

Article

The Typical Set and Entropy in Stochastic Systems with Arbitrary Phase Space Growth

Rudolf Hanel ^{1,2} and Bernat Corominas-Murtra ^{3,*}
¹ Complexity Science Hub Vienna, Josefstädter Strasse 39, 1080 Vienna, Austria

² Section for Science of Complex Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria

³ Institute of Biology, University of Graz, Holteigasse 6, 8010 Graz, Austria

* Correspondence: bernat.corominas-murtra@uni-graz.at

Abstract: The existence of the *typical set* is key for data compression strategies and for the emergence of robust statistical observables in macroscopic physical systems. Standard approaches derive its existence from a restricted set of dynamical constraints. However, given its central role underlying the emergence of stable, almost deterministic statistical patterns, a question arises whether typical sets exist in much more general scenarios. We demonstrate here that the typical set can be defined and characterized from general forms of entropy for a much wider class of stochastic processes than was previously thought. This includes processes showing arbitrary path dependence, long range correlations or dynamic sampling spaces, suggesting that typicality is a generic property of stochastic processes, regardless of their complexity. We argue that the potential emergence of robust properties in complex stochastic systems provided by the existence of typical sets has special relevance to biological systems.

Keywords: entropy; non-exponential phase space growth; typical set; asymptotic equipartition property; extensivity


Citation: Hanel, R.;

Corominas-Murtra, B. The Typical Set and Entropy in Stochastic Systems with Arbitrary Phase Space Growth. *Entropy* **2023**, *25*, 350. <https://doi.org/10.3390/e25020350>

Academic Editor: Philip Broadbridge

Received: 24 October 2022

Revised: 3 February 2023

Accepted: 5 February 2023

Published: 14 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Many living systems are characterized by a high degree of internal stochasticity and display processes that form organization of growing complexity [1–4]. Such complexification processes exist on various scales, from the evolutionary scale [5,6], to the scale of single organisms [7]. The increase of complexity of the different forms of living entities triggered the debate whether the existence of *open-endedness* is a defining trait of biological evolution [8–14], with the resulting challenge of finding a potential statistical-physics-like characterization of it. At the single organism scale, the developmental process consists in the emergence of an adult multicellular organism from a single cell [7]. This astonishingly fast process of complexification sets challenges in many directions. One of these challenges is the characterization of the evolution of the space of potential configurations the system may acquire in time: In early embryo morphogenesis, for example, not only does the number of cells increase exponentially, resulting in the corresponding increase of potential configurations, but also cells differentiate into specialized cell types and create collective structures implying, in statistical physics language, that new states enter the system. This process is almost completely irreversible and, although highly precise, it is known to have a strong stochastic component [15–17]. Away from biology, non stable configuration spaces and processes of complexification have been identified in systems with innovation [18–20]. On the other hand, one can consider processes where the potential number of configurations decreases with time. Away from biology, recent advances in decay dynamics in nuclear physics were achieved considering a mathematical framework based on the stochastic collapse of the phase space [21–23]. In Figure 1 we schematically show the kind of processes we are exploring, all taking place in *dynamic phase spaces*, that is, processes where the amount of available possibilities may change (either grow or decrease)

in time. Despite the ubiquity of such phenomena, a comprehensive characterization of systems with dynamic phase spaces, in terms equivalent to the ensemble theory of statistical mechanics, is lacking.

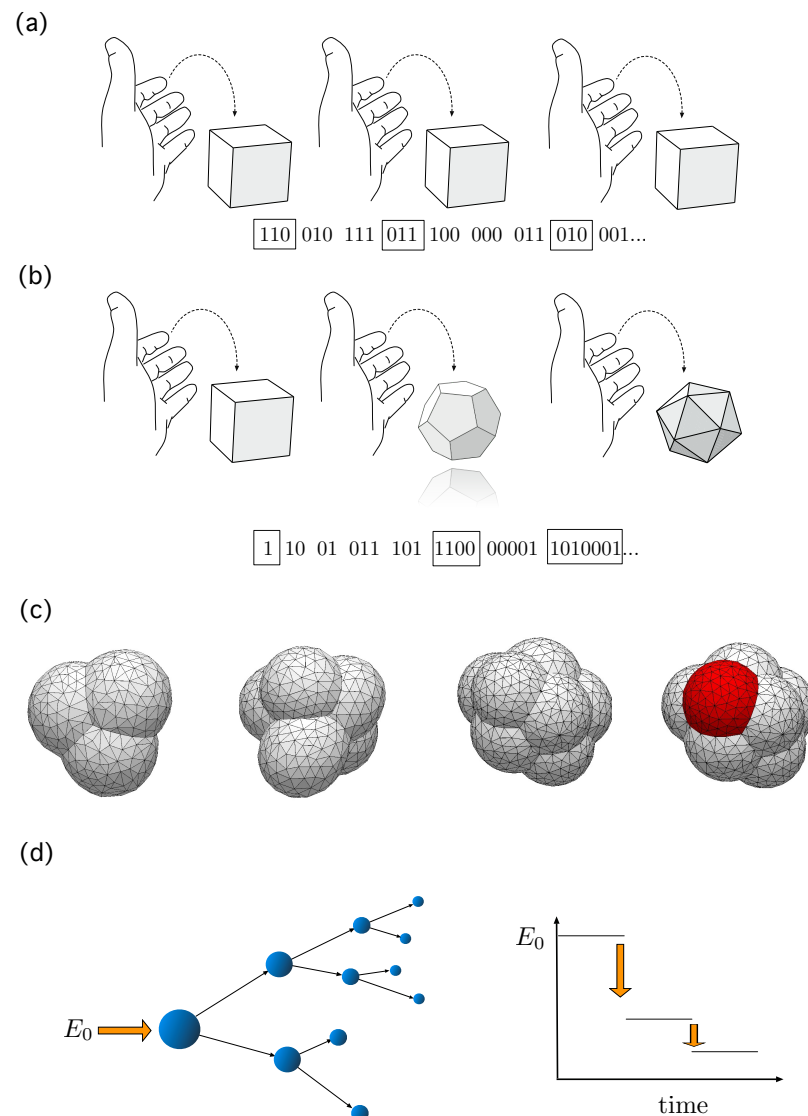


Figure 1. (a) Independent draws of the same dice, either fair or biased, define a i.i.d. stochastic processes whose typical set is well defined and grows following approximately an exponential trend [24]. (b) An example of a system whose typical set may show a super-exponential growth: At every drawing we update the dice by adding, e.g., a new face. (c) Potential configurations of early embryo development resembles, intuitively, the picture of the dice with growing number of faces. In this biological setting, new cells appear and, with that, new configurations but, on top of that, cells differentiate into new types—shown here in red—adding new states to the system that were not there before. Interestingly, even highly reproducible, the whole process displays a strong stochastic component [15]. (d) Nuclear disintegration can be studied from the framework of collapsing phase spaces [23]. In these processes, the amount of potential configurations of the system shrinks as long as the process unfolds. Toy models of embryo packings in (c) have been drawn using the *evolver* software package.

Ensemble formalism in statistical mechanics can be grounded in the concept of *typicality* [24–28]. Informally speaking, given the set of all potential sequences of events resulting from a stochastic process, a subset, the *typical set*, carries most of the probability [24,25]. This should not be confused with the set of most probable sequences: in the case of the biased

coin, for example, the most probable sequence is not in the typical set. In other words, for long enough sequences, the probability that the observed sequence or state belongs to the subset of sequences forming the typical set goes asymptotically to 1. Generalizing to continuous systems, this is known as the *concentration of measure* phenomenon [29]. Accordingly, a typical property for a stochastic system is robust and acts as a strong, almost deterministic attractor as the process unfolds [27], and one may expect to observe it in the vast majority of cases. Moreover, if such a typical property exists, one can use this single property to—at least partially—characterize the system, and hence avoid a detailed microscopic description of all of the system's components. Arguably, considerations based on typicality drive the connection between microscopic dynamics and macroscopic observables and underlie the existence of the thermodynamic limit [28,30,31]. In the context of information theory, the existence of the typical set for a given information source has deep consequences in the context of data compression [24,25].

The size of the typical set gives us valuable information on how the stochastic process is filling the phase space. In equilibrium systems or for information sources drawing independently from identically distributed (i.i.d.) random variables, the Gibbs–Shannon entropic functional arises naturally in the characterization of the typical set [24,25], establishing a clear connection between thermodynamics and phase space occupation. In systems/processes with collapsing or exploding phase spaces, path dependence or strong internal correlations [4,18–21,32–37], the phase space may grow super- or sub-exponentially, and the emergence of the Shannon–Gibbs entropic functional derived from phase space volume occupancy considerations is no longer guaranteed. The same situation may arise in cases dealing with non-stationary information sources [38–42]. Generalized forms for entropies have been proposed to encompass these more general scenarios [43–51], some of them explicitly linking the entropic functional to the expected evolution of the phase space volumes [36,46,52–56]. Despite the notable advances reported also for systems with physical significance [37,57,58], the concept of typicality has not been yet explored for systems/processes with exploding or shrinking phase spaces, which display path dependent dynamics or are subject to emergent internal constraints and correlations.

The purpose of this paper is to fill this important gap in the theory of stochastic processes, providing results with potential implications in the theory of non-equilibrium systems, data compression and coding strategies. As we shall see, the typical set can be defined for processes arbitrarily away from the i.i.d. framework, by only assuming a very generic convergence criteria, satisfied by a broad class of stochastic processes, that we refer to here as *compact stochastic processes*.

2. Results

2.1. Compact Stochastic Processes

Let us consider a general class of stochastic processes η [59,60]. This class encompasses almost any discrete stochastic process that can be conceived. A realization of t steps of the process is denoted as $\eta(t)$:

$$\eta(t) = \eta_1 \eta_2 \dots \eta_{t-1} \eta_t,$$

where $\eta_1, \eta_2, \dots, \eta_{t-1}, \eta_t$ are random variables themselves. Note that, in different realizations of t steps of the process, the sequence of random variables can be different, as the process may display path dependence, long term correlations, or changes of the state space dynamics itself (either shrinking or expanding). We denote a particular trajectory/path the process may follow as:

$$x(t) \equiv x_1 x_2 \dots x_{t-1} x_t \in \Omega(t),$$

$\Omega(t)$ being the set of all possible paths of the process η up to time t . We focus on the family of stochastic processes where there exists (i) a positive, strictly concave and strictly

increasing function $\Lambda \in \mathcal{C}^2$ in the interval $[1, \infty)$, such that $\Lambda(1) = 0$, and (ii) a positive, strictly increasing, $g \in \mathcal{C}^2$, in the interval $(1, \infty)$, by which:

$$\lim_{t \rightarrow \infty} \frac{1}{g(t)} \Lambda \left(\frac{1}{p(\eta(t))} \right) = 1, \quad (1)$$

where the convergence is in probability [60]. We will call this family of stochastic processes *compact stochastic processes* (CSP). Given a CSP process η , a pair of functions Λ, g by which Equation (1) is satisfied define a *compact scale* of the CSP process η . Note that these two functions may not be unique for a given process, meaning that the process can, in principle, have several compact scales.

