An information-theoretic approach to self-organisation: Emergence of complex interdependencies in coupled dynamical systems

Fernando Rosas 1,2*, Pedro A.M. Mediano 3, Martin Ugarte 4 and Henrik J. Jensen 1,5

1 Centre of Complexity Science and Department of Mathematics, Imperial College London, UK
2 Department of Electrical and Electronic Engineering, Imperial College London, UK
3 Department of Computing, Imperial College London, UK
4 CoDE Department, Université Libre de Bruxelles, Belgium
5 Institute of Innovative Research, Tokyo Institute of Technology, Japan

* Correspondence: f.rosas@imperial.ac.uk; Tel.: +44 (0)20 7589 5111

Abstract: Self-organisation lies at the core of fundamental but still unresolved scientific questions, and holds the promise of de-centralised paradigms crucial for future technological developments. While self-organising processes have been traditionally explained by the tendency of dynamical systems to evolve towards specific configurations, or attractors, we see self-organisation as a consequence of the interdependencies that those attractors induce. Building on this intuition, in this work we develop a theoretical framework for understanding and quantifying self-organisation based on coupled dynamical systems and multivariate information theory. We propose a metric of global structural strength that identifies when self-organisation appears, and a multi-layered decomposition that explains the emergent structure in terms of redundant and synergistic interdependencies. We illustrate our framework on elementary cellular automata, showing how it can detect and characterise the emergence of complex structures.

Keywords: Self-organisation; multivariate information theory; coupled dynamical systems; partial information decomposition; high-order correlations; multi-layer complexity

1. Introduction

1.1. Context

It is fascinating how some systems acquire organisation spontaneously, evolving from less to more organised configurations in the absence of centralised control or an external driver. In a world constricted by the second law of thermodynamics and driven by “no free lunch” principles, self-organisation phenomena dazzle us by creating structure seemingly out of nowhere. Besides this aesthetic dimension, self-organisation plays a key role at the core of out-of-equilibrium statistical physics [1], developmental biology [2], and neuroscience [3]. Additionally, self-organisation serves as inspiration for new paradigms of de-centralised organisation where order is established spontaneously without relying on an all-knowing architect or a predefined plan, such as with the Internet of Things [4,5] and blockchain technologies [6]. In this context, self-organisation is regarded as an attractive principle for enabling robustness, adaptability and scalability into the design and management of large-scale complex networks [7–9].

Originally, the notion of self-organisation was introduced in the field of cybernetics [10,11]. These seminal ideas quickly propagated to almost all branches of science, including physics [1,12], biology [2,13], computer science [14,15], language analysis [16,17], network management [18,19], behavioral analysis [20,21] and neuroscience [22,23], to name a few. Despite this success, most working
definitions of self-organisation still avoid formal definitions and rely on intuitions following an “I know when I see it” logic, which might eventually prevent further systematic developments [24]. Formulating formal definitions of self-organisation is challenging, partly because self-organisation has been used in diverse contexts and with different purposes [25], and partly due to the fact that the basic notions of self and organisation are already problematic themselves [26].

The absence of an agreed formal definition, combined with the relevance of this notion for scientific and technological advances, generates a need for further explorations about the principles of self-organisation.

1.2. Scope of this work and contribution

In the spirit of Ref. [27], we explore to what extent an information-theoretic perspective can illuminate the inner workings of self-organising processes. Due to the connections between information theory and thermodynamics [28,29], our approach can be seen as an extension of previous works that relate self-organisation and statistical physics (see e.g. [30–32]). In previous research, self-organisation is typically associated with a reduction in the system’s entropy [30,33,34] —in contrast, we argue that entropy reduction alone is not a robust predictor of self-organisation, and additional metrics are required.

One contribution of this work is to characterise self-organising processes using multivariate information-theoretic tools —or, put differently, to provide a more fine-grained description of the underlying phenomena behind entropy reduction. We argue that self-organising processes are driven by spontaneous creation of interdependencies, while the reduction of entropy is a mere side effect of this. Following this rationale, we propose the binding information [35] as a metric of the strength of the interdependencies in out-of-equilibrium dynamical systems.

Another contribution of our framework is to propose a multi-layered metric of organisation, which combines quantitative and qualitative aspects. Most proposed metrics of “complexity” try to map the whole richness of possible structures into a single dimension [36]. In contrast, drawing inspiration from theoretical neuroscience [37,38], we put forward a multi-dimensional framework that allows for finer and more subtle taxonomy of complex systems. Our framework builds on ideas based on the Partial Information Decomposition (PID) framework [39], which distinguishes various information sharing modes in which the binding information is distributed across the system. This fundamental distinction overcomes counterintuitive issues of existent multiscale metrics for structural complexity, such as the one reported in Refs. [40,41], including negative information values that do not have operational meaning.

A final contribution of our work is to establish a novel connection between information theory and dynamical systems. The standard bridge between these two disciplines is established using symbolic dynamics and corresponding tools, including the Kolmogorov-Sinai entropy, Rényi dimensions and related concepts [42]. In contrast, in this paper we propose to apply information-theoretic analyses over the statistics induced by invariant measures over the attractors. In this way, attractors can be seen as statistical structures that generate interdependencies between the system’s coordinates. This statistical perspective can enrich and complement standard analyses of attractors based on fractal dimensions and other geometrical concepts.

The rest of this paper is structured as follows. First, Section 2 briefly introduces the key ideas of this work. Then, Section 3 discusses fundamental aspects of the definition of self-organisation and coupled dynamical systems. Section 4 presents the key ideas of our information-theoretic approach, which are then developed quantitatively in Section 5. Our framework is illustrated in Section 6 with an application to elementary cellular automata. Finally, Section 7 summarises our findings and presents our main conclusions.
2. Key intuitions

This section introduces the key ideas of our framework in an intuitive fashion. These ideas are made rigorous in the following sections.

2.1. The marble sculpture analogy

The overall configuration of a group of agents can be represented by a probability distribution over the set of their possible configurations. Independent agents who are maximally random are characterised by “flat” distributions (technically, distributions that satisfy the maximum entropy principle [28]). The temporal evolution of the system then shapes this distribution, in the same way as a sculptor shapes a flat piece of marble into a sculpture. In our view, the shape of the resulting distribution encodes the key properties that emerge from the temporal dynamics of the system, and a substantial part of our framework is to provide tools to measure and describe various types of sculptures. Importantly, just as the sculptor reveals a figure by removing the superfluous marble that is covering it, the temporal evolution generates interdependencies not by adding anything but by reducing the randomness/entropy of the system.

2.2. Self-organisation versus dissipation

Consider two agents with random and uncorrelated initial states, as in the analogy above. Their joint entropy, which quantifies their collective randomness, can be depicted as two circles, the size of the circle being proportional to how random each agent is. The circles are shown disjoint to reflect the fact that the agents are initially uncorrelated. From this initial situation, there are two qualitatively different ways in which their joint entropy can decrease: the state of each agent could become less random in time, while their independency is preserved; or the agents could become correlated while their individual randomness is preserved. Although both cases show overall entropy reduction, one needs to distinguish finer features of the shape of the resulting distribution to discriminate between genuine self-organisation in the latter scenario and mere dissipation in the former.

![Figure 1](commons.wikimedia.org). Left: Two flow functions corresponding to two dynamical systems, denoted by $\Phi$ and $\Phi'$, seen as a sculptor who takes away “superfluous” entropy/marble to let structures appear from inside. The figure shows The Atlas and The Bearded Slave (circa 1525–30) by Michelangelo Buonarroti, who was famous for letting his figures emerge from the marble “as though surfacing from a pool of water” [43] (pictures taken from commons.wikimedia.org). Right: Likewise, the joint entropy of two (or more) agents could decrease either because they become less random individually ($\Phi'$), or because they become correlated ($\Phi$). In this article we provide tools to measure how complex systems shape distributions as entropy is reduced—or marbled is carved out—from the initial state.
3. The goal and constraints of self-organisation

Which dynamical properties enable agents to self-organise? Beyond superficial differences, most studies on self-organisation agree that in order to have proper self-organisation three fundamental principles must be satisfied:

1. **Global structure**: the system evolves from less to more structured collective configurations.
2. **Autonomy**: agents evolve in the absence of external guidance.
3. **Horizontality**: no single agent can determine the evolution of a large number of other agents.

The principles of autonomy and horizontality constitute constraints, in the sense that a system that is not autonomous or horizontal cannot be called self-organising. Conversely, the principle of global structure is closer to a goal to be achieved. Hence, one could reformulate the above definition of self-organisation as the following optimisation problem:

\[
\text{generate} \quad \text{Global structure} \\
\text{subject to} \quad \text{Autonomy and horizontality.}
\] (1)

The following subsections provide a formalisation of these three fundamental principles.

3.1. Multiple agents as a coupled dynamical system

An elegant way to formalise these ideas is provided by the literature of coupled dynamical systems. Loosely speaking, a dynamical system is a process that evolves in time, such that its present configuration determines its future evolution following a deterministic rule [44]. Differential equations and finite difference equations are examples of dynamical systems. Furthermore, a collection of dynamical systems are said to be coupled if the future state of each process is affected not only by its own state but also by the state of other processes.

