Chapter 1

Emergence, Intrinsic Structure of Information, and Agenthood

Daniel Polani
Adaptive Systems Research Group
School of Computer Science, University of Hertfordshire, UK
d.polani@herts.ac.uk

Emergence is a central organizing concept for the understanding of complex systems. Under the manifold mathematical notions that have been introduced to characterize emergence, the information-theoretic are of particular interest since they provide a quantitative and transparent approach and generalize beyond the immediate scope at hand.

We discuss approaches to characterize emergence using information theory via the intrinsic temporal or compositional structure of the information dynamics of a system. This approach is devoid of any external constraints and purely a property of the information dynamics itself. We then briefly discuss how emergence naturally connects to the concept of agenthood which has been recently defined using information flows.

1.1 Introduction

The concept of *emergence* is of central importance to understand complex systems. Although there seems to be quite an intuitive agreement in the community which phenomena in complex systems are to be viewed as "emergent", similarly to the concept of *complexity*, it seems difficult to construct a universally accepted precise mathematical notion of emergence. Unsurprisingly one is thus faced with a broad spectrum of different approaches to define emergence.

The present paper will briefly discuss a number of notions of emergence and then focus on the information-theoretic variants. Due to its universality, information theory spawned a rich body of concepts based on its language. It provides power of quantification, characterization and of prediction. The paper will discuss how existing information-theoretic notions of emergence can be connected to issues of intrinsic structure of information and the concept of "agenthood" and thus provide new deep insights into the ramifications, and perhaps the reason why emergence plays such an important role.

1.2 Some Notions of Emergence

Of the broad spectrum of notions for emergence we will give an overview over a small selection representing a few particularly representative approaches, before concentrating on the information-theoretic notions which form the backbone of the philosophy of the present paper.

A category-theoretic notion of emergence has been brought forward in [11]. While having the advantage of mathematical purity, category theory does not lend itself easily for the practical use in concrete systems. One of the difficulties is that the issue of identifying the emergent levels of description is exogenous to the formalism. These have to be formulated externally to be verified by the formalism. As is, the approach provides no (even implicitly) constructive way of finding the emergent levels of description.

The difficulty of identifying the right levels of description for emergence in a system has brought up the suspicion that emergence would have to be considered only "in the eye of the beholder" [6]. In view of the fact that human observers typically agree on the presence of emergence in a system, it is often felt that it would rather be desirable to have a notion of emergence that does not depend on the observer, but is a property that would arise naturally from the dynamics of the system.

In an attempt to derive emergent properties of a system, a pioneering effort to describe organizing principles in complex systems is the approach taken by *synergetics* [4]. The model attempts to decompose nonlinear dynamic systems in a natural fashion. In the vicinity of fixed points, dynamical systems decompose naturally into stable, central and unstable manifolds. Basically, this decomposes a system into fast and slow moving degrees of freedom (*fast foliations* and *slow manifolds*).

Since the lifetime of the slow degrees of freedom exceeds that of the fast ones, Haken termed the former *master modes* as compared to the latter which he termed *slave modes*. The main tenet of synergetics is that these master modes dominate the dynamics of the system. In the language of synergetics, the master modes correspond to emergent degrees of freedom. An information-theoretic approach towards the concept of synergetics is presented in [5].

The synergetics view necessarily couples the concept of emergence to the existence of significantly different timescales. In addition, the applicability of above decomposition is limited to the neighbourhood of a fixed point. Under certain conditions, however, it is possible to achieve a canonical decomposition of chaotic dynamical systems even without separate time scales into weakly coupled or decoupled subsystems [13]. In addition to above, a significant number of other approaches exist, of which we will briefly mention a few in §1.3.5 in relation to the information-theoretic approaches to be discussed in §1.3.3 and §1.3.4.

1.3 Information-Theoretic Concepts of Emergence

Among the possible formalizations of emergence, the information-theoretic ones are particularly attractive due to the universality of information theory and the power of expression, description and prediction it provides, as well as the potential to provide paths for explicitly constructing the structures of relevance (see e.g. §1.4.1).