It is straightforward to check that, if η is a sequence of i.i.d. random variables $X_1, \dots, X_t \sim X$, $\Lambda = \log$ and $g(t)$ is t times the Shannon entropy of a single realization, $H(X)$, the above condition holds, as it recovers the standard formulation of the Asymptotic Equipartition Property (AEP) [24,25]. Therefore, the drawing of i.i.d. random variables $\sim X$ is a CSP with compact scale $(\log, H(X)t)$. However, the range of potential processes that are CSPs is, in principle, much broader. In consequence, the first question we ask concerns the constraints that the convergence condition (1) imposes on Λ . Assuming that (1) holds, one finds that the candidates to characterize CSPs are the Λ s satisfying the following condition; see Proposition A1 of the Appendix B.2 for details:

$$\lim_{z \rightarrow \infty} \frac{\Lambda(\lambda z)}{\Lambda(z)} = 1, \quad \forall \lambda \in \mathbb{R}^+. \quad (2)$$

Typical candidates for Λ are of the form $\Lambda(z) = c(\log(z))^d$, where c, d are two positive, real valued constants or, more generally:

$$\Lambda(z) = c_1(\log(1 + c_2(\log(1 + c_3(\log(\dots))^{d_3}))^{d_2}))^{d_1},$$

where c_1, \dots and d_1, \dots are positive, real valued constants. In previous approaches, these constants have been identified as scaling exponents, enabling us to classify the different potential growth dynamics of the phase space [54]. We observe that the existence of the inverse function of Λ , Λ^{-1} , by which $(\Lambda^{-1} \circ \Lambda)(z) = z$, is guaranteed by the assumption made in the definition of CSPs that Λ is a strictly monotonically growing function.

The convergence condition defining CSPs has a direct consequence on how probabilities are distributed along the set of all potential paths. Indeed, from Equation (1) it follows that there exist two non-increasing sequences of positive numbers $\epsilon_1, \dots, \epsilon_t, \dots, \delta_1, \dots, \delta_t, \dots$, with $\lim_{t \rightarrow \infty} \epsilon_t = \lim_{t \rightarrow \infty} \delta_t = 0$, from which one can define a sequence of subsets of paths $A[\epsilon_1] \dots A[\epsilon_t]$ (such that $A[\epsilon_1] \subseteq \Omega(1), \dots, A[\epsilon_t] \subseteq \Omega(t)$) as follows: For all $x(t) \in A[\epsilon_t]$:

$$\frac{1}{\Lambda^{-1}((1 + \epsilon_t)g(t))} \leq p(x(t)) \leq \frac{1}{\Lambda^{-1}((1 - \epsilon_t)g(t))}, \quad (3)$$

and the probability of a given path to belong to $A[\epsilon_t]$ is bounded as:

$$\mathbb{P}(x(t) \in A[\epsilon_t]) > 1 - \delta_t, \quad (4)$$

where:

$$\mathbb{P}(x(t) \in A[\epsilon_t]) = \sum_{x(t) \in A[\epsilon_t]} p(x(t)).$$

We call the sequence of subsets $A[\epsilon_1] \dots A[\epsilon_t]$ of the respective sampling spaces $\Omega(1) \dots \Omega(t)$ a *sequence of typical sets* of η . Informally speaking, Equation (4) tells us that, for large enough t , the probability of observing a path that does not belong to the typical set becomes negligible. As a consequence, *the typical set can be identified for CSPs*: Given a CSP, the typical set $A[\epsilon_t]$ absorbs all the probability, in the limit $t \rightarrow \infty$. We summarize the above considerations in a theorem:

Theorem 1. If η is a CSP in (Λ, g) , then there exist (a) a non-increasing sequence of positive numbers $\epsilon_1, \epsilon_2, \dots, \epsilon_{t-1}, \epsilon_t, \dots$ with limit $\lim_{t \rightarrow \infty} \epsilon_t = 0$, associated to η and hence (b) the respective sequence of typical sets $A[\epsilon_1], \dots, A[\epsilon_t], \dots$ by which:

$$\lim_{t \rightarrow \infty} \mathbb{P}(x(t) \in A[\epsilon_t]) = 1.$$

Proof. Since η is a CSP we know, by assumption, that, as $t \rightarrow \infty$ also $\Lambda(1/p(\eta(t)))/g(t) \rightarrow 1$ (in probability). We can rewrite this condition by stating that, for every $\epsilon, \delta > 0$, there exists a t_0 by which, for each $t > t_0$ [60]:

$$\mathbb{P}\left(\left|\frac{1}{g(t)}\Lambda\left(\frac{1}{p(x(t))}\right) - 1\right| > \epsilon\right) < \delta.$$

Let $\tau(\epsilon, \delta)$ be the smallest such t_0 . We can then use two arbitrary strictly monotonic decreasing functions ϵ_n^* and δ_n^* that converge to zero and construct a monotonically increasing sequence of times $t_n = \tau(\epsilon_n^*, \delta_n^*)$ such that for all $t \geq t_n$ it is true that

$$\mathbb{P}\left(\left|\frac{1}{g(t)}\Lambda\left(\frac{1}{p(x(t))}\right) - 1\right| > \epsilon_n^*\right) < \delta_n^*.$$

From that, it is straightforward to define a non-increasing sequence $\epsilon_1, \dots, \epsilon_t, \dots$ converging to 0, by just taking:

$$(\forall t : t_n \leq t < t_{n+1}), \epsilon_t = \epsilon_n^*.$$

Finally, the condition:

$$\lim_{t \rightarrow \infty} \mathbb{P}(x(t) \in A[\epsilon_t]) = 1,$$

follows as a direct consequence of the construction of the sequence of typical sets $A[\epsilon_1], \dots, A[\epsilon_t], \dots$, thereby concluding the proof. \square

We omitted a direct reference to the process η in the notation of the typical set (i.e., $A[\epsilon_t] \equiv A[\epsilon_t](\eta)$) for the sake of readability, and we will explicitly refer to it only if it is strictly necessary. In the next section we provide more details on the specific bounds in size by studying a subclass of the CSPs, namely, the class of *simple* CSPs. For them, the characterization of the typical set can be achieved using generalized forms of entropy.

2.2. The Typical Set and Generalized Entropies

Equation (1) can be related to a general form of path entropy:

$$S_\Lambda(\eta(t)) = \sum_{x(t) \in \Omega(t)} p(x(t)) \Lambda\left(\frac{1}{p(x(t))}\right), \quad (5)$$

It can be proven that S_Λ satisfies three of the four Shannon–Khinchin axioms expected by an entropic functional [25,61,62] in Khinchin’s formulation [62], to be referred to as SK1, SK2, SK3. In particular SK1 states that entropy must be a function of the probabilities, which is satisfied by S_Λ , by construction. SK2 states that S_Λ is maximized by the uniform distribution q over $\Omega(t)$, i.e.,

$$q(x(t)) = \frac{1}{|\Omega(t)|}.$$

Finally, SK3 states that, if $p(x(t)) = 0$, then $p(x(t))$ does not contribute to the entropy, which implies:

$$\lim_{p(x(t)) \rightarrow 0} p(x(t)) \Lambda\left(\frac{1}{p(x(t))}\right) = 0,$$

satisfied as well for any Λ considered in the definition of the CSPs.

We further observe that S_Λ is a monotonically increasing function as well, in the case of uniform probabilities: Let us suppose two CSPs η and η' that sample uniformly their

respective sampling spaces, $\Omega(t), \Omega'(t)$, such that $|\Omega(t)| < |\Omega'(t)|$. Let, in consequence, q and q' be the uniform distributions over $\Omega(t)$ and $\Omega'(t)$, respectively. Then:

$$S_{\Lambda}(q) = \Lambda(|\Omega(t)|) < \Lambda(|\Omega'(t)|) = S_{\Lambda}(q'),$$

where $S_{\Lambda}(q), S_{\Lambda}(q')$ are the generalized entropies as defined in Equation (5) applied to distributions q and q' . In the Proposition A3 of the Appendix C we provide details of the above derivations. We observe that SK4 is not generally satisfied: This axiom states that $S(AB) = S(A) + S(B|A)$, and one can only guarantee its validity in the case of Shannon entropy, where $\Lambda = \log$. In the general case, this condition may not be satisfied. A different arithmetic rule can substitute SK4 to accommodate other entropic forms [46]. Notice, however, that the use of Shannon (path) entropy, i.e., $\Lambda = \log$, in the compact scale of a CSP may be used in a broad spectrum of cases, including systems with correlations or super-exponential sample space growth, as we will see in Section 2.3.