Let us consider a system composed by \( N \) parts or subsystems, which we call “agents” adopting the terminology from the robotics and multi-agent systems literature. The set of possible states for the \( k \)-th agent is denoted as \( \Omega_k \), and hence the set of possible configurations of the system is \( \Omega := \prod_{k=1}^{N} \Omega_k \), henceforth called “phase space.” The configuration of the system at time \( t \in T \subset [0, \infty) \) is determined by the vector \( x_t = (x_1^t, \ldots, x_N^t) \in \Omega \), where \( x_k^t \in \Omega_k \) is the corresponding state of the \( k \)-th agent and \( T \) is a collection of time indices. By assuming that the agents constitute coupled dynamical systems, the evolution of the group of agents is determined by a collection of flow functions \( \{\phi_t^{(h)}\} \) with \( h \geq 0 \), where \( \phi_t^{(h)} : \Omega \to \Omega \) drives the evolution of the system such that \( x_{h+t} = \phi_t^{(h)}(x_h) \). Intuitively, \( h \) corresponds to an initial time and \( t \) is the length of the evolution process.

Please note that the choice of (deterministic) coupled dynamical systems has basis of our framework has been made for simplicity of the presentation. The generalisation of our ideas and methods to stochastic dynamics is straightforward.

3.2. Formalising self-organisation

We now discuss aspects of the formalisation of (1) based on the language of dynamical systems.

(A) Autonomy

Intuitively, we say that a system is autonomous if it has no architect or “mastermind” controlling its evolution from the outside. Using the dynamical systems language introduced above, we can readily define necessary conditions for the autonomy of a system: we say that a system is autonomous if the collection of flow functions \( \{\phi_t^{(h)}\} \) are time-invariant —i.e. if its temporal evolution looks the same independently of the start time \( h \). Technically, autonomy requires that \( \phi_t^{(h_1)}(x_0) = \phi_t^{(h_2)}(x_0) \) for any \( h_1, h_2 \in T \) and \( x_0 \in \Omega \). This symmetry ensures that there is no organising influence guiding the
system from outside. In the rest of this manuscript time translation symmetry is assumed, which allows us to disregard the starting time and drop the superscript \((h)\), using \(\phi_i\) as a shorthand notation.\(^1\)

(B) Horizontality and locality

Intuitively, horizontality implies a similar restriction within the system itself, in the sense that no small set of units should control or influence the behaviour of the rest of the system. However, in contrast with the simplicity with which autonomy can be addressed, the formalisation of horizontality is substantially more challenging. Our approach here is to take a stronger condition than horizontality, namely:

3b. **Locality**: agents can only interact with a small number of other agents.

Locality is a sufficient condition for horizontality, since if no agent can interact with many other agents then the direct influence of each agent is limited. Conveniently, locality can be elegantly addressed within the framework of coupled dynamical systems. To do this, let us first introduce the notation \(\phi_i^k\) for the \(k\)-th coordinate of the flow function \(\phi_i\), i.e. \(\phi_i(x) = (\phi_i^1(x), \ldots, \phi_i^N(x))\). Then, one can define the interaction network between agents as follows: there exists a link from agent \(i\) to agent \(j\) if \(\phi_j^i(x)\) is affected by changes in the values of \(x_i\), the \(i\)-th coordinate of \(x\).

These directed networks can be encoded by an \(N \times N\) adjacency matrix \(A = [a_{ij}]\), where \(a_{ij} = 1\) if the \(i\)-th agent is connected with the \(j\)-th agent and zero otherwise. Locality is, hence, equivalent to \(A\) having sparse rows, imposing a fixed bound restricting the number of non-zero entries in each row.

In the following we assume locality, and leave the formalisation of horizontality for future work.

(C) Structure

One of the biggest challenges in the formalisation of self-organisation is to address the notion of **structure**. A large portion of the literature employs this concept without developing a formal definition of it, relying only on intuitive understanding. Furthermore, authors from different fields point towards this same intuition using related but different concepts, including global behaviour, organisation, coordination, or pattern.

Existent approaches to attempt a formalisation of the notion of structure use either attractors, or minimal description length and Kolmogorov complexity. These approaches, and their drawbacks, are discussed in Appendix A. Our own approach, which relies in multivariate information theory, is presented in the next Section.

4. Structure as multi-layered statistical interdependency

This section presents our framework to study emergence of structure in coupled dynamical systems. The key idea in our approach is to understand structure as statistical interdependency. Therefore, we regard patterns as deviations from statistical independence, i.e. as interdependent random variables. As argued below, these statistical interdependencies are best described using tools from multivariate information theory.

Adopting an information-theoretic perspective requires a step of abstraction, namely to place the analysis not in trajectories but in ensembles, as explained in Section 4.1. Then, Section 4.2 explores the relationship between the dynamics of the joint Shannon entropy and the increase of statistical interdependency. This discussion is further developed by introducing a decomposition of the Shannon entropy, developed in Sections 4.3 and 4.4.

---

\(^1\) Additionally, autonomous flows allow simple descriptions. In effect, thanks to the property \(\phi_i(\phi_j(x)) = \phi_j \circ \phi_i(x)\), autonomous evolutions in discrete time can be characterised by the single mapping \(\phi := \phi_1\) by noting that \(\phi_n = (\phi)^n\), while autonomous evolutions in continuous time can be characterised by a set of time-invariant differential equations.
For simplicity of exposition, in the rest of the paper we will focus in the case of discrete phase space $\Omega$. However, most of our results still hold for continuous dissipative systems.

### 4.1. From trajectories to stochastic processes

Traditionally, the study of dynamical systems is fundamentally built on how individual trajectories explore the space of possible system configurations. However, the information-theoretic perspective works not over trajectories but over ensembles (i.e. probability distributions). As a matter of fact, associating entropy values to trajectories is usually problematic, as it involves a number of ad-hoc (and often unacknowledged) assumptions. We make our assumptions explicit, and develop our analysis on an ensemble of systems initialised with stochastic initial conditions. The technicalities behind this approach are developed in the sequel.

Let us consider the case where the initial condition of the system is not a particular configuration $x_0 \in \Omega$, but an ensemble of configurations described by a probability distribution $\mu_0$. Interestingly, the flow operator $\phi_t$ not only induces a flow on the space of configurations $\Omega$, but it also induces another flow on the space of all possible probability distributions over $\Omega$, denoted as $\mathcal{M}(\Omega)$. Consider, as an example, the discrete distribution $\mu_0 = \sum_{i=1}^{\infty} c_i \delta_{x_i}$ where $\delta_{x_i}$ is the Kroneker delta. For this measure, the probability of a subset of configurations $A \subset \Omega$ is calculated by $\mu_0(A) = \sum_{i=1}^{\infty} c_i \mathds{1}_{x_i}(A)$, where $\mathds{1}_{x_i}(A) = 1$ if $x_i \in A$ and zero otherwise. A natural time-evolution of this probability distribution is given by $\mu_t = \sum_{j=1}^{N} c_j \mathds{1}_{\phi_t(x_j)}$. One can generalise this construction for an arbitrary initial probability distribution $\mu_0$, by introducing a flow operator $\Phi_t \{ \cdot \} : \mathcal{M}(\Omega) \rightarrow \mathcal{M}(\Omega)$ defined as

$$\Phi_t \{ \mu_0 \}(A) := \mu_0(\phi_t^{-1}(A)) = \mu_0(\{ x \in \Omega | \Phi_t(x) \in A \}) .$$

Note that $\{ \Phi_t \{ \cdot \}, t \in T \}$ induce a flow over $\mathcal{M}(\Omega)$, and hence constitute a new dynamical system.

The collection of probability distributions $\{ \mu_t = \Phi_t \{ \mu_0 \}, t \in T \}$ induces a corresponding multivariate stochastic process $X_t = (X^1_t, \ldots, X^N_t) = \phi_t(X_0)$, which follows a joint probability distribution $p_{X_t} = \mu_t$ (for the complete statistics of $X_t$ and technical details of this correspondence, see Appendix B). Note that the properties of this stochastic process are completely determined by the initial distribution $\mu_0$ and the flow function $\phi_t$. Each sub-process $X^k_t$ describes the uncertainty related to the state of the agent $k$ at time $t$, the statistics of which are found by marginalising the joint statistics of $p_{X_t}$. The aim of the next subsections is to explore the statistical interdependencies that can exist among these sub-processes.

### 4.2. Information dynamics

The joint Shannon entropy of the system at time $t$, given by $H(X_t) := - \sum_{x \in \Omega} p_{X_t}(x) \log p_{X_t}(x)$, corresponds to the information required to resolve the uncertainty about the state of the system at time $t$ (see Appendix C). The uncertainty reflected by this entropy has two sources. One source is stochasticity in the initial condition, i.e. when the initial configuration of the system at time $t = 0$ is not fully determined, but only prescribed by a probability distribution. The second source of uncertainty are stochastic dynamics (also known as “dynamical noise”), i.e. when the time evolution could make the system potentially transit from a single starting configuration to two or more different future

---

2 As a matter of fact, the measure-theoretic objects that are more studied within dynamical system theory (namely, invariant measures [45]) are distributions that are derived from mean values over trajectories.

