Comon, P. (1991). Independent component analysis. In *Proc. Intl. Signal Processing Workshop on Higher-order Statistics, Chamrousse, France*, pages 111–120.

Crutchfield, J. P. (1994). The calculi of emergence: Computation, dynamics, and induction. *Physica D*, pages 11–54.

Crutchfield, J. P. and Young, K. (1989). Inferring statistical complexity. *Phys. Rev. Lett.*, 63:105–108.

Emmeche, C., Køppe, S., and Stjernfelt, F. (2000). Levels, emergence, and three versions of downward causation. In Andersen, P. B., Emmeche, C., Finnemann, N. O., and Christiansen, P. V., editors, *Downward Causation. Minds, Bodies and Matter*, pages 13–34. Århus: Aarhus University Press.

Golubitsky, M. and Stewart, I. (2003). *The Symmetry Perspective*. Birkhäuser.

Grassberger, P. (1986). Toward a quantitative theory of self-generated complexity. *Int. J. Theor. Phys.*, 25:907–938.

Haken, H. (1983). *Advanced synergetics*. Springer-Verlag, Berlin.

Harvey, I. (2000). The 3 es of artificial life: Emergence, embodiment and evolution. Invited talk at Artificial Life VII, 1.-6. August, Portland.

Helbing, D., Buzna, L., Johansson, A., and Werner, T. (2005). Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions. *Transportation Science*, 39(1):1–24.

Hoyle, R. (2006). *Pattern Formation*. Cambridge University Press.

Jetschke, G. (1989). *Mathematik der Selbstorganisation*. Vieweg, Braunschweig.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2004). Organization of the information flow in the perception-action loop of evolved agents. In *Proceedings of 2004 NASA/DoD Conference on Evolvable Hardware*, pages 177–180. IEEE Computer Society.

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2005). Empowerment: A universal agent-centric measure of control. In *Proc. IEEE Congress on Evolutionary Computation, 2-5 September 2005, Edinburgh, Scotland (CEC 2005)*, pages 128–135. IEEE.

Landauer, R. (1961). Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5:183–191.

Mees, A. I. (1981). *Dynamics of feedback systems*. John Wiley & sons, Ltd.

Meinhardt, H. (1972). A theory of biological pattern formation. *Kybernetik*, 12:30–39.

Meinhardt, H. (1982). *Models of Biological Pattern Formation*. Academic Press.

Pask, G. (1960). The natural history of networks. In (Yovits and Cameron 1960).

Polani, D. (2003). Measuring self-organization via observers. In Banzhaf, W., Christaller, T., Ziegler, J., Dittrich, P., Kim, J. T., Lange, H., Martinetz, T., and

Schweitzer, F., editors, *Advances in Artificial Life (Proc. 7th European Conference on Artificial Life, Dortmund, September 14-17, 2003)*.

Polani, D. (2004). Defining emergent descriptions by information preservation. *InterJournal Complex Systems,* 1102.

Polani, D. (2006). Emergence, intrinsic structure of information, and agenthood. *InterJournal Complex Systems*, 1973.

Prigogine, I. and Nicolis, G. (1977). *Self-Organization in Non-Equilibrium Systems: From Dissipative Structures to Order Through Fluctuations*. J. Wiley & Sons, New York.

Prokopenko, M., Gerasimov, V., and Tanev, I. (2006). Evolving spatiotemporal coordination in a modular robotic system. In Nolfi, S., Baldassarre, G., Calabretta, R., Hallam, J. C. T., Marocco, D., Meyer, J.-A., Miglino, O., and Parisi, D., editors, *From Animals to Animats 9: 9th International Conference on the Simulation of Adaptive Behavior (SAB 2006), Rome, Italy*, volume 4095 of *Lecture Notes in Computer Science*, pages 558–569. Springer.

Rasmussen, S., Baas, N., Mayer, B., Nilsson, M., and Olesen, M. W. (2001). Ansatz for dynamical hierarchies. *Artificial Life*, 7:329–353.

Reichl, L. (1980). *A Modern Course in Statistical Physics*. University of Texas Press, Austin.

Shalizi, C. R. (2001). *Causal Architecture, Complexity and Self-Organization in Time Series and Cellular Automata*. PhD thesis, University of Wisconsin-Madison.

Shalizi, C. R. and Crutchfield, J. P. (2002). Information bottlenecks, causal states, and statistical relevance bases: How to represent relevant information in memoryless transduction. *Advances in Complex Systems*, 5(1):91–95.

Shalizi, C. R., Shalizi, K. L., and Haslinger, R. (2004). Quantifying self-organization with optimal predictors. *Physical Review Letters*, 93(11):118701.

Slonim, N., Friedman, N., , and Tishby, T. (2001). Agglomerative multivariate information bottleneck. In *Neural Information Processing Systems (NIPS 01)*, pages 929–936.

Slonim, N., s. Atwal, G., Tkačik, G., and Bialek, W. (2005). Estimating mutual information and multi-information in large networks. arXiv:cs.IT/0502017.

Spitzner, A. and Polani, D. (1998). Order parameters for self-organizing maps. In Niklasson, L., Bodén, M., and Ziemke, T., editors, *Proc. of the 8th Int. Conf. on Artificial Neural Networks (ICANN 98), Skövde, Sweden*, volume 2, pages 517–522. Springer.

Tishby, N., Pereira, F. C., and Bialek, W. (1999). The information bottleneck method. In *Proc. 37th Annual Allerton Conference on Communication, Control and Computing, Illinois*.

Tononi, G., Sporns, O., and Edelman, G. M. (1994). A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proc. Natl. Acad. Sci. USA*, 91:5033–5037.

Turing, A. M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. B*, 327:37–72.

Walter, W. G. (1951). A machine that learns. *Scientific American*, pages 60–63.

Yovits, M. C. and Cameron, S., editors (1960). *Self-Organizing Systems – Proceedings of an Interdisciplinary Conference, 5-6 May 1959*, Computer Science and Technology and their Application. Pergamon Press.