If the contributions to the above entropy of the paths belonging to the complementary set of $A[\epsilon_t]$, that is, $\Omega(t) \setminus A[\epsilon_t]$, are negligible in the limit of $t \rightarrow \infty$, then we call the CSP *simple*. For simple CSPs with compact scale (Λ, g) the following convergence condition holds:

$$\frac{S_{\Lambda}(\eta(t))}{g(t)} \rightarrow 1. \quad (6)$$

Moreover, in simple CSPs, the typical set has the largest contribution to the entropy. First we define the contribution of the typical set to the entropy, $S_{\Lambda}(A[\epsilon_t])$, as:

$$S_{\Lambda}(A[\epsilon_t]) = \sum_{x(t) \in A[\epsilon_t]} p(x(t)) \Lambda\left(\frac{1}{p(x(t))}\right).$$

We then demonstrate the above claim with the following proposition:

Theorem 2. *If η is a simple CSP with compact scale (Λ, g) , there exists a non-increasing sequence of positive numbers $\epsilon_1, \epsilon_2, \dots, \epsilon_{t-1}, \epsilon_t, \dots$ with limit $\lim_{t \rightarrow \infty} \epsilon_t = 0$ and its corresponding sequence of typical sets $A[\epsilon_1], \dots, A[\epsilon_t], \dots$, such that:*

$$\lim_{t \rightarrow \infty} \frac{S_{\Lambda}(\eta(t))}{g(t)} = \lim_{t \rightarrow \infty} \frac{S_{\Lambda}(A[\epsilon_t])}{g(t)} = 1.$$

Proof. We will start with the second equality, namely:

$$\lim_{t \rightarrow \infty} \frac{S_{\Lambda}(A[\epsilon_t])}{g(t)} = 1.$$

From the definition of typical sets we know that, for paths $x(t) \in A[\epsilon_t]$, it is true that:

$$\frac{1}{\Lambda^{-1}((1 + \epsilon_t)g(t))} \leq p(x(t)) \leq \frac{1}{\Lambda^{-1}((1 - \epsilon_t)g(t))}.$$

In consequence, given that $\mathbb{P}(x(t) \in A[\epsilon_t]) > 1 - \delta_t$, one can bound $S_{\Lambda}(A[\epsilon_t])$ as:

$$(1 - \delta_t)(1 - \epsilon_t) < \frac{S_{\Lambda}(A[\epsilon_t])}{g(t)} < (1 - \delta_t)(1 + \epsilon_t).$$

Since, by construction $\lim_{t \rightarrow \infty} \epsilon_t = \lim_{t \rightarrow \infty} \delta_t = 0$, this second part of the theorem is proven. From that, the statement of the theorem:

$$\lim_{t \rightarrow \infty} \frac{S_{\Lambda}(\eta(t))}{g(t)} = \lim_{t \rightarrow \infty} \frac{S_{\Lambda}(A[\epsilon_t])}{g(t)} = 1,$$

follows directly given the assumption of simplicity. \square

In consequence, the typical set can be naturally defined for simple CSPs in terms of the generalized entropy S_Λ . To see that, we first reword condition (1) for simple CSPs as:

$$\lim_{t \rightarrow \infty} \frac{1}{S_\Lambda(\eta(t))} \Lambda\left(\frac{1}{p(\eta(t))}\right) = 1,$$

(in probability). We can rewrite the above condition in a more convenient form: Given a simple CSP η , there are two non-increasing sequences of positive numbers $\epsilon_1, \dots, \epsilon_t, \dots$, $\delta_1, \dots, \delta_t, \dots$, with $\lim_{t \rightarrow \infty} \epsilon_t = \lim_{t \rightarrow \infty} \delta_t = 0$, by which:

$$\mathbb{P}\left(\left|\frac{1}{S_\Lambda(\eta(t))} \Lambda\left(\frac{1}{p(x(t))}\right) - 1\right| > \epsilon_t\right) < \delta_t. \quad (7)$$

Then, for each $t > 0$ there is a set of paths, the *typical set* $A[\epsilon_t] \subseteq \Omega(t)$, such that for all $x(t) \in A[\epsilon_t]$:

$$\begin{aligned} \frac{1}{\Lambda^{-1}((1 + \epsilon_t)S_\Lambda(\eta(t)))} &\leq p(x(t)) \\ &\leq \frac{1}{\Lambda^{-1}((1 - \epsilon_t)S_\Lambda(\eta(t)))}, \end{aligned}$$

by which $\mathbb{P}(x(t) \in A[\epsilon_t]) > 1 - \delta_t$. Notice that, now, the typical set is characterized using the generalized entropy S_Λ .

The next obvious question refers to the cardinality of the typical set $|A[\epsilon_t]|$. We will see that it can be bounded from above and below in a way analogous to the standard one [24]. The first bound is obtained by observing that:

$$\begin{aligned} 1 - \delta_t &\leq \sum_{x(t) \in A[\epsilon_t]} p(x(t)) \\ &\leq \frac{|A[\epsilon_t]|}{\Lambda^{-1}((1 - \epsilon_t)S_\Lambda(\eta(t)))}, \end{aligned}$$

where Λ^{-1} is the inverse function of Λ , i.e., $(\Lambda^{-1} \circ \Lambda)(z) = z$, which exists given the assumption that Λ is a monotonically growing function made in the definition of CSPs. From that, it follows that the cardinality of the typical set is bounded from below as:

$$|A[\epsilon_t]| \geq (1 - \delta_t) \Lambda^{-1}((1 - \epsilon_t)S_\Lambda(\eta(t))). \quad (8)$$

For the upper bound, we observe that:

$$\begin{aligned} 1 &\geq \sum_{x(t) \in A[\epsilon_t]} p(x(t)) \\ &\geq \frac{|A[\epsilon_t]|}{\Lambda^{-1}((1 + \epsilon_t)S_\Lambda(\eta(t)))}, \end{aligned}$$

leading to:

$$|A[\epsilon_t]| \leq \Lambda^{-1}((1 + \epsilon_t)S_\Lambda(\eta(t))). \quad (9)$$

Given the bounds provided in Equations (8) and (9), one can (roughly) estimate the cardinality of the typical set as:

$$|A[\epsilon_t]| \approx \Lambda^{-1}(S_\Lambda(\eta(t))). \quad (10)$$

We present a rigorous version of the above result as a proposition.

Proposition 1. Let η be a simple CSP in (g, Λ) with some typical localizer sequence $\epsilon_1, \dots, \epsilon_t, \dots$, then:

$$\lim_{t \rightarrow \infty} \frac{\Lambda(|A[\epsilon_t]|)}{S_\Lambda(\eta(t))} = 1.$$

Proof. From Equations (8) and (9), one can derive the following chain of inequalities:

$$\begin{aligned}\Lambda((1 - \epsilon_t)\Lambda^{-1}((1 - \epsilon_t)S_\Lambda(\eta(t)))) &< \Lambda(|A[\epsilon_t]|) \\ &< (1 + \epsilon_t)S_\Lambda(\eta(t)).\end{aligned}$$

The last term poses no difficulties. To explore the behavior of the first one, we just rename the term:

$$z \equiv \Lambda^{-1}((1 - \epsilon_t)S_\Lambda(\eta(t))),$$

and rewrite the first term of the inequality:

$$\Lambda((1 - \epsilon_t)\Lambda^{-1}((1 - \epsilon_t)S_\Lambda(\eta(t)))) = \Lambda((1 - \epsilon_t)z).$$

We know, from Proposition A1, that the functions we are dealing with behave such that:

$$\lim_{z \rightarrow \infty} \frac{\Lambda(\lambda z)}{\Lambda(z)} \rightarrow 1, \quad (\forall z > 0).$$

As a consequence:

$$\frac{\Lambda((1 - \epsilon_t)\Lambda^{-1}((1 - \epsilon_t)S_\Lambda(\eta(t))))}{S_\Lambda(\eta(t))} \rightarrow 1.$$

Therefore, since also the third term goes trivially to $\sim S_\Lambda(\eta(t))$, we can conclude that:

$$\frac{\Lambda(|A[\epsilon_t]|)}{S_\Lambda(\eta(t))} \rightarrow 1,$$

as we wanted to demonstrate. \square

The above proven asymptotic equivalence gives us the opportunity of rewriting the entropy in a *Boltzmann-like* form:

$$S_\Lambda(\eta(t)) \sim \Lambda(|A[\epsilon_t]|).$$

This identifies the cardinality of the typical set with Boltzmann's W , the number of alternatives the system can effectively display: Finally, we notice that we can (roughly) approximate the typical probabilities as:

$$p(x(t)) \approx \frac{1}{\Lambda^{-1}(S_\Lambda(\eta(t)))}.$$

We thus provided a general proof that the typical set exists and that it can be properly defined for a wide class of stochastic processes, the CSPs, those satisfying convergence condition (1). Moreover, we show that its volume can be bounded and fairly approximated as a function of the generalized entropy emerging from the convergence condition, S_Λ , as defined in Equation (5).

2.3. Example: A Path Dependent Process

We briefly explore the behavior of the typical set and its associated entropic forms through a model displaying both path dependence and unbounded growth of the phase space. The process η works as follows: Let us suppose we have a restaurant with an infinite number of tables m_1, \dots, m_n, \dots . At $t_0 = 0$ a customer enters the restaurant and sits at table m_1 . At time t a new customer enters the restaurant where already $m(t)$ tables are occupied; the occupation number of each table is unbounded. The customer can choose either to sit at an already occupied table from the $m_1, \dots, m_{m(t)}$ occupied tables, each with equal probability $\frac{1}{m(t)+1}$, or in the next unoccupied one, $m_{m(t)+1}$, again with probability $\frac{1}{m(t)+1}$. This process is a version of the so-called *Chinese restaurant process* [33,63], exhibiting

a simple form of memory/path dependence. Hence, we refer to it as the Chinese restaurant process with memory (CRPM). In Figure 2 we sketch the rules of this process. Crucially, as $t \rightarrow \infty$, the random variable accounting for the number of tables $m(t)$ has the following convergent behavior; see Proposition A5 of the Appendix D.2 for details:

$$\frac{m(t)}{\sqrt{2t}} \rightarrow 1.$$

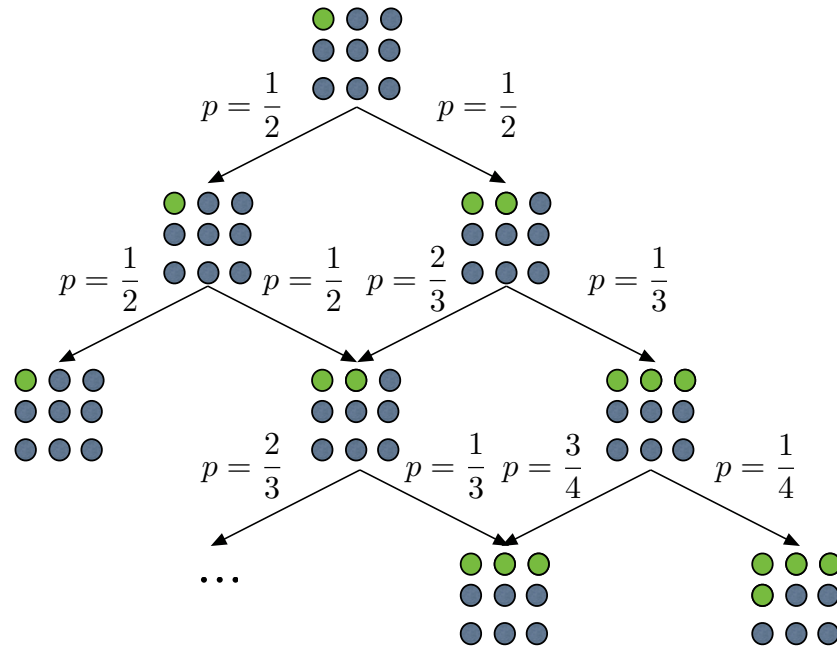


Figure 2. The rules of the Chinese restaurant process with memory. Here green circles represent occupied tables and grey circles empty table. Notice that in the mathematical formulation of the problem the number of tables is infinite. Arrows depict the possible transitions of the process and the associated probabilities.