3 Technically speaking, a specific sequence of symbols has no Shannon entropy or mutual information because it involves no uncertainty. The literature usually associates a value of entropy by considering a stochastic model which most likely generated the sequence. However, this practice relies on strong assumptions (e.g. ergodicity or independence of successve symbols), which might not hold in practice.

4 Interestingly, there exists a subset of $\mathcal{M}(\Omega)$ that is isomorphic to $\Omega$, namely the set of distributions of the form $\{ \mu_x = \mathds{1}_x | x \in \Omega \}$. Therefore, it is consistent to call $\mathcal{M}(\Omega)$ a generalised state space, which corresponds to the notion of “state” that is used by quantum mechanics [46].
configurations. Dynamical systems have deterministic transitions, and hence only exhibit the first type of uncertainty.

When considering discrete phase spaces, the deterministic dynamics guarantee that the uncertainty due to random initial conditions cannot increase; it can only decrease or be conserved. As a simple example, let us consider a dynamical system with a single point attractor: even if one does not know where a trajectory starts, one knows that the trajectory ends in the attracting point. In this case, any information encoded in the initial condition is vanished by the dynamics, as one cannot find out where trajectories are coming from. We call this phenomenon “information dissipation,” which mathematically can be stated as:

$$H(X_t) \geq H(X_{t+h}) \quad \text{for all } h > 0 \ .$$ (3)

Due to the deterministic nature of the time evolution, Eq. (3) is guaranteed by the data-processing inequality \[48\]. This decrease in entropy is not in contradiction with the second law of thermodynamics, as these systems are generally open and connected to an environment \[46,49\]. The equality in Eq. (3) is attained by closed systems, being this a direct consequence of the well-known Liouville theorem \[50\].

Information dissipation is directly related with the action of attractors. For a given attractor \(A\), its basin of attraction \(B(A)\) is the largest subset of \(\Omega\) such that \(\lim_{n \to \infty} \phi_t(x) = A\) for all \(x \in B(A)\). Intuitively, any trajectory starting in \(B(A)\) asymptotically runs into \(A\). Similarly, the evolution of an initial distribution \(\mu_0\) supported on \(B(A)\) eventually ends up being supported almost only on \(A\) when \(t\) is large enough; correspondingly, its Shannon entropy tends to decrease due to the reduced portion of the phase space where the system is confined to dwell. As such, information dissipation (i.e. entropy decreasing due to the action of attractors) is a necessary condition for self-organisation.

It is tempting to postulate entropy reduction as a strong indicator of self-organisation, based on a loose interpretation of entropy as a metric of disorder. However, the relationship between entropy and disorder is problematic, as disorder has different meanings in various contexts and there exists no single widely accepted definition for it. Moreover, entropy reduction is not a sufficient condition for self-organisation \[24\]. For example, consider a group of uncoppied damped oscillators initialised for self-organisation \[24\]. For example, consider a group of uncopled damped oscillators initialised
Following the above rationale, the total “non-shared information” in the system is nothing more than the sum of the non-shared information of every agent, and corresponds to the residual entropy \[35\]:

\[
R(X_t) := \sum_{j=1}^{N} H(X_j^t | X_{-j}^t) .
\] (5)

One can verify that the agents are statistically independent at time \( t \) if and only if \( H(X_t) = R(X_t) \). The complement of the residual entropy corresponds to the binding information \[33\], which quantifies the part of the joint Shannon entropy that is shared among two or more agents. This can be computed as

\[
B(X_t) := H(X_t) - R(X_t) = H(X_t) - \sum_{j=1}^{N} H(X_j^t | X_{-j}^t) .
\] (6)

Note that the above formula corresponds to a multivariate generalisation of the information-theoretic identity \( I(X;Y) = H(X,Y) - H(X|Y) - H(Y|X) \), which captures linear and non-linear dependencies that might exist between two or more agents. As such, the binding information is one of several multivariate generalisations of the mutual information, and is the only one known to enable a non-negative decomposition of the joint entropy \[54\].

In summary, the binding information provides a natural metric of the strength of the statistical interdependencies within a system. Interestingly, although the deterministic nature of deterministic dynamical systems constrains \( H(X_t) \) to be non-increasing, both \( B(X_t) \) and \( R(X_t) \) can increase or decrease. In contrast with the entropy, an increase in binding information is an unequivocal sign that statistical structures are being generated within the system by its temporal evolution.

4.4. The anatomy of the interdependencies

Although the binding information provides an attractive information-theoretic metric of organisation strength, a one-dimensional description is not rich enough to describe the range of phenomena observed in self-organising agents. To obtain a more detailed picture we use the Partial Information Decomposition (PID) framework, which allows us to develop a finer decomposition of the binding information and distinguish between different modes of information sharing. Originally, PID was introduced to study various aspects of information-theoretic inference, which consider a target variable predicted using the information provided by a number of information sources (see Refs. \[39,55–57\] and references therein). A key intuition introduced by these works is to distinguish between various information modes: in particular, redundant information corresponds to information about the target variable that can be retrieved from more than one source, and synergistic information corresponds to information that becomes available only when two or more sources are accessed simultaneously.

Traditional PID approaches divide the variables between target and sources, having each of them a very different role in the framework. Nevertheless, it is possible to propose symmetric decompositions of the joint Shannon entropy using PID principles that avoid these dialectic labellings \[54,58,59\]. In this case, the total information encoded in the system’s configuration is decomposed in redundant, unique and synergistic components. Redundancy takes place when measuring a single agent allows the observer to predict the state of other agents. Synergy corresponds to high-order statistical effects that can constrain groups of variables without imposing low-order restrictions. This idea of synergistic information is a generalisation of the well-known fact that random variables can be pairwise independent while being jointly interdependent. The relationship between synergistic information and high order correlations in the context of statistical physics has been explored in Refs. \[54,55\].

The work reported in \[54\] describes a decomposition for the binding information for the case of systems of \( N = 3 \) agents. In the sequel, we extend these ideas postulating three formal decompositions of the binding information for larger system sizes. Please note that our approach here is not to establish
precise formulas for computing the value of the components of these decompositions for arbitrary underlying probability distributions. Instead, in Section 5.1 we provide universal upper and lower bounds for these components, which need to be satisfied irrespective of the chosen functional form. Moreover, these bounds can be used in some cases to determine exact values of the decomposition’s components, as illustrated in Section 6.

Decomposition by extension of sharing

Since the binding information is the information shared by two or more agents, it is natural to discriminate exactly how many agents are involved in the sharing. Following this rationale, we propose the following decomposition:

\[ B(X_i) = \sum_{n=2}^{N} b^n(X_i), \quad (7) \]

where \( b^n(X_i) \) measures the portion of the binding information that is shared among exactly \( n \) agents. The index \( n \) refers to the amount of agents that are linked by the corresponding relationship. Therefore, \( b^n(X_i) \) quantifies the strength of interdependencies that link groups of \( n \) agents.

To illustrate these ideas, let us explore some simple examples where this decomposition can be computed directly from our desiderata.

**Example 1.** Consider two independent Bernoulli random variables \( U \) and \( V \) with parameter \( p = 0.5 \) (i.e. \( H(U) = H(V) = 1 \)). Then,

(i) If \( (X_1^1, X_2^1, X_3^1) = (U, U, U) \), then \( R(X_i) = 0 \) and \( B(X_i) = H(U) \). Furthermore, because of the triple identity \( b^3(X_i) = H(U) \), and hence \( b^2(X_i) = 0 \).

(ii) If \( (X_1^2, X_2^2, X_3^2) = (U, U, V) \), then \( R(X_i) = H(V) \) and \( B(X_i) = H(U) \). In this case, \( b^2(X_i) = H(U) \) and hence \( b^3(X_i) = 0 \).

(iii) If \( (X_1^3, X_2^3, X_3^3) = (U, V, U \lor V) \), then \( R(X_i) = 0 \) and \( B(X_i) = H(U) + H(V) \). Furthermore, due to the triple interdependency \( b^3(X_i) = H(U) + H(V) \), and hence \( b^2(X_i) = 0 \).

Decomposition by sharing modes

Following Ref. [54], we distinguish between redundant and synergistic information sharing modes. Redundancy, in this context, refers to information that is disclosed as soon as any of the agents who participate in the sharing are measured. Put differently, if agents are engaged in a redundant information sharing mode then measuring one of them allows the observer to make inferences on the states of the others. Conversely, synergistic information sharing takes place when accessing the state of one agent is not enough to obtain predictive power, i.e. to infer the state of the other agents involved. The key element is then how many agents need to be measured in order to obtain information about the other agents. Synergistic relationships require two or more, implying high-order statistical effects.