1.3.1 Notation

We introduce some notation. For random variables use capital letters such as X, Y, Z, for the values they assume use letters such as x, y, z, and for the sets they take values in use letters such as $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$. For simplicity of notation, we will assume that such a set $\mathcal{X}$ is finite. A random variable X is determined by the probabilities $\Pr(X = x)$ assumed for all $x \in \mathcal{X}$. Similarly, joint variables (X, Y) are determined via $\Pr(X = x, Y = y)$, and conditional variables via $\Pr(Y = y|X = x)$. If there is no danger of confusion, we will prefer writing the probabilities in the shorthand form of $p(x)$, $p(x, y)$ and $p(y|x)$ instead of the more cumbersome explicit forms above.

Define the *entropy* of a random variable X by

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

and the *conditional entropy* of Y given X as

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

where $H(Y|X = x) := -\sum_{y\in\mathcal{Y}} p(y|x) \log p(y|x)$ for $x \in \mathcal{X}$. The *joint entropy* $H(X, Y)$ is the entropy of the random variable (X, Y) with jointly distributed X and Y. The *mutual information* of random variables X and Y is defined as

$$I(X; Y) := H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y) \ .$$

1.3.2 Projections

In analogy to (regular) maps between sets, we define a probabilistic map $\mathcal{X} \to \mathcal{Y}$ via a conditional $p(y|x)$. If the probabilistic map is deterministic, we call it a *projection*. A given probability distribution $p(x)$ on X induces a probability distribution $p(y)$ on Y via the probabilistic map $X \to Y$ in the natural way.

1.3.3 Emergence as Improved Predictive Efficiency

Based on the epsilon-machine concept, a notion of emergence in time series has been developed in [1, 12]: a process emerges from another one if it has a greater *predictive efficiency* than the second. This means that, the ratio between prediction information (excess entropy) and the complexity of the predicting epsilon-machine is better in the emerging process than the original process. This gives a natural and precise meaning to the perspective that emergence should represent a simpler coarse-grained view of a more intricate fine-grained system dynamics. In the following, we will review the technical aspects of this idea in more detail.

For this, we generally follow the line [12]. Consider a random variable X together with some projection $\mathcal{X} \to \hat{\mathcal{X}}$. Then define the *statistical complexity* of the induced variable $\hat{X}$ as $C_\mu(\hat{X}) := H(\hat{X})$.

Let random variables X, Y be given, where we wish to predict Y from X. Define an equivalence relation $\sim_\epsilon$ on $\mathcal{X}$ of equivalent predictiveness with respect to Y via

$$\forall x, x' \in \mathcal{X} : x \sim_\epsilon x' \quad \text{iff} \quad \forall y \in \mathcal{Y} : p(y|x) = p(y|x') \; . \tag{1.1}$$

The equivalence relation $\sim_\epsilon$ induces a partition $\tilde{\mathcal{X}}$ of $\mathcal{X}$ into equivalence classes[1]. Each $x \in \mathcal{X}$ is naturally member into one of the classes $\tilde{x} \in \tilde{\mathcal{X}}$, and thus there is a natural projection of $\mathcal{X}$ onto $\tilde{\mathcal{X}}$. This induces a probability distribution on $\tilde{\mathcal{X}}$ and makes $\tilde{X}$ a random variable which is is called a *causal state*.

Consider now an infinite sequence $\dots S_{-1}, S_0, S_1, \dots$ of random variables. Furthermore, introduce the notation $S_{[t,t']}$ for the subsequence $S_t, S_{t+1}, \dots, S_{t'-1}, S_{t'}$, where for $t = -\infty$ one has a left-infinite subsequence and for $t' = \infty$ a right-infinite subsequence. We consider only stationary processes, i.e. processes where $p(s_{[t,\infty]}) = p(s_{[t',\infty]})$, for any t, t'. Then, without loss of generality, one can write $\overleftarrow{S} := S_{[\infty,t]}$ for the *past* of the process and $\overrightarrow{S} := S_{[t,\infty]}$ for the *future* of the process, as well as $\overleftrightarrow{S}$ for the whole sequence.