In Figure 3a we see that the prediction $m(t) \sim \sqrt{2t}$ is quite accurate when compared to numerical simulations of the process. This property enables us to demonstrate that the CRPM we are studying is actually a CSP with compact scale $(\log, \frac{t}{2} \log t)$; see Theorem A1 of the Appendix D.3. In particular, Equation (1) is satisfied, in this particular case as:

$$\lim_{t \rightarrow \infty} \frac{1}{\frac{t}{2} \log t} \log \left(\frac{1}{p(\eta(t))} \right) = 1,$$

in probability. In addition, the process is *simple*; see Theorem A2. Since we are using $\Lambda = \log$, the entropy form that will arise is Shannon path entropy, by direct application of Equation (5), i.e., $S_\Lambda(\eta(t)) = H(\eta(t))$, with $H(\eta(t))$ defined as:

$$H(\eta(t)) = - \sum_{x(t) \in \Omega(t)} p(x(t)) \log p(x(t)). \quad (11)$$

It directly follows that:

$$\frac{H(\eta(t))}{\frac{t}{2} \log t} \rightarrow 1. \quad (12)$$

Given the compact scale used, one can estimate the evolution of the size of the typical set as:

$$|A[\epsilon_t]| \sim \sqrt{\Gamma(t)}, \quad (13)$$

where Γ is the standard Γ -function [64]. We see that the growth of the typical set as shown in Equation (13) is clearly faster than exponential. The phenomenon of *concentration of measure* [29] is clearly manifested here, as the size of the typical set vanishes in relation to the size of the whole potential set of outcomes, $\Omega(t)$. A rough estimation leads to $|\Omega(t)| \sim \Gamma(t)$, leading to:

$$\frac{|A[\epsilon_t]|}{|\Omega(t)|} \sim \frac{1}{\sqrt{\Gamma(t)}} \rightarrow 0,$$

despite $\mathbb{P}(x(t) \in A[\epsilon_t]) \rightarrow 1$.

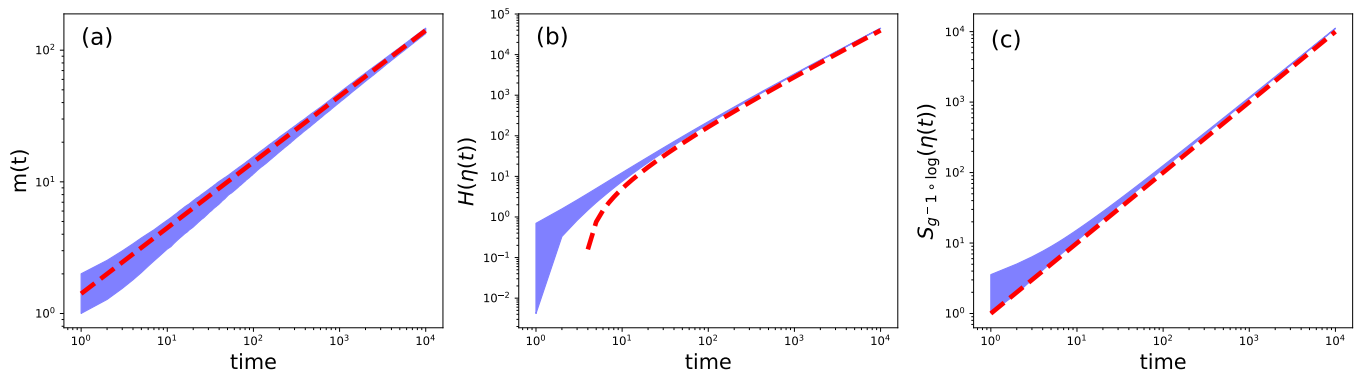


Figure 3. Numerical simulations for the *Chinese Restaurant process with memory*. The blue cloud represents actual numerical outcomes, dashed orange line the theoretical prediction. Time is given in arbitrary coordinates, representing a step in the process. In (a) we show the evolution of the amount of occupied tables against the prediction $m(t) \sim \sqrt{2t}$. (b) The evolution of Shannon path entropy for the CRPM, being the prediction given in (11). The dashed red line shows the function $g(t) \sim \frac{t}{2} \log t$. (c) Evolution of the generalized path entropy S_Λ , with Λ as defined in Equation (14). Numerical outcomes have been obtained from 1000 replicas of the whole CRPM process up to $t = 10^4$ steps.

Additionally, in Figure 3b we see that the prediction made in Equation (12) fits perfectly with the numerical realizations of the process. Note that we have shown the dependence on Shannon path entropy for the clarity in the exposition. Indeed, as pointed out above, a CSP η may have several compact scales. For example, taking the compact scale that led to Shannon entropy, $(\log, g(t))$, with $g(t) = \frac{t}{2} \log t$, one can construct another compact scale for the CRPM by composing g^{-1} (which, by assumption, exists) with both functions. In consequence, one will have a new compact scale (Λ, \tilde{g}) , defined as:

$$\Lambda(t) = (g^{-1} \circ \log)(t) \sim \frac{2 \log(t)}{W(2 \log(t))}, \quad \tilde{g}(t) = t, \quad (14)$$

where W is the positive, real branch of the *Lambert function* [64]. In Figure 3c we see that $S_\Lambda(\eta(t))$ fits perfectly $g(t) \sim t$, proving that (Λ, t) is a compact scale for the CRPM; see also Appendix D.5. We observe that this particular compact scale makes the path entropy S_Λ *extensive* when applied to the CRPM.

3. Discussion

We demonstrated that, for a very general class of stochastic processes, which we refer to as *compact stochastic processes*, the typical set is well defined, as the probability measure tends to concentrate in clearly identifiable regions of the space of possible outcomes of the process. These processes can be path dependent, contain arbitrary internal correlations or display dynamic behavior of the phase space, showing sub- or super-exponential growth on the effective number of configurations the system can achieve. The only requirement is that there exist two functions Λ, g for which Equation (1) holds. Along the existence of the typical set, a generalized form of entropy naturally arises, from which, in turn, the cardinality of the typical set can be computed.

The existence of the typical set in systems with arbitrary phase space growth opens the door to a proper characterization, in terms of statistical mechanics, of a number of processes, mainly biological, where the number of configurations and states changes over time. In particular, it paves the path towards the statistical-mechanics-like understanding of processes showing open-ended evolution on the basis of typicality. For example, this could encompass thermodynamic characterizations of—part of—the developmental paths in early stages of embryogenesis. The existence of the typical set, even in some extreme scenarios of stochasticity and phase space behavior, leads us to the speculative hypothesis that typicality may underlie the astonishing reproducibility and precision of some biological processes. In this scenario, stochasticity would drive the system to the set of *correct* configurations—those belonging to the typical set—with high accuracy. Selection, in turn, would operate on typical sets, thereby promoting certain stochastic processes over others. More specific scenarios are nevertheless required in order to make this intriguing hypothesis more sound. Further works should clarify the potential of the proposed probabilistic framework to accommodate generalized, consistent forms of thermodynamics and explore the complications that can arise due to the loss of ergodicity that is characteristic for some of the processes that are nonetheless compatible with the above description.

Importantly, our results provide a potential starting point for an ensemble formalism for systems with sub- or super-exponential phase space growth. This opens the possibility to extend the concept of thermodynamic limit to these systems without requiring further conditions such as microscopic detailed balance, which cannot be justified in a broad range of out-of-equilibrium processes. Questions like the definition of free energies or the possible need of extensivity, which have to be answered in order to progress towards a complete and consistent thermodynamic picture remain, however, open. For tentative answers to those questions, connections to early proposals could be drawn, both on the thermodynamic grounds; see, e.g., [36,65,66], and from the perspective of entropy characterization; see, e.g., [37,44,46,53,54]. In this paper we meet an equivalence relation underlying compact scales, which allows us to transform between compact scales that do not differ too strongly, i.e., not more than by a power, without essentially changing the structure of the typical set sequence, a fact that one can utilize to give the entropic functional particular properties. For instance making the entropy associated with the compact scale extensive, as we demonstrated above for the CRPM. However, this also introduces the possibility that processes may have more than one inequivalent compact scale. We are aware that, in order to demonstrate the full potential of the theory one should aim at examples more extreme than the Chinese restaurant process we present. However, this would require an additional inquiry into the potentially hierarchic structure of potentially inequivalent compact scales that both in technical terms and conceptual terms go beyond the scope of this paper. This intriguing issue may be related to different levels of coarse graining, and deserves further investigations.

We finally point out the implications of our results for the study of information sources, given the fundamental role the typical set plays in optimal coding and data compression. The existence of the typical set in these broad class of information sources, where, simply put, the information flow is not constant, may open the possibility of new compressing strategies. These strategies could be based, for example, on properties of the specific CSP representing the information source and the compact scale Λ, g that characterize the process.

Author Contributions: Conceptualization, R.H. and B.C.-M.; Formal analysis, R.H. and B.C.-M.; Writing—original draft, R.H. and B.C.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors want to thank Petr Jizba and Artemy Kolchinsky for the helpful discussions that enabled us to improve the quality of the manuscript. B.C.-M. wants to acknowledge the hints provided by Daniel R. Amor and the support of the field of excellence *Complexity of Life, Basic Research and Innovation* of the University of Graz.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In this appendix we provide the mathematical details required to completely justify the results contained in the main text of the manuscript “The typical set and entropy in stochastic systems with arbitrary phase space growth”.

Appendix B. Compact Categorical Processes

We start by providing a general framework for the processes considered in the paper. In Figure A1, below, we outline the hierarchy that our study induces over stochastic processes.