Based on these ideas, we postulate the following decomposition:

\[ b^n(X_i) = \sum_{i=1}^{n-1} l^n_i(X_i), \quad (8) \]

where \( l^n_i(X_i) \) denotes information that is shared between \( n \) agents, and becomes fully available after accessing \( i < n \) of the agents involved in the sharing. In other words, \( i \) is the smallest number of agents that enables the use of the information that corresponds to \( l^n_i(X_i) \) for predicting the state of the remaining \( n - i \) agents. Note the use of upperscripts and lowerscripts differentiate between group sizes and order of the sharing mode. This decomposition introduces a range of \( (i, n) \)-interdependencies, where \( n \) is the extension of the interdependency (how many agents are involved) while \( i \) is the “degree
of synergy.” With this notation, redundancies correspond to \( i = 1 \)-interdependencies, while \( I^i_t (t) \) for \( i \geq 2 \) are synergies of order \( i \).

Based on these ideas, another way of decomposing the binding information is by focusing on the possible information sharing modes, i.e. ways in which information can be shared among the agents according to \( i \). By combining Eqs. (7) and (8), one can then present the following decomposition:

\[
B(X_t) = \sum_{n=2}^{N} \sum_{i=1}^{n-1} I^i_t (X_t) = \sum_{i=1}^{N-1} m_i(X_t) ,
\]

where \( m_i(X_t) := \sum_{\alpha \notin I} I^\alpha_t (X_t) \) corresponds to information sharing modes that are fully accessed when measuring sets of \( i \) agents. In particular, \( m_1(X_t) \) collects all the “redundancies” of the system, i.e. sharing modes that are fully accessed by measuring only one of the agents involved in the sharing. Correspondingly, the terms \( m_i(X_t) \) for \( i \geq 2 \) convey the strength of synergies and high-order effects.

To contrast these ideas with the previous decomposition, we study the same scenarios from Example 1 under this new perspective.

**Example 2.** Consider \( U \) and \( V \) as defined in Example 1. Then,

(i) If \( (X_1^t, X_2^t, X_3^t) = (U, U, U) \), then \( m_1(X_t) = H(U) \), as the information contained in any variable allows to predict the others, while \( m_2(X_t) = 0 \).

(ii) If \( (X_1^t, X_2^t, X_3^t) = (U, U, V) \), then similarly as above \( m_1(X_t) = H(U) \) and \( m_2(X_t) = 0 \). Both cases are redundancies (same \( i \) of disimilar extension (different \( n \)).

(iii) If \( (X_1^t, X_2^t, X_3^t) = (U, V, U (aor V)) \), then only when measuring two agents one becomes able to predict the third one (c.f. Ref. [54]). This implies that \( m_2(X_t) = H(U) + H(V) \), and hence \( m_1(X_t) = 0 \).

5. A quantitative method to study time-evolving organisation

In this section we leverage the ideas discussed in Section 4 to develop a method to conduct a quantitative and qualitative analysis of self-organisation in dynamical systems. The goal of this method is twofold: to detect when self-organisation is taking place, and to characterise it as redundancy- or synergy-dominated. For this, Section 5.1 first develops upper and lower bounds for the terms of the decompositions of the binding information presented in Section 4.4. Then, Section 5.2 outlines a protocol of four steps that can be applied in practical scenarios.

5.1. Upper and lower bounds for the information decompositions

5.1.1. Bounds for the decomposition for extension

Let us define \( a_L = (a_1, \ldots, a_L) \) to be a vector of integer indices with \( 1 \leq a_1 < a_2 < \cdots < a_L \leq N \), and \( B(X^{a_L}_t) \) to be the binding information of the agents that correspond to those indices at time \( t \), i.e.

\[
B(X^{a_L}_t) = H(B(X^{a_L}_t)) - \sum_{j=1}^{N} H(X^{a_L}_t | X^{a_1}_t, \ldots, X^{a_{j-1}}_t, X^{a_{j+1}}_t, \ldots, X^{a_L}_t) ,
\]

where \( X^{a_L}_t = (X^{a_1}_t, \ldots, X^{a_L}_t) \). Also, let us denote as \( \mathcal{I}_L \) the set of all index vectors \( a_L \) of length \( L \), which correspond to the possible subsets of \( L \) agents of length \( L \), with cardinality \( |\mathcal{I}_L| = \binom{N}{L} \).

Recall that \( b^L(X_t) \) corresponds to information that is shared exactly by \( n \) agents, and hence \( \sum_{n=2}^{L} b^n(X_t) \) is the information shared by \( L \) or less agents. As \( B(X^{a_L}_t) \) corresponds to the information shared between agents \( a_1, \ldots, a_L \), it is clear that for any \( L \in \{2, \ldots, N\} \) the following bounds hold:

\[
\max_{a_L \in \mathcal{I}_L} B(X^{a_L}_t) \leq \sum_{n=2}^{L} b^n(X_t) \leq \sum_{a_L \in \mathcal{I}_L} B(X^{a_L}_t) \leq \binom{N}{L} \max_{a_L \in \mathcal{I}_L} B(X^{a_L}_t) .
\]
Although the last bound might not be tight, these inequalities show that \( \max_{\alpha L \in I_L} B_{\alpha L}(t) \) is a good proxy for sizing the value of \( \sum_{i=2}^{L} b^\alpha(t) \). In particular, if \( \max_{\alpha L \in I_L} B_{\alpha L}(t) = 0 \) then \( b^\alpha(X_t) = 0 \) for all \( n = 2, \ldots, L \), which due to Eq. (7) would imply that \( B(X_t) = \sum_{n=L+1}^{N} b^\alpha(X_t) \).

5.1.2. Bounds for the decomposition by sharing modes

Let us recall that \( m_i(X_t) \) accounts for the information about other agents that is obtained when measuring groups of \( i \) agents, but not less. Similarly, \( \sum_{i=1}^{L} m_i(X_t) \) is the predictability about other agents that is obtained when accessing \( L \) or less agents. Therefore, one can provide the following bounds valid for any \( L \in \{1, \ldots, N-1\} \):

\[
\psi_L(t) \leq \sum_{i=1}^{L} m_i(X_t) \leq \sum_{j=1}^{N} \sum_{\substack{\alpha L \in I_L \ \alpha_i \neq \alpha_j}} I(X_t^\alpha L; X_t^\alpha_j) \leq N \binom{N-1}{L} \psi_L(t) ,
\]

where we have used the shorthand notation

\[
\psi_L(t) := \max_{j \in \{1, \ldots, N\}} \max_{\alpha L \in I_L \ \alpha_i \neq \alpha_j} I(X_t^\alpha L; X_t^\alpha_j).
\]

As in Eq. (11), this shows that \( \psi_L(t) \) can be used as a proxy for estimating the relevance of \( \sum_{i=1}^{L} m_i(X_t) \). In particular, if \( \psi_L(t) = 0 \) then \( \sum_{i=1}^{L} m_i(X_t) = 0 \). Therefore, by using Eq. (9), if \( \psi_L(t) = 0 \) then all the binding information is composed by synergies of order \( L + 1 \) or more.

This discussion suggests that the properties of \( \psi_L(t) \) can reveal the distribution of sharing modes across the system. First, note that \( \psi_L(t) \) is a non-decreasing function of \( L \): information (in the Shannon sense) “never hurts,” and hence having larger groups of agents for making predictions cannot reduce predictive power. Secondly, in most scenarios \( \psi_L(t) \) is concave: the additional predictability obtained by including one more agent usually shows diminishing returns as \( L \) grows. In effect, the most informative agents are normally selected first, and hence for large values of \( L \) one can just add agents with weak informative power, which can also be redundant with the agents already considered. Accordingly, scenarios where \( \psi_L(t) \) as function of \( L \) is concave are called redundancy-dominated. In contrast, scenarios in which \( \psi_L(t) \) is convex are called synergy-dominated. Intuitively, in synergy-dominated scenarios agents might be uninformative by themselves, but become informative when grouped together. Therefore, a convex \( \psi_L(t) \) is a sign of a synergistic system, one that has larger predictability gains when \( L \) grows.

5.2. Protocol to analyse self-organisation in dynamical systems

Wrapping up these results, we propose the following definitions for self-organisation. Note that these are aimed at quantifying organisation, while the constraints of “self” are guaranteed by restricting to autonomous flow functions with sparse interaction matrices (see Section 3.2).

**Definition 1.** Consider a coupled dynamical system with autonomous evolution and a bounded number of non-zero elements per row in its interaction matrix. Then, the system is self-organising if \( B(X_t) \) is an increasing function of \( t \). Moreover, the value of \( B(X_t) \) is used as a metric of organisation strength.

**Definition 2.** A self-organising process is said to be synergy-dominated if, for a large \( t \), \( \psi_L(t) \) is convex as function of \( L \). If \( \psi_L(t) \) is concave, then the process is said to be redundancy-dominated.