Following (1.1), introduce an equivalence between different pasts $\overleftarrow{s}, \overleftarrow{s}'$ in predictiveness with respect to the future $\overrightarrow{S}$ (for detailed technical treatment of the semi-infinite sequences, see [12]). This induces a causal state $\tilde{S}$. Then, in [12] it is shown that, if a realization $\tilde{s}$ of the causal state induced by the past $s_{[-\infty,t]}$ is followed by a realization s of S_{t+1}, the subsequent causal state induced by $s_{[-\infty,t+1]}$ is uniquely determined. This induces an automaton on the set of causal states, called the ϵ-machine. Then the statistical complexity $C_\mu(\tilde{S})$ (the entropy of the ϵ-machine) measures how much memory the process stores about its past.

As opposed to that, one can consider the *excess entropy* of the process, defined by $E = I(\overleftarrow{S}; \overrightarrow{S})$. The excess entropy effectively measures how much information the past of a process contains about the future. It can be easily

[1] Note that the partition consists of subsets of $\mathcal{X}$. However, we will use the partition itself later as a new state space and therefore the individual equivalence classes $\tilde{x}$ are both subsets of $\mathcal{X}$ and states of the new state space $\tilde{\mathcal{X}}$.

shown that $E \leq C_\mu(\tilde{S})$. In other words, the amount the past of the process "knows" at a point about its future cannot exceed the size $C_\mu(\tilde{S})$ of the internal memory of the process.

Armed with these notions, in [12] the following definition of emergence is suggested: define $E/C_\mu(\tilde{S}) \in [0,1]$ as a measure of *predictive efficiency*, that is, how much of the internal process memory is used to actually predict what is going to happen. If we consider a *derived process* induced by a projection applied to each member of the sequence $\overleftrightarrow{S}$, this derived process is then called *emergent* if it has a higher predictive efficiency than the process it derives from. In particular, emergence is an intrinsic property of the process and does not depend on a subjective observer.

1.3.4 Emergent Descriptions

The *emergent description* model developed by the author in [10] takes an approach that, while related to predictive efficiency, differs from it in some important aspects. In the emergent descriptions model, consider again a sequence of random state variables $\overleftrightarrow{S}$. Assume that there exists a collection of k probabilistic mappings each inducing a sequence of random variables $\overleftrightarrow{S}^{(i)}$, with $i = 1 \ldots k$, forming a decomposition of the original system $\overleftrightarrow{S}$.

Then [10] defines $\overleftrightarrow{S}^{(k)}$ to be an *emergent description* for $\overleftrightarrow{S}$ if the decomposition $S_t^{(i)}$, fulfils three properties $\forall i = 1 \ldots k$:

1. the decomposition represents the system fully:
   $I\big(S_t; S_t^{(1)}, \ldots S_t^{(k)}\big) = H(S_t))$;

2. the individual substates are independent from each other:
   $I\big(S_t^{(i)}; S_t^{(j)}\big) = 0$ for $i \neq j$;

3. and they are individually information conserving through time
   $I\big(S_t^{(i)}; S_{t+1}^{(i)}\big) = H\big(S_{t+1}^{(i)}\big)$.

Similarly to the predictive efficiency from §1.3.3, the emergent description formalism considers the predictivity of a time series which is measured by mutual information. However, the emergent description model only deals with a system without a past, unlike the predictive efficiency model which uses ϵ-machines and thus includes full causal histories. However, a much more important difference is that the emergent description model explicitly considers a *decomposition* of the total dynamical system into individual independent informational components. Rather than considering the system as a unstructured "bulk", this view perceives it as having an inner informational dynamics and a natural informational substructure. Similar to the emergence notion from §1.3.3 this substructure is not externally imposed, but rather an intrinsic property of the system. It is, however, not necessarily unique. Fig. 1.1 shows schematically the decomposition into emergent descriptions.