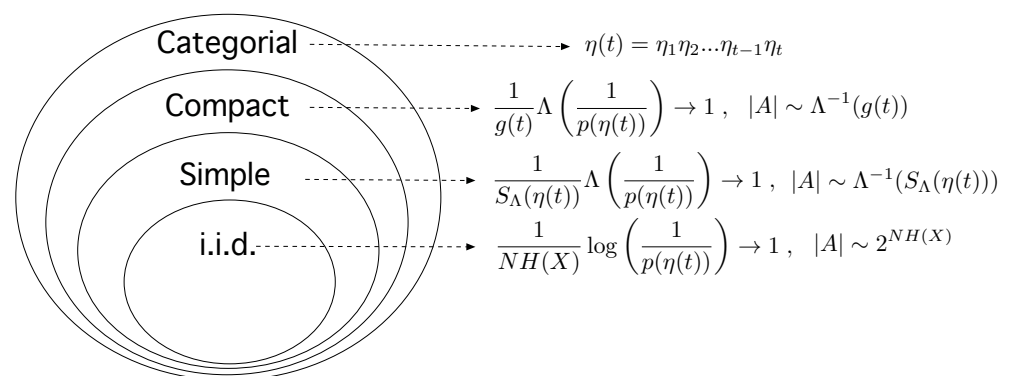


Figure A1. A potential hierarchy of discrete stochastic processes. The largest class would correspond to the *categorical* processes, which comprise almost anything that can be conceived as a discrete stochastic process. A subclass of categorical processes are the *compact* processes, by which the convergence condition stated in the Equation (1) of the main text holds and, therefore, a sequence of typical sets can be identified. Inside the compact processes, we identify the subclass of *simple* processes, where the sequence of typical sets can be defined in terms of a general form of entropy. Finally, the simplest subclass is the one defined by stochastic processes defined by sequences of independent, identically distributed random variables $\sim X$, by which the sequence of typical sets can be defined from the entropy in Shannon-like form. In turn, in i.i.d. systems, the path entropy up to time t can be written as t times the contribution of a single event [24]. Note that no assumptions of independence or stability of the sampling space are needed in the two first subclasses. Even the typical set can be consistently identified. In addition, as we will see in Appendix D, the use of Shannon entropy to characterize the sequence of typical sets is not restricted to i.i.d. systems.

Appendix B.1. Categorical Processes

Categorical processes η are processes that at any time t sample one state from a finite number of distinguishable states collected in the set Ω_t , called the *sample space* of the process at time t . If we look at a discrete time line $T = 1, 2, 3, \dots$ we represent the process η up to time t as a sequence of random variables η_t , i.e.,

$$\eta(t) = \eta_1\eta_2 \cdots \eta_{t-1}\eta_t.$$

The process η neither needs to consist of statistically independent random variables η_t nor does the sample space of the process need to be constant. That is, the local sample spaces

Ω_t of the variable η_t can differ from the sample space $\Omega_{t'}$ of another variable $\eta_{t'}$. If we sample $\eta(t)$ this provides us with a particular path:

$$x(t) = x_1 x_2 \cdots x_{t-1} x_t,$$

with $x(t) \in \Omega(t)$, being $\Omega(t)$ defined as:

$$\Omega(t) = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_{t-1} \times \Omega_t, \quad (\text{A1})$$

in other words, the Cartesian product of all local sample spaces $\Omega_{t'}$ up to time t .

In principle it could be that also the sample space Ω_t depends on the path $x(t-1)$ the process has taken up to time $t-1$. In consequence, $\Omega(t)$ could contain many paths that are not possible for the process η and, therefore, have zero probability of being sampled. However, in such a case we can always consider the non-empty subset $\mathring{\Omega}(t) \subseteq \Omega(t)$, which contains all paths $x(t)$, and only those paths of η , with $p(x(t)) > 0$.

Definition A1. Let η be a categorical process. We call $p(x(t))$ the probability of paths $x(t) \in \Omega(t)$ to be sampled by η and, then, we call:

$$\mathring{\Omega}(t) = \{x(t) \in \Omega(t) : p(x(t)) > 0\}, \quad (\text{A2})$$

the *well formed interior* of $\Omega(t)$ or the *path sample space* of the process.

Let us call $\Omega_t[x(t-1)]$ the set of states of Ω_t that can be sampled at time t provided that our trajectory up to time $t-1$ was $x(t-1)$. The set of potential states that can be visited at time t , will then be:

$$\Omega_t = \bigcup_{x(t-1) \in \mathring{\Omega}(t-1)} \Omega_t[x(t-1)],$$

i.e., Ω_t contains all states the process could possibly sample at time t after all possible histories the process could have sampled up to time $t-1$. In this way we can always assume that we can find $\Omega(t)$ and its well formed interior $\mathring{\Omega}(t)$ by pruning all ill formed sequences from $\Omega_t \times \mathring{\Omega}(t-1)$. All information on how the sample space Ω_t gets sampled then solely resides in the hierarchy of transition probabilities $p(x_t|x(t-1))$, where $x_t \in \Omega_t$ and $x(t-1) \in \Omega(t-1)$. We therefore get:

$$\mathring{\Omega}(t) = \{x(t) \in \Omega_t \times \mathring{\Omega}(t-1) : p(x_t|x(t-1)) > 0\}.$$

In the context of categorical processes, $p(x(t))$ is clearly a monotonic decreasing function in time bounded below by zero, $p(x(t)) \geq 0$. In consequence, $\lim_{t \rightarrow \infty} p(x(t))$ converges for all possible paths $x(t)$ of the process η . This also means that for almost all paths the probability $\lim_{t \rightarrow \infty} p(x(t)) = 0$ even though some finite number of paths could have non-zero probabilities even in the limit $t \rightarrow \infty$, although convergence is guaranteed (Unlike for processes, for *systems*, e.g., particles in a box, it is not guaranteed that adding a new particle, as analog of to a new sample step, $p(x(t)) \leq p(x(t-1))$. In consequence, for systems, we cannot guarantee convergence of $p(x(t))$ as the system size $t \rightarrow \infty$). After this general description of categorical processes, we can start by characterizing the subclass of them we are interested in. In the following we provide the condition of compactness that we impose to categorical processes in order to ensure the existence of a typical set.

Definition A2. Let η be a categorical process η and let $\eta(t)$ denote the process up to time t and $\Omega(t)$ denote the path sample space of $\eta(t)$ (as discussed above). Let us consider pairs of functions (Λ, g) , such that Λ is a twice continuously differentiable, strictly monotonic increasing and strictly concave function on the interval $[1, \infty)$ with $\lim_{t \rightarrow \infty} \Lambda(t) = \infty$ and $\Lambda(1) = 0$; and g is a twice

continuously differentiable strictly monotonically increasing function on the interval $[0, \infty)$. If we can **associate** such a pair of functions (Λ, g) with the process η , such that:

$$\lim_{t \rightarrow \infty} \frac{1}{g(t)} \Lambda\left(\frac{1}{p(\eta(t))}\right) = 1,$$

(in probability), then we call the process η **compact** in (Λ, g) , and (Λ, g) a **compact scale** of η .

We will call these processes *Compact Stochastic Processes* (CSP). Note that compact scales (Λ, g) associated to a given CSP need not be unique (The fact that we can find different pairs of functions, (Λ, g) , in which a process η is compact, leads to questions related to equivalence relations on the space of pairs, (Λ, g) . As it turns out, following up the idea of typical sets, that we are to explore below, a process η induces an equivalence relation on this space, partitioning the space into monads of equivalent compact scales, (Λ, g) , of the process. At the same time this means that there exist inequivalent compact scales, which can be thought of as different "scales of resolution" to look at a process.).

Appendix B.2. Properties of Λ

The next step is to check what kind of functions Λ can be expected when dealing with CSPs. To that end, we will impose a condition to the CSP, namely, that the CSP is *filling*. From this, very mild, condition, we will then check which functions enable the convergence criteria to be fulfilled. First of all, we need to introduce some technical terms.

Definition A3. Let η be a CSP in (Λ, g) . Let $\epsilon_1, \dots, \epsilon_t, \dots$ be a non-increasing sequence of real numbers such that $\lim_{t \rightarrow \infty} \epsilon_t = 0$, from which one can construct a sequence of typical sets $A[\epsilon_1] \dots A[\epsilon_t]$ of the respective sample spaces $\Omega(1) \dots \Omega(t)$ such that $\mathbb{P}(x(t) \in A[\epsilon_t]) \rightarrow 1$. We define the upper and lower **typical ratios**, $r_{\pm}(t)$, in (Λ, g) as:

$$r_{\pm}(t) = \frac{\Lambda^{-1}((1 \pm \epsilon_t)g(t))}{\Lambda^{-1}(g(t))},$$

where Λ^{-1} is the inverse function of Λ , which exists due to the strict monotonicity of Λ .

Note that, by construction, $r_{+}(t) \geq 1$ and $r_{-}(t) \leq 1$.

Definition A4. We call a CSP η **filling** in (Λ, g) if its typical ratios have the property $\lim_{t \rightarrow \infty} r_{+}(t) = \infty$ and $\lim_{t \rightarrow \infty} r_{-}(t) = 0$.

We can now say something about the shape of functions Λ .

Proposition A1. If η is a filling CSP in (Λ, g) , then it follows that:

$$\lim_{z \rightarrow \infty} \frac{\Lambda(\lambda z)}{\Lambda(z)} = 1,$$

for all $\lambda > 0$.

Proof. We note that for filling CSP η it is true that:

$$\frac{1}{g(t)} \Lambda(r_{\pm}(t) \Lambda^{-1}(g(t))) \rightarrow 1,$$

as $t \rightarrow \infty$. We can rewrite $z_t = \Lambda^{-1}(g(t))$ and for $\lambda > 1$ we find a t_0 such that $\lambda = r_+(t_0)$. Therefore for all $t > t_0$ we find:

$$1 \leq \frac{\Lambda(\lambda z_t)}{\Lambda(z_t)} \leq \frac{\Lambda(r_+(t)z_t)}{\Lambda(z_t)} \rightarrow 1.$$

The other case, $0 < \lambda < 1$, we prove analogously using r_- instead of r_+ , and the proposition follows. \square

To get an idea which kind of functions satisfy this condition we can look at the following example:

Proposition A2. For any $c > 0$ the function $\Lambda(z) = \log(z)^c$ has the property $\lim_{z \rightarrow \infty} \Lambda(\lambda z) / \Lambda(z) = 1$ for all $\lambda > 0$.

Proof. We can prove this by direct computation, i.e.,

$$\begin{aligned} \frac{\Lambda(\lambda z)}{\Lambda(z)} &= \left(\frac{\log(\lambda z)}{\log(z)} \right)^c \\ &= \left(\frac{\log(\lambda) + \log(z)}{\log(z)} \right)^c \\ &= \left(1 + \frac{\log(\lambda)}{\log(z)} \right)^c. \end{aligned}$$

Since $\log(z) \rightarrow \infty$, one can easily read from the last line that the example family of functions Λ fulfills the proposition. \square

In general we can say that candidates for Λ of filling CSPs are of the form $\Lambda(z) = c \log(z)^d$ for some positive constants c and d or even slower growing functions of the form:

$$\Lambda = c_1 \log(1 + c_2 \log(1 + c_3 \log(\dots)^{d_3})^{d_2})^{d_1}.$$

where c_1, \dots and d_1, \dots are positive, real valued constants.