Following these definitions, we propose the following protocol for analysing a given dynamical system. The steps are:

1. Check that the flow functions satisfy autonomy and locality (Section 3.2).
(1) Consider a random initial condition given by a uniform distribution over the phase space, \( \mu_0 \), and use it to drive the coupled dynamical system. This involves initialising the system in the least biased initial configuration, i.e. with maximally random and independent agents.

(2) Compute the evolution of the probability distribution given by \( \mu_t = \Phi_t(\mu_0) \). This can be done directly using the flow function, a master equation [60], or in the case of a finite phase space by computing numerically all the trajectories.

(3) Compute the joint Shannon entropy \( H(X_t) \), the residual information \( R(X_t) \) and the binding information \( B(X_t) \) as a function of \( t \).

(4) For values of \( t_0 \) at which \( B(X_{t_0}) > 0 \), compute \( \psi_L(t_0) \) for \( L = 1, \ldots, N \).

Note that by considering a flat initial condition in step (1), one ensures that the system initially has no correlations, i.e. \( B(X_0) = 0 \). Therefore, if one finds that \( B(X_t) > 0 \) for some \( t > 0 \), one can be sure that these interdependencies were entirely created by the dynamics of the system. Also, while step (3) clarifies if self-organisation is taking place following Definition 1, (i.e. by checking if \( B(X_t) > 0 \) for some \( t > 0 \)), step (4) discriminates between redundant and synergistic organisation structures according to Definition 2.

6. Proof of concept: Cellular automata

Cellular Automata (CA) are a well-known class of discrete coupled dynamical systems widely used in the study of complex systems and distributed computation [61]. A CA is a multi-agent system in which every agent has a finite set of possible states, and evolves in discrete time steps following a set of simple rules based on its own and other agents’ states.

CA are a natural candidate for our measures, since they have been often used in other studies of self-organisation [62], some of them are capable of universal computation [63], and they provide an rich testbed for theories of distributed computation and collective behaviour in complex systems [64].

6.1. Method description

Our analysis focuses on Elementary Cellular Automata (ECA), which constitute a particular subclass of CA. In ECA, agents (or cells) are arranged in a one-dimensional cyclic array (or tape). The state of each cell at a given time step has two possible states, 0 or 1, and is a boolean function of the state of itself and its immediate neighbours at the previous time step. The same boolean function dictates the time evolution of all agents, inducing a spatial translation symmetry. Hence, each of the 256 different boolean functions of three binary inputs induces a different evolution rule. Rules are then enumerated from 0 to 255 and each ECA, irrespective of its number of agents, can be classified by its rule. For a more detailed description of ECA and their numbering system, see Ref. [62].

In our simulations, we followed the protocol outlined in Section 5.2 over arrays of \( N \) cells that followed one ECA rule. We initialised one copy of the ECA in each of the \( 2^N \) possible initial conditions and numerically computed the temporal evolution of each one of them. As is standard in the ECA literature, the automata were simulated under periodic boundary conditions. The probability distribution at time \( t, \mu_t \), was calculated until the distribution reached a pseudo-stationary regime, using methods outlined in Appendix D. These methods allowed us to consider arrays up to size \( N = 17 \).

Our analysis of the ECAs included the following elements:

a) The temporal evolution of \( H(X_t), B(X_t) \) and \( R(X_t) \). These plots show if the ECA shows signs of self-organisation according to Definition 1, and if the joint entropy decreases or remains constant (c.f. Section 4.2).

b) The interdependency between individual cells through time, as given by the mutual information between a single cell at time \( t = 0 \) and all other cells in the same and successive times (i.e.
\( I(X_0^t, X_k^t) \) for \( t \in \{0, 1, \ldots \} \) and \( k \in \{1, \ldots, N\} \). This reflects the predictive power of the state of a cell in the initial condition over the future evolution of the system.\(^5\)

c) The mutual information between every pair of cells for the pseudo-stationary distribution.

Because of the spatial translation symmetry of ECA, it suffices to take any cell and compute its mutual information with each other cell. We call this “spatial correlation,” as it measures interdependencies between cells at the same time \( t \).

e) The curve \( \psi_L \) introduced in Section 5, that is used to characterise a self-organising system as either redundancy- or synergy-dominated as per Definition 2. This curve can also be interpreted as how much of a cell can be predicted by the most informative set of \( L \) other cells.

6.2. Results

Now we present and discuss the profiles of three well-known rules, which illustrate paradigmatic behaviour. Figures show results of ECAs with \( N = 17 \) agents. Extended versions of these results for all rules with \( N = 4, \ldots, 17 \) agents can be found in https://cellautomata.xyz.

**Strong redundancy: rule 232**

Rule 232 is commonly referred to as the majority rule, as one cell’s next state is 1 if and only if two or more of its predecessors are 1. The dynamics of this rule when starting from a random initial condition are governed by interactions between nearest neighbours, which are resolved after few steps into stable configurations (Figure 2a). As a result of this brief interaction, the dynamics generate binding information while decreasing the joint entropy, as shown in Figure 2d.

In agreement with those observations, it is found that one cell at the initial condition has high predictive power over the state of itself and its nearest neighbours in the future (Figure 2b). Correspondingly, the profile of pairwise mutual information terms between cells at the pseudo-stationary regime shows exponentially decaying correlations as a function of cell distance (Figure 2c).

The curve of \( \psi_L \) shows a concave shape, growing strongly for the first two (nearest) neighbours, growing slightly for the third and fourth nearest neighbours, and remaining then essentially flat (Figure 4). This means that remote neighbours are practically independent, which is consistent with the pairwise correlation profile. Note that knowing all the other cells provides an 80% prediction over a given cell, meaning that there is a non-negligible amount of residual entropy.

**Synergistic profile: rule 30**

Rule 30 is known for generating complex geometric patterns, and has a sensitive dependence to initial conditions [65]. This rule, among others, has provided key insights to understand how simple rules can generate complex structures. For example, similar patterns can be found in the shell of the *conus textile* cone snail species. Rule 30 has also been proposed as a stream cipher for cryptography [66], and has been used as a pseudo-random number generator [67].

Visual inspection suggests that the information processing done by this rule is much more complex than Rule 232. In effect, Figure 3d shows that this rule generates high \( B(X_1^t) \) through a much longer mixing time. Intriguingly, the predictive information of a single cell seems to disappear after very few steps (Figure 3b), meaning that knowing the state of a single cell of the initial condition is not useful for predicting the state of any cell at later stages. Even more intriguingly, the pseudo-stationary regime shows that each pair of cells is practically independent (Figure 3b), in direct contrast with the high value of \( B(X_1^t) \).

\(^5\) To use an analogy, one can think of the information content of a cell as a drop of ink that is thrown into the river of the temporal evolution of the system.
Figure 2. Combined results for rule 232. (a) Example of evolution starting from random initial conditions. (b) The predictability (measured by the mutual information) enabled by knowing the initial condition of one cell allows to predict the same cell in the future and it nearest neighbours. (c) Profile of pairwise mutual information terms between cells at the pseudo-stationary regime shows a typical exponential decay. (d) Time evolution generates interaction reflected by $B(X_t)$, which is of the same order of magnitude than $R(X_t) = H(X_t) - B(X_t)$. Both $B$ and $H$ are reported in bits.

These apparent paradoxes are solved when one considers high-order correlations by studying the behaviour of $\psi_L$ (Figure 4). In effect, the convex shape of the curve shows a pronounced synergistic structure: groups of less than 8 cells show no interdependency, but groups of 15 allow almost perfect prediction! This proves that the self-organisation driven by Rule 30 generates high-order structures. In particular, for arrays of 17 cells, the fact that $\sum_{j=1}^{9} m_j(X_t) \approx 0$ implies that the interdependencies are synergies of order 10 or higher.

**Pure synergy: rule 90**

Rule 90 consists of concatenated xor logic gates: the future state of each cell corresponds to the xor of its two predecessors. When started from a single active cell, Rule 90 generates a Sierpinsky triangle, while when started from a random initial condition it generates irregular triangular patterns. Rule 90 is known for having connections with number theory, as discussed in Ref. [62].

Rule 90 was found to be the most synergistic rule of all 256 ECA. In fact, for an array of $N$ cells started with random initial conditions, after the second step any group of $N - 1$ cells or less is statistically independent. This implies that $\psi_L = 0$ for all $L < N$, and therefore $m_L(X_t) = 0$ for all $L < N - 1$ (Figure 4). However, our calculations show that $R(X_t) = 0$ while $B(X_t) = H(X_t) = N - 1$, indicating that the binding information of Rule 90 corresponds exclusively to synergy of the highest order, i.e. $B(X_t) = m_N(X_t)$.

The fact that Rule 90 was found to have the highest synergy is consistent with the crucial role played by xor gates in cryptography. For a discussion on this connection, see Ref. [54, Section 4.2].
Figure 3. Combined results for rule 30. (a) Example of evolution starting from random initial conditions. (b) The predictability enabled by knowing the initial condition of one cell vanishes quickly, being useless to predict future states of single cells. (c) At the pseudo-stationary regime, there exists no mutual information between any pair of cells. (d) Despite having no significant pairwise correlations, the dynamics generate large amounts of interdependency between the cells reflected by a high value of $B(X_t)$. Both $B$ and $H$ are reported in bits.