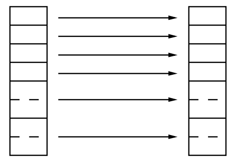

Figure 1.1: Schematic structure of emergent description decomposition into independent modes.

The emergent description model has the advantage that it can be explicitly constructed due to its quantitative characterisation (§1.4.1). This is a considerable advantage to more conceptual abstract models (such e.g. the category-theoretic approach mentioned in §1.2).

1.3.5 Other Related Approaches

Two further related approaches should be mentioned. The authors in [9] suggest emergence as higher-level prediction models for partial aspects of a systems, based on entropy measures. This model can be viewed as a simplified version both of the predictive efficiency and of the emergent description model. Compared with Crutchfield/Shalizi's predictive efficiency, it does not consider causal states, and compared with our emergent description model, it does not consider a full decomposition into independent information modes.

As opposed to that, [7] construct a decomposition into dynamical hierarchies based on smooth dynamical systems. This model is close in philosophy to the emergent descriptions approach, except for the fact that it is not based on information theory.

1.4 Emergent Descriptions: Construction and Generalizations

1.4.1 Construction

The quantitative character of the emergent description model provides an approach to construct (at least in principle) an emergent description (or at least an approximation) for a given system. Consider a system with 16 states, starting with a equally distributed random state and with a deterministic evolution rule $s_{t+1} := s_t + 1 \mod 16$, i.e. it acts as a counter modulo 16. We attempt to find an emergent description of the system into 2 subsystems of size 4 (these values have been chosen manually), applying a multiobjective Genetic Algorithm (GA) [2] to find projections that maximize system representation (criterion 1) and individual system prediction (criterion 3). With the given parameters, the optimization implicitly also optimizes criterion 2.

The multiobjective optimization fully achieves criterion 1 and comes close in maximizing criterion 3[2]. The search is far from fully optimal and can easily be improved upon. However, it provides a proof-of-principle and it demonstrates several issues of relevance that are discussed below. The dynamics of one of the emergent descriptions found is shown in Fig. 1.2.

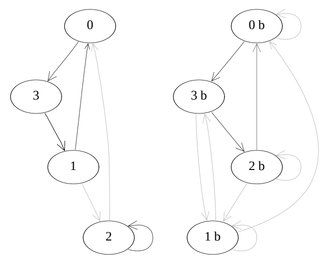

Figure 1.2: Automaton found in multiobjective GA search. The two groups of states belong to the two components of the emergent description, and the arrows indicate the stochastic transitions, where darker arrows indicate higher transition probabilities.

The left automaton, if the GA had been fully successful, would have shown a perfect 4-cycle, i.e. a counter modulo 4, with only deterministic transitions; the failure of the GA to find this solution is due to the deceptiveness of the problem. However, the right counter, the lower-level counter, can never be fully deterministic according to the model from §1.3.4. Like a decadic counter, perfectly predicting the transition to the next state in right counter would ideally depend on the carry from the left counter; but the indepence criterion does not allow the right counter to “peek” into the left, thus always forcing a residual stochasticity.

1.4.2 Hierarchical Emergent Descriptions

It turns out that this observation has a highly relevant relation to the algebraic Krohn-Rhodes semigroup decomposition [3]. Here, it turns out that the most general decomposition of a semigroup has a particular hierarchical structure that comprises a high-level substructure (group or flip-flop) which does not depend on anything else) and then a successive hierarchy of substructures each of which may depend on all the structures above them, the simplest example illustrated by a counter such as above[3].

To incorporate this insight into the emergent description model, one could modify the conditions from §1.3.4 to respect the possible Krohn-Rhodes structure of the system. Schematically, this would correspond to a decomposition of the kind shown in Fig. 1.3[4].

[2]In fact, the GA fails to find the best solution since the problem is GA-deceptive.
[3]It also bears some algebraic relation to the Jacobian decomposition discussed in [7].
[4]It is evident how to formalize this diagram in the spirit of §1.3.4.