Appendix C. Properties of the Generalized Entropies

The generalized entropy associated to a CSP η in (Λ, g) is defined as:

Definition A5. Let η be a CSP in (Λ, g) , then we call the measure, S_Λ of $\eta(t)$, defined as:

$$S_\Lambda(\eta(t)) = \sum_{x(t) \in \Omega(t)} p(x(t)) \Lambda\left(\frac{1}{p(x(t))}\right), \quad (\text{A3})$$

a **generalized path entropy** associated with η . Note that for a set $B \subset \Omega(t)$ the generalized entropy measure, $S_\Lambda(B)$, is given by $S_\Lambda(B) = \sum_{x(t) \in B} p(x(t)) \Lambda(1/p(x(t)))$.

In the following proposition we see that the above defined entropy satisfies three of the four Shannon–Khinchin axioms for an entropy measure [62] (**SK1**, **SK2**, **SK3**). The fourth axiom (**SK4**) is not generally satisfied.

Proposition A3. The entropy functional S_Λ defined in Equation (A3) satisfies the first three of the four Shannon–Khinchin axioms for an entropy measure as formulated in [62]:

SK1 S_Λ is a continuous function only depending on the probabilities $p(x(t))$.

SK2 S_Λ is maximized if $(\forall p(x(t))) p(x(t)) = \frac{1}{|\Omega(t)|}$, i.e., equiprobability.

SK3 If $p(x(t)) = 0$, then: $p(x(t)) \Lambda\left(\frac{1}{p(x(t))}\right) = 0$, i.e., events with zero probability have no contribution to the entropy.

Proof. To demonstrate **SK1**, it is enough to observe that S_Λ is only function of the probabilities and to take into account that, by assumption, $\Lambda \in \mathcal{C}^2$, therefore, S_Λ continuous. To demonstrate **SK2**, we need to maximize the functional ψ , defined as:

$$\psi = S_\Lambda(\eta(t)) - \alpha \left(\sum_{x(t) \in \Omega(t)} p(x(t)) - 1 \right),$$

where α is a Lagrangian multiplier implementing the normalization constraint. Maximizing ψ with respect to a $p(x(t))$ yields:

$$0 = \frac{\partial \psi}{\partial p(x(t))} = \Lambda(z) - z\Lambda'(z) - \alpha,$$

where $z = 1/p(x(t))$ and Λ' is the first derivative of Λ . Note that if the equation does not depend explicitly on $x(t)$ and if it has a unique solution then the proposition is proved, since all $p(x(t))$ have the same value. To see that a unique solution exists we need to show that $f(z) = \Lambda(z) - z\Lambda'(z)$ is strictly monotonic. To see that, it is enough to note that the first derivative of f is given by $f'(z) = -z\Lambda''(z) > 0$, since, by definition of compactness, $\Lambda \in \mathcal{C}^2$ and strictly concave; and therefore $\Lambda''(z) < 0$.

Finally, to demonstrate that S_Λ satisfies **SK3** we apply the l'Hopital rule. First, by defining $y = 1/z$, one has that:

$$\lim_{z \rightarrow 0} z\Lambda\left(\frac{1}{z}\right) = \lim_{y \rightarrow \infty} \frac{1}{y}\Lambda(y).$$

Then, considering that, by definition, $\Lambda \in \mathcal{C}^2$ is a strictly growing and concave function, one is led, after application to the l'Hopital rule for the limit, to:

$$\lim_{y \rightarrow \infty} \frac{1}{y}\Lambda(y) = \lim_{y \rightarrow \infty} \Lambda'(y) = 0,$$

thereby concluding the proof. \square

We observe that **SK3** enables us to safely perform the sum for the entropy over the whole set of paths $\Omega(t)$, since $\sum_{x(t) \in \hat{\Omega}(t)}(\dots) = \sum_{x(t) \in \Omega(t)}(\dots)$.

Appendix D. The Chinese Restaurant Process

We now turn to analyzing the version of the Chinese restaurant process with memory (CRPM) discussed in the main body of the paper. The version presented here is a variation of the standard Chinese Restaurant process as found in [33,63].

Appendix D.1. Definition and Basics

Suppose a restaurant with an infinite set of tables m_1, \dots, m_n, \dots each with infinite capacity. The first customer enters and sits at the first table m_1 . The second customer now has a choice to also sit down at the first table m_1 together with the first customer or to choose a free table m_2 , each with probability $1/2$. Let $m(t-1)$ be the number of occupied tables at $t-1$. If the t 'th customer finds that $m(t-1)$ tables are already occupied by some guests, then again the customer will choose one of the non-empty tables:

$$m_1, \dots, m_{m(t-1)},$$

each with probability $1/(m(t-1) + 1)$, in which case $m(t) = m(t-1)$, or the next empty table $m_{m(t-1)+1}$, also with probability $1/(m(t-1) + 1)$. In this latter case $m(t) = m(t-1) + 1$. The key point is therefore the number of occupied tables $m(t)$. We observe that the amount of occupied tables can be rewritten as a stochastic recurrence:

$$m(t+1) = m(t) + \zeta(m(t)), \quad (\text{A4})$$

where $\zeta(m(t))$ is a random variable by which:

$$p(\zeta(m(t)) = 0) = \frac{m(t)}{m(t+1)}, \quad p(\zeta(m(t)) = 1) = \frac{1}{m(t+1)}.$$

Clearly,

$$m(t) = 1 + \sum_{t' \leq t} \zeta(m(t')),$$

is a non-decreasing function in t . Now let us define $M_k(t)$ as a random variable taking values uniformly at random over the set m_1, \dots, m_k, m_{k+1} at time t . The sequence of random variables accounting describing the CRPM $\eta(t)$ can be written as:

$$\eta(t) = M_1(1), M_1(2), M_{m(2)}(3), \dots, M_{m(t)}(t+1), \dots,$$

leading to paths $x(t)$ of the kind:

$$x(t) = m_1, m_1, m_2, m_1, m_2, m_3, m_2, m_2, m_1, m_3, \dots$$

We emphasize that *the tables visited are distinguishable and can be visited repeatedly*. The CRPM however does not fill up $\Omega(t)$, i.e., there exist elements in $\Omega(t) = \times_{t'=1}^t \Omega_{t'}$, that are not potential paths of the CRP. This includes all sequences that select a table m_i without ever having chosen some table m_j , with $j < i$ before. For example, the path $x(t) = (m_1, m_2, m_1, m_2, m_3, m_1, m_5, m_4, m_3, \dots)$ is not possible, because m_5 is chosen before m_4 . The CRPM therefore gives us the opportunity to introduce sampling spaces conditional to a particular well formed path. In this particular case, it is enough to observe that the sampling space some well formed path $x(t) \in \Omega(t-1)$ sees at time t is given by:

$$\Omega_t[x(t-1)] = \{m_1, m_2, \dots, m_{m[x](t-1)+1}\}, \quad (\text{A5})$$

where $m[x](t-1)$ is the number of different tables the well formed CRPM path $x(t-1)$ has sampled at time $t-1$. By convention, we define $m[x](1) = 1$.

Appendix D.2. Statistics of the CRPM

We start computing the probability of a particular path $x(t) \in \hat{\Omega}(t)$.

Proposition A4. *Let process $\eta(t)$ be the CRPM and $x(t) \in \hat{\Omega}(t)$ be a given path of the process up to time t . Let $m[x](t')$ be the number of occupied tables in the restaurant associated with the path $x(t')$ at time t' . Then the probability to observe the particular sequence of tables, $p(x(t))$ is given by:*

$$p(x(t)) = \prod_{t'=1}^t (m[x](t') + 1)^{-1}.$$

Proof. The proposition follows from direct calculation. \square

Now we will see that the sequence corresponding to the number of occupied tables converges to a tractable functional form.

Proposition A5. *Given the sequence of occupied tables of the CRPM $m(1), \dots, m(t)$ as defined in Equation (A4), then:*

$$\frac{m(t)}{\sqrt{2t}} \rightarrow 1,$$

in probability.

Proof. Consider the random variable Δt_k denoting the amount of steps by which $m(t) = k$. Δt_k is a geometric random variable with associated law:

$$p(\Delta t_k = i) = \left(1 - \frac{1}{k}\right)^{i-1} \frac{1}{k}, \quad \langle \Delta t_k \rangle = k, \quad \sigma^2(\Delta t_k) = k^2 - k.$$

Now we construct a new set of renormalized random variables, $\delta t_1, \dots, \delta t_k$ as:

$$\delta t_k \equiv \frac{\Delta t_k}{k}, \quad \langle \delta t_k \rangle = 1, \quad \sigma^2(\delta t_k) = 1 - \frac{1}{k}.$$

In that context, the sum of $\delta t_1, \dots, \delta t_k$ is the sum of k random variables with mean 1 and $\sigma^2 < 1$. Therefore, there exists a monotonically increasing function, $\mu(t)$ by which, by the law of large numbers, for each pair $\epsilon, \delta < 0$, there exists t_0 such that, for $t > t_0$:

$$\mathbb{P}\left(\left|\frac{\sum_{k \leq m(t)} \delta t_k}{\mu(t)} - 1\right| > \epsilon\right) < \delta.$$

Notice that, in this setting, we have that the deviations behave close to a discrete random walk centered at 0 and with step length $\sigma^2 < 1$. Since for all δt_k , $\langle \delta t_k \rangle = 1$:

$$\frac{m(t)}{\mu(t)} \rightarrow 1, \quad (\text{in probability}),$$

and, since, by construction:

$$\sum_{k \leq \mu(t)} k \delta t_k = t,$$

one has that:

$$\mu(t) \sim \sqrt{2t}.$$

where “ \sim ” means asymptotically equivalent, as we wanted to demonstrate. \square

Appendix D.3. The Typical Set of the CRP

Theorem A1. The CRPM is compact with compact scale $(\Lambda, g(t)) = (\log, \frac{t}{2} \log t)$.