7. Discussion

This paper presents an information-theoretic framework to study self-organisation in multi-agent systems, which explores how statistical structures are spontaneously generated by the evolution of coupled dynamical systems. To guarantee the absence of centralised control guiding the process, we restrict ourselves to autonomous systems where each agent can interact directly with only a small number of other agents. To isolate structures that are purely created by the system’s dynamics, we consider the evolution of agents that are initially maximally random and independent.

A fundamental insight behind our framework is the fact that deterministic dynamical systems are able to create correlations by destroying information. In effect, we saw that while the temporal evolution of many dynamical systems reduce their joint Shannon entropy, this condition can be the consequence of two qualitatively opposite scenarios: in one case interdependency is created while the stochasticity of each agents is preserved; and in the other mere information dissipation occurs (each agent becomes less random while remaining independent of each other). Following this line of thought, and diverging from the standard literature, we propose to attribute self-organisation to processes where the strength of interdependencies increases with time. In this work we use the binding information as metric of global interdependency strength.

As a second step, we propose a multi-layered description of the attained organisation based on synergies and redundancies of various orders. The key idea is to decompose the information stored in the system, as quantified by the joint Shannon entropy, considering two principles: extension (how many agents are linked), and sharing mode (how many agents need to be measured in order to obtain
Rule 232: Redundancy

Rule 30: Weak synergy

Rule 90: Strong synergy

**Figure 4.** While the concave shape of $\psi_L$ for Rule 232 shows that correlations are mostly redundant, the convex shape for Rule 30 shows the dominance of synergies of order 10 or more. Rule 90 is the only rule that generates purely synergistic structure of the highest order.

predictive power). The information sharing mode of order 1 corresponds to redundancy, which takes place when by measuring only one agent one can (partially) predict the state of a number of other agents. Synergy takes place when such predictive power is accessible only when measuring two or more agents. We proposed these decompositions as formal structures, without providing an explicit way to compute the values of their components for arbitrary probability distributions. Nevertheless, upper and lower bounds for these components are provided, which in some cases can allow a complete determination of the decomposition.

Our results enable a way of understanding self-organising processes that is consistent with the Bayesian interpretation of information theory, as described in Ref. [28]. Using our framework, this work is the first —to the best of our knowledge— to demonstrate cases of high-order synergy in relatively large systems. In particular, we showed that the ECA that corresponds to rule 90 generates maximal synergy, which, according Ref. [54, Section 4.2], could enable the development of interesting cryptographic applications.

An interesting extension of this work would be to use some of the recently proposed measures of synergy (see e.g. [68–70]) to build exact formulas for the proposed decompositions. However, this could prove to be challenging as most of these metrics are designed for systems of three variables with their extensions to larger systems not being straightforward. Another interesting extension is to apply the presented framework to study continuous coupled dynamical systems. Interestingly, while the entropy of continuous systems can be negative, the binding information is still a non-negative quantity and hence its decomposition is still meaningful. It would be of interest to explore if well-known chaotic attractors can be explained in terms of the synergies and redundancies they induce in the corresponding coordinates, which could hopefully provide a new link between chaos theory and multivariate information theory. These developments could, in turn, also allow us to build a bridge between the ideas presented in this paper and the extensive literature of self-organising coupled oscillators [71,72].

Finally, it is worth emphasising that the statistical character of our proposed framework makes it orthogonal to some well-established self-organisation principles, such as the enslaving principle for multi-scale systems [73] or the free energy principle for autopoietic organisms [23]. As a matter of fact, it remains to be explored to what extent those principles can be enriched by including multi-layered decompositions in terms of redundancies and synergies. Also, please note that the presented approach to self-organisation is restricted to structures that are generated within known possibilities, which is related to the idea of “weak emergence” [74]. An attractive extension would be to include phenomena related to “strong emergence,” i.e. processes in which evolution can affect the state space itself, generating entirely new configurations for the system to explore. An attractive way of attempting
this extension could be to combine the presented framework with the notion of super-exponentially growing phase spaces presented in Ref. [75].

**Author Contributions:** All authors participated in the development of the concepts, wrote and revised the manuscript. The numerical evaluations over ECA were carried out by M.U. All authors have read and approved the final manuscript.

**Funding:** FR was supported by the European Union’s H2020 research and innovation programme, under the Marie Skłodowska-Curie grant agreement No. 702981.

**Acknowledgments:** The authors thank Carlos Gershenson, Robin Ince, Ryan James and Michael Lachmann for helpful discussions.

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

**Appendix A. Alternative approaches to formalising global structure**

**Structure as geometrical properties of attractors**

A natural way to attempt a formalisation of structure is by relating it to the notion of attractor from the dynamical systems literature. To define what an attractor is, let us note that the set \( \{ x_s | x_s = \phi_s(x_0) \} \) corresponds to the trajectory that originates from the initial condition \( x_0 \in \Omega \). A set \( B \subset \Omega \) is stable if it contains the trajectories of all its elements, i.e. for all \( y_0 \in B \) and \( t \in \mathbb{N} \) we have that \( y_t = \phi_t(y_0) \in B \). An attractor is a set \( A \) that is stable and has no stable proper (non-empty) subsets. Common attractors are fixed points, limit cycles (i.e. periodic trajectories) and strange attractors [76].

Since the early efforts of Ashby [10,77], it has been noted that self-organisation is a consequence of the tendency of dynamical systems to evolve towards attractors. The system, hence, becomes more "selective" as time passes. Following this rationale, one can argue that the distinctive properties of these attracting configurations are the ones that emerge within the time-evolution.

Could one relate the attractor’s geometrical structure to properties of structure and organisation? An attractive fact is that “interesting” dynamics are usually associated with non-linear equations, which in turn generate strange attractors with exotic geometric properties; while on the other hand the attractors of linear dynamics have uninteresting geometrical structure. Following this line of thought, one could attempt to establish relationships between the geometrical structure of attractors (e.g. in terms of fractal structure) and properties of the organisation attained by the agents. Although this is plausible, the route to develop such endeavour is not straightforward.

**Structure as pattern complexity**

Interesting approaches for formalising the concept of structure or pattern can be found in the computer science and signal processing literature. One such approach is to relate pattern strength with the Kolmogorov complexity (KC) [78], which is the length of the shortest computer program that is able to generate the pattern as an output. In this way, pattern strength is inversely proportional to the value of the corresponding KC: very structured configurations can be generated by short programs, while random configurations with no structure can only be the output of a program of the same length of the sequence itself.

Using the KC to measure pattern strength is attractive due to its intuitiveness, and because its quantitative nature can allow comparisons between heterogenous structures [79]. Unfortunately, the KC has been proven to be not computable (this imposibility being related to Gödel’s incompleteness theorem [80]), which hinders its practical value\(^6\). Another weakness of this approach is that the KC does not have, to the best of our knowledge, properties about the relationship between the complexity

---

\(^6\) Please note that practical estimation of the KC can be attempted via upper bounds, which can be calculated using lossless compression algorithms [79].
of a system and the complexity of it parts, nor properties of how the KC evolves in time under diverse dynamical conditions. These two limitations are overcome by adopting an information-theoretic framework, which is done in the following section.

Appendix B. From a dynamical system to a stochastic process

In order to consider probabilities defined over the phase space $\Omega$, one needs to introduce a collection of "events," denoted by $B$, which is a set of measurable subsets of $W$. It is natural to ask this collection to have $\sigma$-field properties, so that if $F, G \in B$ then both $F \cup G \in B$ and $F \cap G \in B$ are guaranteed. A probability measure $\mu$ is a function $\mu : B \mapsto [0, \infty)$ such that $\mu(\Omega) = 1$ and $\mu \left( \bigcup_{j=1}^{\infty} B_j \right) = \sum_{j=1}^{\infty} \mu(B_j)$ if $B_j \in B$ and $\bigcap_{j=1}^{\infty} B_j = \emptyset$.

When considering a flow function $\phi_t$ over the phase space, it is natural to require $\phi_t$ and $B$ to match together appropriately, i.e. for $F \in B$ then $\phi_t^{-1}(F) = \{ x \in \Omega | \phi_t(x) \in F \} \in B$. In that way, one can guarantee the consistency of the definition given in (2).