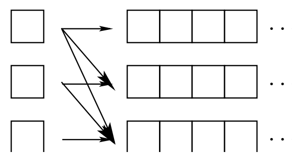

Figure 1.3: Emergent description with hierarchical dependence of states similar to Krohn-Rhodes decomposition.

1.4.3 Emergent Descriptions with History

In the present model system there is, however, a way to recover the independence of modes and maintain an optimal predictiveness. Note that in the emergent description model we completely banished state history. If we readopt it similar in spirit to component-wise ϵ-machines, then the components can count individually whether a carry is required or not. The idea is schematically represented in Fig. 1.4.

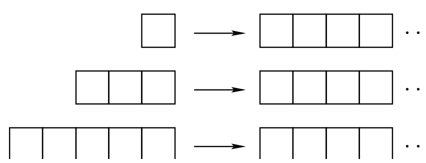

Figure 1.4: Emergent description with state histories.

1.4.4 Discussion

We have contrasted the ϵ-machine approach to characterize emergence with that of the emergent descriptions. The approaches are in many respects orthogonal, as the ϵ-machine creates a relation of the two full half-axes of the temporal coordinate without any decomposition of the states itself, while the emergent description approach limits itself to a single time slice, however suitably decomposing the state into independent modes. This approach has however been shown to lose some predictivity even in the very simple counter scenario. As a remedy one can introduce either a hierarchical form of emergent descriptions, inspired by the Krohn-Rhodes decomposition, or else aim for an ϵ-machine like history memory for the individual modes which is a kind of marriage of the emergent prediction and the ϵ-machine models.

In particular, this observation suggests the hypothesis that it might be possible to formulate a trade-off: one the one hand the memory requirements that a "serial" computation model such as the ϵ-machine needs to compute the future from the past; on the other hand the information processing resources required by the "parallel" computation model such as the hierarchical emergent descriptions which involves combining information from different components of the decomposition to compute a component's future. It is quite possible that universal trade-offs may exist here, offering the possibility for resource optimization and also for studying generalized forms of ϵ-machines where computational resources can be shifted more-or-less freely between temporal and compositional degrees of freedom.

1.5 Agenthood

As a final comment, it should be mentioned that in [8] it has been shown that the perception-action loop of an agent acting in an environment can be modeled in the language of information. This is particularly interesting for above considerations, as the agent/environment system is a generalization of a time series (a time series can be considered an agent without the ability to select an action, i.e. without the capacity for "free will").

Using infomax principles, above agent/environment system can be shown to structure the information flows into partly decomposable information flows, a process that can be interpreted as a form of concept formation. This gives a new interpretation for the importance of emergence as the archetypical mechanism that allows the formation of concept in intelligent agents and may thus provide a key driving the creation of complexity in living systems.

Bibliography

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

[2] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6:182–197, 2002.

[3] A. Egri-Nagy and C. L. Nehaniv. Making sense of the sensory data — coordinate systems by hierarchical decomposition. In *Proc. KES 2006*, 2006.

[4] H. Haken. *Advanced synergetics*. Springer-Verlag, Berlin, 1983.

[5] H. Haken. *Information and Self-Organization*. Springer Series in Synergetics. Springer, 2000.

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

[7] M. N. Jacobi. Hierarchical organization in smooth dynamical systems. *Artificial Life*, 11(4):493–512, 2005.

[8] A. S. Klyubin, D. Polani, and C. L. Nehaniv. 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, 2004.

[9] S. McGregor and C. Fernando. Levels of description: A novel approach to dynamical hierarchies. *Artificial Life*, 11(4):459–472, 2005.

[10] D. Polani. Defining emergent descriptions by information preservation. In *Proc. of the International Conference on Complex Systems*. NECSI, 2004. Long abstract, full paper under review in InterJournal.

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

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

[13] S. Winter. Zerlegung von gekoppelten Dynamischen Systemen (Decomposition of Coupled Dynamical Systems). Diploma thesis, Johannes Gutenberg-Universität Mainz, 1996. (In German).