Proof. We need to demonstrate that:

$$\frac{1}{\frac{t}{2} \log t} \log\left(\frac{1}{p(\eta(t))}\right) \rightarrow 1, \quad (\text{in probability}).$$

We first note that, according to the definition of the sequence of the number of occupied tables given in Equation (A4), and the statement of Proposition A4, one can rewrite the logarithmic term of the condition for compactness as:

$$\log\left(\frac{1}{p(\eta(t))}\right) = \sum_{t' \leq t} \log(m(t')).$$

Now, let us define a new random variable $z(t)$ as follows:

$$z(t) = m(t) - \mu(t),$$

with $\mu(t) = \sqrt{2t}$. Notice that, according to Proposition A5, we have that:

$$\frac{z(t)}{\mu(t)} \rightarrow 0, \quad (\text{in probability}).$$

We can then rewrite condition of compactness, with $g(t) = \sum_{t' \leq t} \log(\mu(t'))$, as:

$$\begin{aligned} \frac{1}{\sum_{t' \leq t} \log(\mu(t'))} \log\left(\frac{1}{p(\eta(t))}\right) &= \frac{1}{\sum_{t' \leq t} \log(\mu(t'))} \sum_{t' \leq t} \log(m(t')) \\ &= \frac{1}{\sum_{t' \leq t} \log(\mu(t'))} \sum_{t' \leq t} \log(\mu(t') + z(t')) \\ &= \frac{1}{\sum_{t' \leq t} \log(\mu(t'))} \sum_{t' \leq t} \log(\mu(t')) + \log\left(1 + \frac{z(t')}{\mu(t')}\right) \\ &= 1 + \frac{1}{\sum_{t' \leq t} \log(\mu(t'))} \sum_{t' \leq t} \log\left(1 + \frac{z(t')}{\mu(t')}\right). \end{aligned}$$

It remains to see that the second term of the sum goes to 0. Clearly, by Proposition A5:

$$\frac{\log\left(1 + \frac{z(t)}{\mu(t)}\right)}{\log(\mu(t))} \rightarrow 0, \quad (\text{in probability}).$$

Consequently:

$$\frac{1}{\sum_{t' \leq t} \log(\mu(t'))} \log\left(\frac{1}{p(\eta(t))}\right) \rightarrow 1, \quad (\text{in probability}).$$

Finally, we need to compute the asymptotic form of $\sum_{t' \leq t} \log(\mu(t'))$. Observing that we have a Riemann sum, one can consider:

$$\sum_{t' \leq t} \log(\mu(t')) \sim \int_0^t \log(\sqrt{2t'}) dt' \sim \frac{t}{2} \log t,$$

thus concluding the proof. \square

Appendix D.4. The Entropy of the CRP

We demonstrate here that the CRPM is simple. In consequence, the typical set can be computed as a function of the entropy. In that case, one can show that a suitable choice is $\Lambda = \log$, chosen for the sake of simplicity, implying that the associated entropy is Shannon path entropy. However, we emphasize that this choice is not unique. Given a different choice of g , one could have another Λ by which the process is also simple and, therefore, the typical set could be defined through another form of entropy. We briefly comment on this point in the next section, sketching how another potential pair (Λ, g) functions would work as well.

Theorem A2. *The CRPM is simple in $(\Lambda, g(t)) = (\log, \frac{t}{2} \log t)$.*

Proof. Since the CRPM is compact with compact scale $(\Lambda, g(t)) = (\log, \frac{t}{2} \log t)$, we know that there is a typical localizer sequence $\epsilon_1, \dots, \epsilon_t, \dots$ with associated $\delta_1, \dots, \delta_t, \dots$, such that $\lim_{t \rightarrow \infty} \epsilon_t = \lim_{t \rightarrow \infty} \delta_t = 0$. The paths belonging to the typical set $A[\epsilon_t]$, are those satisfying:

$$e^{-(1+\epsilon_t)\frac{t}{2} \log t} \leq p(x(t)) \leq e^{-(1-\epsilon_t)\frac{t}{2} \log t}.$$

In addition, the measure associated to the typical set $A[\epsilon_t]$ is given by:

$$\mathbb{P}(x(t) \in A[\epsilon_t]) \geq 1 - \delta_t.$$

In consequence,

$$\begin{aligned} (1 - \epsilon_t)(1 - \delta_t) \frac{t}{2} \log t &\leq - \sum_{x(t) \in A[\epsilon_t]} p(x(t)) \log p(x(t)) \\ &\leq (1 + \epsilon_t)(1 - \delta_t) \frac{t}{2} \log t. \end{aligned}$$

Now let us consider that the complement of the typical set $\Omega(t) \setminus A[\epsilon_t]$, whose measure is:

$$\mathbb{P}(x(t) \in \Omega \setminus A[\epsilon_t]) \leq \delta_t,$$

is completely populated by those paths by which:

$$x(t) \in \Omega \setminus A[\epsilon_t], p(x^*(t)) = \min_{\Omega(t)} \{p(x(t))\},$$

therefore, for all paths $x(t) \in \Omega(t)$:

$$-\log p(x(t)) \leq -\log p^*(x(t)).$$

The least probable path is the one that increases the number of tables at every step, having a probability $p^*(x(t))$ of:

$$p^*(x(t)) = \frac{1}{t!},$$

leading to $-\log p^*(x(t)) \sim t \log t$, thanks to the Stirling approximation [64]. In that context:

$$- \sum_{x(t) \in \Omega(t) \setminus A[\epsilon_t]} p(x(t)) \log p(x(t)) \leq \delta t \log t.$$

Collecting the above reasoning, and by observing that the defined entropy is actually the *Shannon path entropy*, $S_\Lambda(\eta(t)) = H(\eta(t))$ [24]:

$$H(\eta(t)) = - \sum_{x(t) \in \Omega(t)} p(x(t)) \log p(x(t)),$$

one has that:

$$\begin{aligned} (1 - \epsilon_t)(1 - \delta_t) \frac{t}{2} \log t &\leq H(\eta(t)) \\ &\leq (1 + \epsilon_t)(1 - \delta_t) \frac{t}{2} \log t + \delta t \log t. \end{aligned}$$

In consequence, by defining $g(t) = \frac{t}{2} \log t$, one is led to:

$$(1 - \delta_t)(1 - \epsilon_t) \leq \frac{H(\eta(t))}{g(t)} \leq (1 - \delta_t)(1 + \epsilon_t) + 2\delta_t.$$

Since we know that $\lim_{t \rightarrow \infty} \epsilon_t = \lim_{t \rightarrow \infty} \delta_t = 0$ we conclude that:

$$\frac{H(\eta(t))}{g(t)} \rightarrow 1,$$

as we wanted to demonstrate. \square

Appendix D.5. A Different Compact Scale (Λ, G) for the CRP

We finally briefly comment how another compact scale made of different (Λ, g) functions could be used to characterize the typical set and the generalized entropy of the CRP. We avoid the technical details, for the sake of simplicity. We start computing the inverse function of g , g^{-1} :

$$g^{-1}(z) \approx \frac{2z}{\mathbf{W}(2z)},$$

where \mathbf{W} is the *Lambert* function [64], where only the positive, real branch is taken into account. Then we compose it with the log function, thereby defining a new function Λ as:

$$\Lambda(z) = (g^{-1} \circ \log)(z) \sim \frac{2 \log(z)}{\mathbf{W}(2 \log(z))}. \quad (\text{A6})$$

We observe that Λ as above defined is a strictly growing, concave function with continuous second derivatives. Clearly, $(g^{-1} \circ g)(t) \sim t$. Therefore, as a direct consequence of Theorem A1, the CRPM is compact in (Λ, t) :

$$\frac{1}{g^{-1}(\frac{t}{2} \log t)} (g^{-1} \circ \log) \left(\frac{1}{p(\eta(t))} \right) = \frac{1}{t} \Lambda \left(\frac{1}{p(\eta(t))} \right) \rightarrow 1.$$

(in probability). In consequence, the size of the typical set and the typical probabilities can be approximated from the following generalized entropy S_{Λ} :

$$S_{\Lambda}(\eta(t)) = \sum_{x(t) \in \Omega(t)} p(x(t)) \frac{2 \log \left(\frac{1}{p(x(t))} \right)}{\mathbf{W} \left[2 \log \left(\frac{1}{p(x(t))} \right) \right]}. \quad (\text{A7})$$

where we explicitly wrote it in terms the functional form of $\Lambda = g^{-1} \circ \log$ as defined in Equation (A6).

References

1. Morowitz, H.J. *Energy Flow in Biology: Biological Organization as a Problem in Thermal Physics*; Academic Press: London, UK, 1968.
2. Bialek, W. *Biophysics: Searching for Principles*; Princeton University Press: Princeton, NJ, USA, 2012.
3. Solé, R.; Goodwin, B. *Signs of Life*; Basic Books; Perseus Group: New York, NY, USA, 2000.
4. Corominas-Murtra, B.; Seoane, L.; Solé, R. Zipf's Law, unbounded complexity and open-ended evolution. *J. R. Soc. Interface* **2018**, *15*, 20180395. [CrossRef] [PubMed]
5. Maynard Smith, J.; Szathmáry, E. *The Major Transitions in Evolution*; Freeman: Oxford, UK, 1995.
6. Bonner, J.T. *The Evolution of Complexity by Means of Natural Selection*; Princeton University Press: Princeton, NJ, USA, 1988.
7. Wolpert, L.; Jessell, T.; Lawrence, P.; Meyerowitz, E.; Robertson, E.; Smith, J. *Principles of Development*, 3rd ed.; Oxford University Press: Oxford, UK, 2007.
8. Schuster, P. How does complexity arise in evolution? In *Evolution and Progress in Democracies. Theory and Decision Library*; Götschl, J., Ed.; Springer: Dordrecht, The Netherlands, 1996; Volume 31.
9. Bedau, M.A.; McCaskill, J.S.; Packard, N.H.; Rasmussen, S.; Adami, C.; Green, D.G.; Ikegami, T.; Kaneko, K.; Ray, T.S. Open Problems in Artificial Life. *Artif. Life* **2000**, *6*, 363–376. [CrossRef] [PubMed]
10. Ruíz-Mirazo, K.; Peretó, J.; Moreno, A. A universal definition of life: Autonomy and open-ended evolution. *Orig. Life Evol. Biosph.* **2004**, *34*, 323–346. [CrossRef] [PubMed]
11. Ruíz-Mirazo, K.; Umérez, J.; Moreno, A. Enabling conditions for “open-ended” evolution. *Biol. Philos.* **2008**, *23*, 67–85. [CrossRef]
12. Day, T. Computability, Gödel's incompleteness theorem, and an inherent limit on the predictability of evolution. *J. R. Soc. Interface* **2012**, *9*, 624–639. [CrossRef] [PubMed]
13. Packard, N.; Bedau, M.A.; Channon, A.; Ikegami, T.; Rasmussen, S.; Stanley, K.O.; Taylor, T. An Overview of Open-Ended Evolution: Editorial Introduction to the Open-Ended Evolution II Special Issue. *Artif. Life* **2019**, *25*, 93–103. [CrossRef]
14. Pattee, H.H.; Sayama, H. Evolved Open-Endedness, Not Open-Ended Evolution. *Artif. Life* **2019**, *25*, 4–8. [CrossRef]
15. Dietrich, J.E.; Hiiragi, T. Stochastic patterning in the mouse pre-implantation embryo. *Development* **2007**, *134*, 4219–4231. [CrossRef] [PubMed]
16. Maitre, J.L.; Turlier, H.; Illukkumbura, R.; Eismann, B.; Niwayama, R.; Nédélec, F.; Hiiragi, T. Asymmetric division of contractile domains couples cell positioning and fate specification. *Nature* **2016**, *536*, 344–348. [CrossRef]