Given a probability distribution $\mu_0$, any measurable function $Y : \Omega \mapsto \mathbb{R}$ can be considered to be a random variable with statistics defined as

$$\mathbb{P}\{ Y \in I \} := \mu(Y^{-1}(I)).$$

Above, $Y^{-1}(I) = \{ x \in \Omega | Y(x) \in I \}$ and $I \subset \mathbb{R}$. Similarly, the multivariate stochastic process $X_t = (X_1^t, \ldots, X_n^t)$ induced by the flow function $\phi$ is defined by the joint statistics are given by

$$\mathbb{P}\{ X_i^t \in I_i, \ldots, X_m^t \in I_m \} := \mu_0 \left( \bigcap_{j=1}^{m} \phi_t^{-1} \left( X_j^{-1}(I_j) \right) \right),$$

where $i_j \in \{1,2,\ldots,n\}$ are a collection of indices, $t_1, \ldots, t_m \in T$ is a collection of time points and $I_j \subset \mathbb{R}$ for $j = 1, \ldots, m$. Above, $\phi_t^{-1} \left( X_j^{-1}(I) \right) = \{ x \in \Omega | X_j(\phi_t(x) \in I) \}$, with $X_j(x)$ being the $i$-th coordinate of $x \in \Omega$. For discrete phase spaces, the joint probability distribution of $X_t$ is given by $p_{X_t}(x) = \mathbb{P}\{ X_t = x \}$ for $x \in \Omega$.

Appendix C. Information and entropy

In this work we follow a Bayesian interpretation of the Shannon information, as described in [28]. The entropy is a functional over the probability distribution that describes the state of knowledge that an observer has with respect to a given system of interest. In this context, uncertainty in the system corresponds to information that can be potentially extracted by performing adequate measurements.

Following this line of thought, the amount of information needed to specify a single configuration within $|\Omega|$ possibilities is $\log |\Omega|$, where the base of the logarithm can be chosen according to the preferred units for counting information (bits, nats, or others). If a system with a phase space of cardinality $|\Omega|$ at time $t$ follows a statistical distribution $p_{X_t}$, then this information gets divided as follows:[54]:

$$\log |\Omega| = H(X_t) + \mathcal{N}(X_t),$$

where $H(X_t) := -\mathbb{E}\{ \log p_{X_t}(X_t) \}$ is the joint Shannon entropy of the system, and $\mathcal{N}(X_t) := \log |\Omega| - H(X_t)$ is the “negentropy”. After an observer comes to know the statistics of the system, as encoded by $p_{X_t}$, the average amount of information needed to specify a particular configuration decreases from $\log |\Omega|$ to $H(X_t)$; therefore, the negentropy corresponds to the bits that are disclosed by the knowledge of the statistics. In contrast, the Shannon entropy measures the information that is not disclosed by the statistics, which can only be obtained when the configuration of the system is actually measured.

Appendix D. Simulation details

This appendix describes the procedure for calculating the evolution of probability distributions over ECAs (c.f. Section 7). The possible states of an ECA with $N$ cells were encoded as a binary
numbers, and hence the phase space corresponds to \( \Omega = \{0, \ldots, 2^N - 1\} \). Probabilistic distributions over the phase space were stored as an array \( L_\mu = (\mu(0), \ldots, \mu(2^N - 1)) \), where \( \mu(k) \geq 0 \) for all \( k \in \{0, \ldots, 2^N - 1\} \) and \( \sum_{k=0}^{2^N-1} \mu(k) = 1 \).

We considered trajectories over \( \Omega \), which correspond to sequences of binary numbers \((s_1, s_2, \ldots)\) such that \( s_{k+1} = \phi(s_k) \) with \( \phi : \Omega \rightarrow \Omega \) being the function that encodes the ECA rule. Our interest was to find the step at which \( \phi(\cdot) \) brings the trajectory back to a state that has been already visited before. Note that all trajectories end up in an periodic attractor, this being a consequence of the finiteness of \( \Omega \). From a trajectory starting at \( s \in \Omega \), we store the pair \((p_s, a_s)\) with \( p_s \) being the length of the trajectory until reaching a state in the periodic attractor, and \( a_s \) being the length of the periodic attractor (i.e. the number of states between the first and the second appearance of a repeated state). The interest of these numbers lays in the fact that

\[
\phi^t(s) = \phi^K(t,s)(s) \quad \forall t \in \{p_s, p_s + 1, \ldots\},
\]

where \( K(t,s) := p_s + (t - p_s) \mod a_s \).

The ECA is said to have reached a pseudo-stationary regime when it has been run for a number of steps \( t_s \) such that, for any initial distribution \( \mu_0 \), a trajectory of distributions \((\mu_0, \mu_1, \ldots, \mu_{t_s})\) obtained by time evolution would reach a distribution that has already been visited before\(^7\). The minimal number of steps needed to reach a pseudo-stationary regime, denoted by \( t_0 \), can be calculated as

\[
t_0 = \text{LCM}(\{a_s\}_{s \in \Omega}) + \max_{s \in \Omega} p_s,
\]

where LCM stands for the least common multiple. Above, the last term ensures that each state entered their periodic attractor, and the former is the smallest number of steps that guarantees a simultaneous full cycle of all the attractors. For the considered ECAs with \( N = 17 \) cells, the largest values found where \( t_0 \approx 1014 \).

In order to be able to study the statistics of ECAs under pseudo-stationary regimes, we developed an efficient way to compute the evolution of a given initial distribution \( \mu \) for very large number of steps. Let us represent the initial distribution \( \mu \) by the array \( L_\mu \), and the resulting distribution after \( t_0 \) steps as \( \mu' \) with its corresponding vectorial representation \( L_{\mu'} \). Our key idea is to compute the trajectory from each \( s \in \Omega \) only for \( K(t_0, s) \) steps —as additional multiples of \( a_s \) correspond to mere cycles over its periodic attractor. The general procedure for computing \( \mu' \) using this idea goes as follows:

1. Initialize the components of \( L_{\mu'} \) with zeros.
2. For each \( s \in \Omega \): compute \( s' = \phi^{K(t_0,s)}(s) \) and then add \( \mu(s) \) to \( \mu'(s') \) (i.e. add \( L_\mu[s] \) to \( L_{\mu'}[s'] \)).

This technique resulted to be very efficient, as the largest values found over all ECA rules for \( N = 17 \) were \( \max_{s \in \Omega} p_s = 1776 \) and \( \max_{s \in \Omega} a_s = 78821 \).

References

1. Haken, H. Synergetics: an introduction. Non-equilibrium phase transition and self-organisation in physics, chemistry and biology, 1983.
2. Camazine, S. Self-organization in biological systems; Princeton University Press, 2003.
3. Tognoli, E.; Kelso, J.S. The metastable brain. Neuron 2014, 81, 35–48.
4. Ding, Y.; Jin, Y.; Ren, L.; Hao, K. An Intelligent Self-Organization Scheme for the Internet of Things. IEEE Computational Intelligence Magazine 2013, 8, 41–53. doi:10.1109/MCI.2013.2264251.

---

\(^7\) As mentioned in Section 4.1, the set of all probability distributions \( \mu \) over \( \Omega \) together with their dynamics form a new dynamical system, which also has periodic attractors. From this point of view, \( t_s \) is the smallest integer such that all trajectories of distributions reach their periodic attractor.
5. Athreya, A.P.; Tague, P. Network self-organization in the Internet of Things. 2013 IEEE International Conference on Sensing, Communications and Networking (SECON), 2013, pp. 25–33. doi:10.1109/SAHCN.2013.6644956.

6. MacDonald, T.J.; Allen, D.W.; Potts, J. Blockchains and the boundaries of self-organized economies: Predictions for the future of banking. In Banking Beyond Banks and Money; Springer, 2016; pp. 279–296.

7. Prokopenko, M., Ed. Guided self-organization: Inception; Vol. 9, Springer Science & Business Media, 2013.

8. Kuze, N.; Kominami, D.; Kashima, K.; Hashimoto, T.; Murata, M. Controlling large-scale self-organized networks with lightweight cost for fast adaptation to changing environments. ACM Transactions on Autonomous and Adaptive Systems (TAAS) 2016, 11, 9.

9. Rosas, F.; Hsiao, J.H.; Chen, K.C. A technological perspective on information cascades via social learning. IEEE Access 2017, 5, 22605–22633.

10. Ashby, W.R. Principles of the self-organizing dynamic system. The Journal of general psychology 1947, 37, 125–128.

11. Foerster, H.v. On self-organizing systems and their environments. Self-organizing systems 1960, pp. 31–50.

12. Haken, H.; Jumarie, G. A macroscopic approach to complex system; Springer, 2006.

13. Crommelinck, M.; Feltz, B.; Goujon, P. Self-organization and emergence in life sciences; Springer, 2006.

14. Heylighen, F.; Gershenson, C. The meaning of self-organization in computing. IEEE Intelligent Systems 2003, 18.

15. Mamei, M.; Menezes, R.; Tolk, R.; Zambonelli, F. Case studies for self-organization in computer science. Journal of Systems Architecture 2006, 52, 443–460.

16. De Boer, B. Self-organization in vowel systems. Journal of phonetics 2000, 28, 441–465.

17. Steels, L. Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation. Approaches to the Evolution of Language 1998, pp. 384–404.

18. Prehofer, C.; Bettstetter, C. Self-organization in communication networks: principles and design paradigms. IEEE Communications magazine 2005, 43, 78–85.

19. Dressler, F. A study of self-organization mechanisms in ad hoc and sensor networks. Computer Communications 2008, 31, 3018–3029.