17. Giammona, J.; Campàs, O. Physical constraints on early blastomere packings. *PLoS Comput. Biol.* **2021**, *17*, e1007994. [[CrossRef](#)] [[PubMed](#)]
18. Tria, F.; Loreto, V.; Servedio, V.D.P.; Strogatz, S.H. The dynamics of correlated novelties. *Sci. Rep.* **2014**, *4*, 5890. [[CrossRef](#)]
19. Loreto, V.; Servedio, V.D.P.; Strogatz, S.H.; Tria, F. Dynamics on expanding spaces: Modeling the emergence of novelties. *Creat. Universality Lang.* **2016**, 59–83. [5](#). [[CrossRef](#)]
20. Iacopini, I.; Di Bona, G.; Ubaldi, E.; Loreto, V.; Latora, V. Interacting Discovery Processes on Complex Networks. *Phys. Rev. Lett.* **2020**, *125*, 248301. [[CrossRef](#)]
21. Corominas-Murtra, B.; Hanel, R.; Thurner, S. Understanding scaling through history-dependent processes with collapsing sample space. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5348–5353. [[CrossRef](#)] [[PubMed](#)]
22. Corominas-Murtra, B.; Hanel, R.; Thurner, S. Sample space reducing cascading processes produce the full spectrum of scaling exponents. *Sci. Rep.* **2017**, *7*, 11223. [[CrossRef](#)]
23. Fujii, K.; Berengut, J.C. Power-Law Intensity Distribution of γ -Decay Cascades: Nuclear Structure as a Scale-Free Random Network. *Phys. Rev. Lett.* **2021**, *126*, 102502. [[CrossRef](#)] [[PubMed](#)]
24. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: New York, NY, USA, 2012.
25. Ash, R.B. *Information Theory*; Dover Publications: New York, NY, USA, 2012.
26. Pathria, R.K. *Statistical Mechanics*; Oxford University Press: Oxford, UK, 2002.
27. Pitowsky, I. Typicality and the Role of the Lebesgue Measure in Statistical Mechanics. In *Probability in Physics*; Ben-Menahem, Y., Hemmo, M., Eds.; Springer: Berlin, Germany, 2012; pp. 41–58.
28. Lebowitz, J.L. Macroscopic Laws, Microscopic Dynamics, Time's Arrow and Boltzmann's Entropy. *Physica A* **1993**, *194*, 1–27. [[CrossRef](#)]
29. Ledoux, M. *The Concentration of Measure Phenomenon*; American Mathematical Society: Providence, RI, USA, 2005.
30. Battermann, R. *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*; Oxford University Press: Oxford, UK, 2001.
31. Frigg, R. Typicality and the Approach to Equilibrium in Boltzmannian Statistical Mechanics. *Philos. Sci.* **2009**, *76*, 997–1008. [[CrossRef](#)]
32. Kač, M. A history-dependent random sequence defined by Ulam. *Adv. Appl. Math.* **1989**, *10*, 270–277. [[CrossRef](#)]
33. Pitman, J. *Combinatorial Stochastic Processes*; Springer: Berlin, Germany, 2006.
34. Clifford, P.; Stirzaker, D. History-dependent random processes. *Proc. R. Soc. Lond. A* **2008**, *464*, 1105–1124. [[CrossRef](#)]
35. Biró, T.; Nédá, Z. Unidirectional random growth with resetting. *Phys. A Stat. Mech. Its Appl.* **2018**, *499*, 335–361. [[CrossRef](#)]
36. Jensen, H.J.; Pazuki, R.H.; Pruessner, G.; Tempesta, P. Statistical mechanics of exploding phase spaces: Ontic open systems. *J. Phys. A Math. Theor.* **2018**, *51*, 375002. [[CrossRef](#)]
37. Korbel, J.; Lindner, S.D.; Hanel, R.; Thurner, S. Thermodynamics of structure-forming systems. *Nat. Commun.* **2021**, *12*, 1127. [[CrossRef](#)]
38. Gray, R.M.; Davisson, L.D. Source coding theorems without the ergodic assumption. *IEEE Trans. Inform. Theory* **1974**, *20*, 502–516. [[CrossRef](#)]
39. Visweswariah, K.; Kulkarni, S.R.; Verdu, S. Universal coding of nonstationary sources. *IEEE Trans. Inf. Theory* **2000**, *46*, 1633–1637. [[CrossRef](#)]
40. Vu, V.Q.; Yu, B.; Kass, R.E. Information in the Non-Stationary Case. *Neural Comput.* **2009**, *21*, 688–703. [[CrossRef](#)] [[PubMed](#)]
41. Boashash, B.; Azemi, G.; O'Toole, J. Time-frequency processing of nonstationary signals: Advanced TFD design to aid diagnosis with highlights from medical applications. *IEEE Signal Process. Mag.* **2013**, *30*, 108–119. [[CrossRef](#)]
42. Granero-Belinchón, C.; Roux, S.G.; Garnier, N.B. Information Theory for Non-Stationary Processes with Stationary Increments. *Entropy* **2019**, *21*, 1223. [[CrossRef](#)]
43. Abe, S. Axioms and uniqueness theorem for Tsallis entropy. *Phys. Lett. A* **2000**, *271*, 74–79. [[CrossRef](#)]
44. Hanel, R.; Thurner, S. A comprehensive classification of complex statistical systems and an axiomatic derivation of their entropy and distribution functions. *Europhys. Lett.* **2011**, *93*, 20006. [[CrossRef](#)]
45. Enciso, A.; Tempesta, P. Uniqueness and characterization theorems for generalized entropies. *J. Stat. Mech. Theory Exp.* **2017**, *12*, 123101. [[CrossRef](#)]
46. Tempesta, P. Beyond the Shannon–Khinchin formulation: The composability axiom and the universal-group entropy. *Ann. Phys.* **2016**, *365*, 180–197. [[CrossRef](#)]
47. Tempesta, P. Formal groups and Z-entropies. *Proc. R. Soc. Lond. A* **2016**, *472*, 20160143. [[CrossRef](#)]
48. Thurner, S.; Corominas-Murtra, B.; Hanel, R. The three faces of entropy for complex systems—Information, thermodynamics and the maxent principle. *Phys. Rev. E* **2017**, *96*, 032124. [[CrossRef](#)] [[PubMed](#)]
49. Jizba, P.; Korbel, J. When Shannon and Khinchin meet Shore and Johnson: Equivalence of information theory and statistical inference axiomatics. *Phys. Rev. E* **2020**, *101*, 042126. [[CrossRef](#)]
50. Korbel, J.; Jizba, P. Maximum Entropy Principle in Statistical Inference: Case for Non-Shannonian Entropies. *Phys. Rev. Lett.* **2019**, *122*, 120601.
51. Jizba, P.; Korbel, J. On the Uniqueness Theorem for Pseudo-Additive Entropies. *Entropy* **2017**, *19*, 605. [[CrossRef](#)]
52. Hanel, R.; Thurner, S. When do generalized entropies apply? How phase space volume determines entropy. *Europhys. Lett.* **2011**, *96*, 50003. [[CrossRef](#)]

53. Jensen, H.J.; Tempesta, P. Group entropies: From phase space geometry to entropy functionals via group theory. *Entropy* **2018**, *20*, 804. [\[CrossRef\]](#)
54. Korbel, J.; Hanel, R.; Thurner, S. Classification of complex systems by their sample-space scaling exponents. *New J. Phys.* **2018**, *20*, 093007. [\[CrossRef\]](#)
55. Korbel, J.; Hanel, R.; Thurner, S. Information geometry of scaling expansions of non-exponentially growing configuration spaces. *Eur. Phys. J. Spec. Top.* **2020**, *229*, 787–807. [\[CrossRef\]](#)
56. Hanel, R.; Thurner, S. Generalized (c,d)-entropy and aging random walks. *Entropy* **2013**, *15*, 5324–5337. [\[CrossRef\]](#)
57. Nicholson, S.B.; Alaghemandi, M.; Green, J.R. Learning the mechanisms of chemical disequilibria. *J. Chem. Phys.* **2016**, *145*, 084112. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Balogh, S.G.; Palla, G.; Pollner, P.; Czégel, D. Generalized entropies, density of states, and non-extensivity. *Sci. Rep.* **2020**, *10*, 15516. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Gardiner, C.W. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*; Springer: Berlin, Germany, 1983.
60. Feller, W. *An Introduction to Probability Theory and Its Applications*; Wiley: New York, NY, USA, 1991; Volumes 1 and 2.
61. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech.* **1948**, *27*, 379–423. [\[CrossRef\]](#)
62. Khinchin, A. *Mathematical Foundations of Information Theory*; Dover: New York, NY, USA, 1957.
63. Bassetti, B.; Zarei, M.; Cosentino Lagomarsino, M.; Bianconi, G. Statistical mechanics of the “Chinese restaurant process”: Lack of self-averaging, anomalous finite-size effects, and condensation. *Phys. Rev. E* **2009**, *80*, 066118. [\[CrossRef\]](#)
64. Abramowitz, M.; Stegun, I. *Handbook of Mathematical Functions. National Bureau of Standards; Applied Mathematics Series 55*; U.S. Government Printing Office: Washington, DC, USA, 1964.
65. Abe, S.; Martínez, S.; Pennini, F.; Plastino, A. Nonextensive thermodynamic relations. *Phys. Lett. A* **2001**, *281*, 126–130. [\[CrossRef\]](#)
66. Abe, S. Temperature of nonextensive systems: Tsallis entropy as Clausius entropy. *Phys. A Stat. Mech. Its Appl.* **2006**, *368*, 430–434. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.