20. Kugler, P.N.; Kelso, J.S.; Turvey, M. On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. Tutorials in motor behavior 1980, 3, 3–47.

21. Kelso, J.S.; Schöner, G. Self-organization of coordinative movement patterns. Human Movement Science 1988, 7, 27–46.

22. Kelso, J.S. Dynamic patterns: The self-organization of brain and behavior; MIT press, 1997.

23. Friston, K. The free-energy principle: a unified brain theory? Nature Reviews Neuroscience 2010, 11, 127.

24. Shalizi, C.R.; Shalizi, K.L.; Haslinger, R. Quantifying self-organization with optimal predictors. Physical review letters 2004, 93, 118701.

25. Gershenson, C. Guiding the self-organization of random Boolean networks. Theory in Biosciences 2012, 131, 181–191.

26. Gershenson, C.; Heylighen, F. When can we call a system self-organizing? European Conference on Artificial Life. Springer, 2003, pp. 606–614.

27. Krakauer, D.; Bertschinger, N.; Olbrich, E.; Ay, N.; Flack, J.C. The information theory of individuality. arXiv preprint arXiv:1412.2447 2014.

28. Jaynes, E.T. Probability theory: the logic of science; Cambridge university press, 2003.

29. Mezard, M.; Montanari, A. Information, physics, and computation; Oxford University Press, 2009.

30. Nicolis, G.; Prigogine, I. Self-Organization in Non-Equilibrium Systems: From Dissipative Structures to Order Through Fluctuations; Wiley, 1977).

31. Heylighen, F.; others. The science of self-organization and adaptivity. The encyclopedia of life support systems 2001, 5, 253–280.

32. Pulselli, R.; Simoncini, E.; Tiezzi, E. Self-organization in dissipative structures: A thermodynamic theory for the emergence of prebiotic cells and their epigenetic evolution. Biosystems 2009, 96, 237–241.

33. Klimontovich, Y.L. Turbulent Motion. The Structure of Chaos. In Turbulent Motion and the Structure of Chaos; Springer, 1991; pp. 329–371.

34. Gershenson, C.; Fernández, N. Complexity and information: Measuring emergence, self-organization, and homeostasis at multiple scales. Complexity 2012, 18, 29–44.
35. Vijayaraghavan, V.S.; James, R.G.; Crutchfield, J.P. Anatomy of a spin: the information-theoretic structure of classical spin systems. *Entropy* 2017, 19, 214.

36. Lloyd, S. Measures of complexity: a nonexhaustive list. *IEEE Control Systems Magazine* 2001, 21, 7–8.

37. Tononi, G.; Sporns, O.; Edelman, G.M. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences* 1994, 91, 5033–5037.

38. Friston, K.J.; Tononi, G.; Sporns, O.; Edelman, G. Characterising the complexity of neuronal interactions. *Human Brain Mapping* 1995, 3, 302–314.

39. Williams, P.L.; Beer, R.D. Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515* 2010.

40. Bar-Yam, Y. Multiscale complexity/entropy. *Advances in Complex Systems* 2004, 7, 47–63.

41. Allen, B.; Stacey, B.C.; Bar-Yam, Y. Multiscale information theory and the marginal utility of information. *Entropy* 2017, 19, 273.

42. Beck, C.; Schögl, F. Thermodynamics of chaotic systems: an introduction; Number 4, Cambridge University Press, 1995.

43. Vasari, G. *The Lives of the Artists*; Vol. 293, Oxford University Press, 1991; pp. 58–59.

44. Robinson, R.C. *An introduction to dynamical systems: continuous and discrete*; Vol. 19, American Mathematical Soc., 2012.

45. Schuster, H.G.; Just, W. Deterministic chaos: an introduction; John Wiley & Sons, 2006.

46. Breuer, H.P.; Petruccione, F.; others. *The theory of open quantum systems*; Oxford University Press on Demand, 2002.

47. Schreiber, T.; Kantz, H. Noise in chaotic data: diagnosis and treatment. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 1995, 5, 133–142.

48. Cover, T.M.; Thomas, J.A. *Elements of information theory*; John Wiley & Sons, 2012.

49. Schuster, H.G.; Just, W. *Time's arrows and quantum measurement*; Cambridge University Press, 1997.

50. Martynov, G. *Liouville’s theorem and the problem of the increase of the entropy*. *Soviet Journal of Experimental and Theoretical Physics* 1995, 80, 1056–1062.

51. Slepian, D.; Wolf, J. Noiseless coding of correlated information sources. *IEEE Transactions on information Theory* 1973, 19, 471–480.

52. El Gamal, A.; Kim, Y.H. *Network information theory*; Cambridge university press, 2011.

53. Te Sun, H. Nonnegative entropy measures of multivariate symmetric correlations. *Information and Control* 1978, 36, 133–156.

54. Rosas, F.; Ntranos, V.; Ellison, C.J.; Pollin, S.; Verhelst, M. Understanding interdependency through complex information sharing. *Entropy* 2016, 18, 38.

55. Olbrich, E.; Bertschinger, N.; Rauh, J. Information decomposition and synergy. *Entropy* 2015, 17, 3501–3517.

56. Barrett, A.B. Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Physical Review E* 2015, 91, 052802.

57. Lizier, J.T.; Bertschinger, N.; Jost, J.; Wibral, M. Information Decomposition of Target Effects from Multi-Source Interactions: Perspectives on Previous, Current and Future Work. *Entropy* 2018, 20.

58. Rosas, F.; Ntranos, V.; Ellison, C.J.; Verhelst, M.; Pollin, S. Understanding high-order correlations using a synergy-based decomposition of the total entropy. Proceedings of the 5th joint WIC/IEEE Symposium on Information Theory and Signal Processing in the Benelux, 2015, pp. 146–153.

59. Ince, R.A. The Partial Entropy Decomposition: Decomposing multivariate entropy and mutual information via pointwise common surprisal. *arXiv preprint arXiv:1702.01591* 2017.

60. Van Kampen, N.G. *Stochastic processes in physics and chemistry*; Vol. 1, Elsevier, 1992.

61. Mitchell, M.; others. Computation in cellular automata: A selected review. *Nonstandard Computation* 1996, pp. 95–140.

62. Wolfram, S. *A new kind of science*; Vol. 5, Wolfram media Champaign, IL, 2002.

63. Wolfram, S. Universality and complexity in cellular automata. *Physica D: Nonlinear Phenomena* 1984, 10, 1–35.

64. Lizier, J. The local information dynamics of distributed computation in complex systems. PhD thesis, University of Sydney, 2010.

65. Cattaneo, G.; Finelli, M.; Margara, L. Investigating topological chaos by elementary cellular automata dynamics. *Theoretical computer science* 2000, 244, 219–241.
66. Wolfram, S. Cryptography with cellular automata. Conference on the Theory and Application of Cryptographic Techniques. Springer, 1985, pp. 429–432.

67. Wolfram, S. Random sequence generation by cellular automata. *Advances in applied mathematics* 1986, 7, 123–169.

68. Ince, R.A. Measuring multivariate redundant information with pointwise common change in surprisal. *Entropy* 2017, 19, 318.

69. James, R.G.; Ellison, C.J.; Crutchfield, J.P. dit: a Python package for discrete information theory. *Journal of Open Source Software* 2018.

70. Makkeh, A.; Theis, D.O.; Vicente, R. BROJA-2PID: A robust estimator for bivariate partial information decomposition. *Entropy* 2018, 20, 271.

71. Pikovsky, A.; Rosenblum, M.; Kurths, J.; Kurths, J. *Synchronization: a universal concept in nonlinear sciences*; Vol. 12, Cambridge university press, 2003.

72. Kuramoto, Y. *Chemical oscillations, waves, and turbulence*; Vol. 19, Springer Science & Business Media, 2012.

73. Haken, H. *Synergetics. Physics Bulletin* 1977, 28, 412.

74. Chalmers, D.J. Strong and weak emergence. *The reemergence of emergence* 2006, pp. 244–256.

75. Jensen, H.J.; Pazuki, R.; Pruessner, G.; Tempesta, P. Statistical mechanics of exploding phase spaces: Ontic open systems. *Journal of Physics A: Mathematical and Theoretical* 2018.

76. Strogatz, S.H. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*; CRC Press, 2018.

77. Ashby, W.R. Principles of the self-organizing system. In *Principles of Self-Organization: Transactions of the University of Illinois Symposium*; Foerster, H.V.; G. W. Zopf, J., Eds.; Springer, 1962; pp. 255–278.

78. Kolmogorov, A.N. Three approaches to the quantitative definition of information’. *Problems of information transmission* 1965, 1, 1–7.

79. Li, M.; Vitanyi, P. *An introduction to Kolmogorov complexity and its applications*, 3rd edition ed.; Springer Publishing Company, Incorporated, 2008.

80. Chaitin, G.J. *Information, randomness & incompleteness: papers on algorithmic information theory*, Vol. 8, World Scientific, 1990.